CONTENTS

CHAPTER 1

Introduction

The goal of this text is to illustrate how tools from dynamical systems can be brought to bear on rigidity questions in Riemannian geometry, in particular questions about

- the lengths of closed geodesics in negatively curved surfaces,
- metrics on the torus without conjugate points, and
- numerical invariants of geodesic flows for negatively curved manifolds.

We will present three results that give insight into these topics: the Otal–Croke proof of marked length spectrum rigidity for surfaces [272, 96], the Burago–Ivanov proof of the Hopf conjecture on tori [72], and A. Katok's proof of the entropy rigidity conjecture on surfaces [207]. We state these results later in this chapter, in §1.2.

First we put these results in a broader context, with an (admittedly personal) view of the role of rigidity in dynamics and geometry. Most of the technical material presented in the next section will be revisited more formally in later chapters, and we have referenced this material accordingly. We have chosen to focus much of this discussion through the lens of surfaces, as many of the basic dynamical and geometric themes emerge already in this restricted setting.

1.1. Rigid Structures in Geometry and Dynamics: Some Background

Rigidity theory concerns itself with finding the sharp edge between predictability and freedom. Perhaps the first rigidity result in mathematics is the classification of Platonic solids, a result credited to the early Greek mathematician Theaetetus. Not only do Platonic solids exist as isolated, gem-like structures, there are only finitely many of them, despite their a priori description as polyhedra composed of fixed regular polygons, with no restriction on the number of sides.

In the geometric context of this text, we concern ourselves with Riemannian manifolds equipped with some additional data, such as the sign of the curvature or topological type. A property of such a manifold is *rigid* if it is destroyed by any nontrivial perturbation of the metric, or more strongly, if it is not found in any other Riemannian manifold within a given class. Rigidity is thus a contextual notion: a property can be rigid within one class of manifolds, but not in another.[1]

[1] The definition of rigidity is itself not rigidly fixed: "The word rigidity is some kind of tarte à la crème, a bit like hyperbolic, you can choose its meaning as you wish."—Étienne Ghys.

The very notion of rigidity has a strong geometric flavor, suggesting stiffness. Diamonds owe their strength to the rigidity of their molecular structure. Controlled rigidity—that is, flexing only along certain directions—allows suspension bridges to survive high winds. Rigidity questions appear early in the history of Riemannian geometry: one of the oldest conjectures in the subject posits that a smoothly embedded, closed surface in $\mathbb{R}^3$ cannot be isometrically deformed. This is known to be true for positively curved surfaces but is open in general.

Rigidity is also a valuable notion in representation theory, where rigid linear representations of discrete groups often correspond to rigