

Introduction

Geometry is the study of spatial relationships, such as the familiar assertion from elementary plane Euclidean geometry that, if two triangles have sides of the same lengths, then they are “congruent.” What does congruent mean here? One possibility, which is rather abstract and very much in the spirit of the axiomatic approach usually attributed to Euclid, is to say:

Call two straight line segments “congruent” if they have the same length. Call two triangles “congruent” if each side of one can be paired with a side of equal length on the other.

A more concrete way to say this is that one can take the first line segment and move it “rigidly” from wherever it is in the plane to wherever the second line segment is, in such a way that it overlies the second exactly; similarly, one can take the first triangle and move it rigidly so that it overlies the second exactly.

One of the key insights of modern geometry is that the rigid motions are precisely those maps from the plane onto itself that preserve lengths of line segments. The point is that it is just the notion of “length” that counts: all the angles, the area and other stuff follow once you preserve lengths.

The simplest rigid motion of the plane is reflection in a line: that is, pick a line and, for every point off the line, draw the perpendicular to the line through the point and find the point on the other side that is the same distance from the line; points on the line itself are fixed.

Clearly, following one rigid motion by another results in a rigid motion, and any rigid motion can be reversed to get back where you started, so, in the language of modern algebra, the rigid motions form a group under composition generated by reflections. Historically, the term “symmetry” is used in place of “rigid motion,” so you will see a large literature on “groups of symmetries.”

This analysis raises an obvious question:

Is there some other notion of “length” with a corresponding rigid motion group that is also “natural” and that leads to a geometry that is different from Euclid’s? What are the analogues of the notions like “polygon”, “area” and “interior angle” that these motions preserve?

The short answer is yes, and the long answer is the heart of this book.

The geometry we will study in this book is called *hyperbolic geometry*. It has the same notion of angle as Euclidean geometry and the rigid motions in this geometry preserve angles.

To motivate our development of this geometry, we show that, by adding a single point to the Euclidean plane, we have another kind of symmetry, “reflection” or “inversion” in a circle: pick any point except the center of the circle and draw the ray from the center through the point; the reflected point is that point on the ray whose distance from the center is the square of the radius of the circle divided by the distance of the given point to the center. This symmetry preserves angles but not Euclidean length. The image of a line under an inversion may be a circle and vice versa so that, in a geometry whose group of motions contains inversions, lines and circles are considered the same kind of object.

In Chapter 2, we begin with the unit disk together with the group of motions that take the disk onto itself and preserve angles. Each such motion is a composition of inversions in circles orthogonal to the boundary of the disk. We develop a notion of length, or distance, for which this group is the group of rigid motions. In this geometry, the shortest path between two points in the disk is along a circle passing through the points and orthogonal to the unit circle. Points on the unit circle are at infinite distance from points inside the disk. The disk with this distance function is called the *hyperbolic plane* and the distance function is called the *hyperbolic metric*.

The most natural way to study this geometry is to use the language of complex numbers and some of the theorems about holomorphic functions – that is, complex valued functions of a complex variable that have a complex derivative. For example, a very important theorem from complex analysis, which we will use over and over again, is called the “generalized Schwarz lemma”. It says that any holomorphic map of the disk into itself that is not a rigid motion decreases hyperbolic distance.

We will also be interested in studying geometries for domains that are open subsets of the Euclidean plane. One natural geometry is obtained by restricting the usual Euclidean notion of length. This geometry, however, doesn’t really capture some of the intrinsic properties of the domain. For example, rigid motions of the plane do not necessarily map the domain to itself; points on

the boundary of the domain will be a finite distance from interior points. These domains do admit other geometries, and in particular a hyperbolic geometry. The idea behind this book is to study a set of related geometries we can put on plane domains that are defined in terms of the derivatives of their distance functions called *densities*. These are more general than the metrics in that they enable us to measure the lengths of paths, the areas of triangles etc. Moreover, the metric may be recovered from the density by integrating over paths. These densities we consider are called “conformal densities” because they have the property that angle measurement is the same for all of the geometries they define.

To begin our discussion on conformal densities for arbitrary plane domains, we need standard tools from complex variable theory and topology. We develop these tools in Chapters 3, 4. We state all of the standard results we need, but we prove only those that are most important or whose proofs involve relevant techniques.

In Chapters 5 and 6 we investigate the symmetry groups of the Euclidean and hyperbolic planes in detail. We show how to use topology to identify plane domains with subgroups of rigid motions. Then in Chapter 7 we define the density that determines the hyperbolic geometry of an arbitrary domain and in Chapters 8, 9 and 10 we define generalizations of this density.

The contents of the first seven chapters of this book are, for the most part, part of the standard mathematical lore. Our approach to defining the hyperbolic metric for an arbitrary plane domain, however, lends itself to generalization and in Chapters 8, 9 and 10 we present this generalization. Some of this material has not appeared before, or has only appeared recently.

In Chapters 11, 12 and 13, we turn to applications. We look at iterated function systems from a given plane domain to a subdomain. The characteristics of the limiting behavior of these systems are controlled by the geometry of the domain and subdomain. This material is the subject of ongoing research and contains both previously unpublished results and open problems.

In general it is a very difficult problem to find an explicit formula for the hyperbolic metric of an arbitrary domain. It is possible, however, to get estimates on the metric by using inclusion mappings. We address this problem in Chapters 14 and 15. In Chapter 14, we get estimates on the hyperbolic metric for various domains by applying the generalized Schwarz lemma to the inclusion map of the domain into the twice punctured plane. We also present an equivalent characterization of the hyperbolic metric which gives another method of finding estimates for the metric. Finally, in Chapter 15 we obtain estimates on the hyperbolic metric for domains called uniformly perfect. In general, one can get estimates using inclusion mappings from the

disk into the domain but they only work in one direction. The densities of uniformly perfect domains are comparable to the reciprocal of the distance to the boundary and for these we get estimates in both directions. The last chapter is an appendix in which we present a brief survey of elliptic functions.

The exercises throughout the book should be considered as an integral part of the text material because they contain the statements of many things that are used in the text. In the later chapters they also contain open problems.

We envision this book to be used in several different ways. The first author has used Chapters 1, 4 and 5 as the basis of a junior level undergraduate course. The authors have used the first seven chapters as a one semester second year graduate course. They are currently using the remaining chapters for the second semester of a second year graduate course designed to introduce graduate students to potential research problems.

This book grew out of a seminar for graduate students at the Graduate Center of the City University of New York. While we were writing this book, many of the students read and lectured on the material and gave us invaluable feedback. We would like to thank all the participants in the seminar for all their input. The members of the seminar are Orlando Alonso, Anthony Conte, Ross Flek, Frederick Gardiner, Sandra Hayes, Jun Hu, Yunping Jiang, Greg Markowsky, Bill Quattromani, Kourosh Tavakoli, Donald Taylor, Shenglan Yuan and Zhe Wang. We would also like to thank Ross Flek for making most of the figures in the book and Fred Gardiner for his encouragement. The book would also never have come to being without the strong support of our families, Jonathan Brezin and Ljiljana and Emily Lakic.

1

Elementary transformations of the Euclidean plane and the Riemann sphere

1.1 The Euclidean metric

In most of this book we will be interested in the hyperbolic metric and its generalizations. The hyperbolic metric has its own inner beauty, but it also serves as an important tool in the study of many different areas of mathematics and other sciences.

Before we start exploring hyperbolic geometry in Chapter 2, we get some orientation from the geometry that everyone is very familiar with, Euclidean geometry. We denote the plane by \mathbb{R}^2 or \mathbb{C} and a point in the plane either by its Cartesian coordinates, (x, y) , by its polar coordinates (r, θ) or by the complex number $z = x + iy = re^{i\theta}$, depending on which is most convenient. We denote the modulus $|z|$ by r and the argument $\arg z$ by θ where $|z| = r = \sqrt{x^2 + y^2}$ and $\arg z = \theta = \arctan(y/x)$. It is also convenient sometimes to think of the point z as the vector from the origin to z .

The argument θ is the angle measured from the positive x -axis to the vector z . The plane has an *orientation*: the argument is positive if the direction from the x -axis to z is counterclockwise and negative otherwise.

The complex conjugate of z is $\bar{z} = x - iy$; it has the same modulus as z and its argument has the opposite sign.

The Euclidean length of the vector z is its modulus and the distance between points $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ in the plane is the modulus of their difference,

$$d(z_1, z_2) = |z_2 - z_1| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

This definition of distance satisfies the three requisite conditions for a distance function:

- $d(z_1, z_2) \geq 0$ with equality if and only if $z_1 = z_2$;

- $d(z_1, z_2) = d(z_2, z_1)$;
- $d(z_1, z_3) \leq d(z_1, z_2) + d(z_2, z_3)$.

Any space X , together with a distance function d_X satisfying these properties, is called a *metric space*.

Definition 1.1 A finite curve in a metric space is a continuous map of the unit interval $[0, 1]$ into a metric space; an infinite curve is a continuous map of the real line $(-\infty, \infty)$ into a metric space; a semi-infinite curve is a continuous map of the half line $[0, \infty)$ into a metric space. The word curve denotes any one of these.

We can use the third condition for the distance function to characterize straight lines in Euclidean geometry.

Definition 1.2 We say a curve is a straight line or geodesic in the Euclidean plane if for every triple of points z_1, z_3, z_2 on the curve with z_3 between z_1 and z_2 we have

$$d(z_1, z_2) = d(z_1, z_3) + d(z_3, z_2).$$

If the curve satisfying this condition is finite we call it a straight line segment or geodesic segment; if it is a semi-infinite curve satisfying the condition it is called a ray or geodesic ray and if it is infinite it is called an infinite geodesic or an extended line.

For readability, when it is clear what kind of straight line or geodesic we mean we simply call it a line or geodesic.

Exercise 1.1 Verify that the function $d(z_1, z_2)$ is a distance function on \mathbb{C} and also on any subdomain $\Omega \subset \mathbb{C}$.

Exercise 1.2 Let $z_1 = 1$ and $z_2 = i$. Evaluate the distance from z_1 to z_2 and find the formula for the geodesic segment joining z_1 and z_2 .

Exercise 1.3 Show that, for any point z and any extended line l , there is a unique point w on l closest to z . The point w is called the projection from z onto l .

1.2 Rigid motions

The Euclidean geometry of the plane is defined by the following maps of the plane onto itself.

Definition 1.3 A rigid motion of the plane is a one to one map f of the plane onto itself such that, for any two points z_1, z_2 ,

$$|f(z_1) - f(z_2)| = |z_1 - z_2|.$$

A rigid motion is also called an isometry.

Compositions of rigid motions are again rigid motions and, since they are one to one and onto, they are invertible. We can therefore talk about the group of rigid motions of the plane.

The reflections in the x - and y -axes respectively are rigid motions given by the maps $R_x : z \mapsto \bar{z}$ and $R_y : z \mapsto -\bar{z}$ respectively. More generally, let l be any line in the plane \mathbb{C} . For any point $z \in \mathbb{C}$, there is a point w on l closest to z , the projection from z to l (see Exercise 1.3). Define the reflection R_l to be the map that sends a point z to the point $R_l(z)$ on the line through z and w on the opposite side of l such that $d(z, w) = d(R_l(z), w)$. Note that R_l sends a point on l to itself. Clearly a reflection is its own inverse; that is $R_l R_l$ is the identity map Id . Any transformation that is its own inverse is called an *involution*.

Two other types of rigid motions are *translations* given by maps of the form $T_{z_0} : z \mapsto z_0 + z$ and *rotations* given by maps of the form $R_{z_0, \alpha} : z \mapsto z_0 + (z - z_0)e^{i\alpha}$.

We state the properties of rigid motions as Exercises 1.4 to 1.12.

Exercise 1.4 If l_1 and l_2 are lines in \mathbb{C} such that the angle from l_1 to l_2 is α , then the angle from $R_x(l_1)$ to $R_x(l_2)$ is $-\alpha$ and similarly for the images under R_y . In other words, these reflections reverse orientation.

Exercise 1.5 Prove the proposition: A rigid motion is uniquely determined by what it does to three points that do not lie in the same line. Hint: First prove that, if a rigid motion fixes two points, it fixes every point on the line joining them.

Exercise 1.6 Prove the proposition: If l and m are two lines in the plane and R_l and R_m are reflections in these lines then the composition $R_m R_l$ is a rotation or a translation depending on whether the lines intersect or are parallel.

Exercise 1.7 Prove the proposition: If l and m are two lines in the plane and R_l and R_m are reflections in these lines then the composition $R_m R_l R_m$ is a reflection about the line $l' = R_m(l)$.

Exercise 1.8 Prove the proposition: Translations preserve orientation.

Exercise 1.9 Prove the proposition: Rotations preserve orientation.

Exercise 1.10 Prove the proposition: Reflections reverse orientation. Hint: Show that every reflection can be written as a conjugation of the reflection R_x by a translation and rotation.

Exercise 1.11 Prove the proposition: If a rigid motion has a single fixed point it is a rotation; if it fixes two points, but does not fix every point, it is a reflection; if it has no fixed points it is either a translation or a translation followed by a reflection.

Definition 1.4 *A rigid motion that is a translation followed by a reflection is called a glide reflection. If the reflection is in a line perpendicular to the vector of translation, the motion reduces to a reflection and the glide reflection is trivial.*

Exercise 1.12 Prove the proposition: Every rigid motion can be written as the composition of at most three reflections. The only rigid motion that cannot be written as a composition of fewer than three reflections is a non-trivial glide reflection. Hint: Write the motion as a composition of reflections and use the fact that reflections about any line are involutions.

Exercise 1.13 Prove the proposition: The group of translations and rotations constitute the full group of orientation preserving rigid motions of the plane.

1.2.1 Scaling maps

In the previous part of this section we looked at the rigid maps of the plane. These maps preserve sizes and shapes. Another map that preserves shapes but changes sizes is the scaling map; that is, contraction or stretching by the same amount in every direction. This map is described by the formula

$$S_c(z) = cz, c(\neq 0) \in \mathbb{C}.$$

The inverse of the scaling map is clearly $S_{1/c}$. Note that the scaling map preserves orientation.

The scaling map can be composed with any rigid motion and the result is also a map that preserves shapes. Such a map is called a *similarity* and the set of such maps forms the group of *similarities* that contains the rigid motions as a subgroup. The subgroup of the similarities that contains all maps of the form $z_0 + cz$, where $z_0, c(\neq 0) \in \mathbb{C}$, is called the *group of complex affine maps* or more simply the group of affine maps.

Exercise 1.14 Show that the similarities take straight lines to straight lines and preserve or reverse the angles between lines.

Exercise 1.15 Show that the affine maps are precisely the subgroup of orientation preserving similarities.

Exercise 1.16 Let $g(z) = 1 + 6z$ be an affine map. What is the image of the unit disk under this map?

1.3 Conformal mappings

Complex affine maps are not the only orientation preserving maps of the plane that preserve angles. Since lines are not necessarily mapped to lines, we need to define what we mean by preservation of angles.

Definition 1.5 A plane domain Ω is an open connected subset of the complex plane \mathbb{C} .

Definition 1.6 If two differentiable curves γ_1 and γ_2 in the domain Ω intersect at the point z_0 , the angle between them is defined as the angle measured from the tangent to γ_1 to the tangent to γ_2 .

We then define

Definition 1.7 A differentiable map f from a plane domain Ω in \mathbb{C} to another plane domain X is called conformal at $p \in \Omega$ if, for every pair of differentiable curves γ_1 and γ_2 intersecting at p , the angle between γ_1 and γ_2 is equal to the angle between the curves $f(\gamma_1)$ and $f(\gamma_2)$ at $f(p)$. If the angle between the curves $f(\gamma_1)$ and $f(\gamma_2)$ at $f(p)$ is the negative of the angle between γ_1 and γ_2 then f is called anti-conformal at p .

In this chapter we will work with a special class of maps.

Definition 1.8 An invertible map f from a plane domain Ω in \mathbb{C} onto another plane domain X is called a homeomorphism if both f and its inverse f^{-1} are continuous.

Definition 1.9 A homeomorphism f from a plane domain Ω in \mathbb{C} onto another plane domain X is called a diffeomorphism if both f and its inverse f^{-1} are differentiable.

Definition 1.10 A diffeomorphism f from a plane domain Ω in \mathbb{C} to another plane domain X is called conformal (or a conformal homeomorphism) if it is conformal at every $p \in \Omega$. It is called anti-conformal if it is anti-conformal at every $p \in \Omega$.

To find a formula to check whether a map is conformal or anti-conformal we need to use coordinates. We will get formulas in both Cartesian and complex coordinates.

Using Cartesian coordinates we denote the points in Ω by (x, y) and points in X by (u, v) and write $f(x, y) = (u(x, y), v(x, y))$.

It is an exercise from calculus that, if all the directional derivatives at a point have the same magnitude, the map preserves the magnitudes of the angles at that point; that is, the map on tangent vectors is a scaling map. In particular, if we write $f(x, y) = u(x, y) + iv(x, y)$, and use the notation f_x, f_y, u_x, u_y , and v_x, v_y for partial derivatives, this means that

$$f_x = u_x + iv_x = -if_y = -iu_y + v_y.$$

Equating real and imaginary parts we get

$$u_x = v_y \text{ and } u_y = -v_x.$$

These equations are called the *Cauchy–Riemann equations* and they are satisfied whenever the map is conformal.

Since we require that f be invertible and orientation preserving, the Jacobian $J(f) = u_x v_y - v_x u_y > 0$. If the orientation is reversed $J(f) < 0$. The condition for anti-conformality is then given by $f_x = if_y$ and we get equations

$$u_x = -v_y \text{ and } u_y = v_x.$$

If we use differentials, the tangents at (x, y) are given by (dx, dy) and the tangents at $(u(x, y), v(x, y))$ are given by (du, dv) and we have

$$du = u_x dx + u_y dy \text{ and } dv = v_x dx + v_y dy.$$

The differential of the map f is

$$df = f_x dx + f_y dy.$$

There is another formal notation that makes these computations neater. Set $z = x + iy$, $\bar{z} = x - iy$, $dz = dx + idy$, $d\bar{z} = dx - idy$ and $f(z) = f(x, y)$. Then write

$$f_z(z) = \frac{1}{2}(f_x(z) - if_y(z)) \text{ and}$$

$$f_{\bar{z}}(z) = \frac{1}{2}(f_x(z) + if_y(z)).$$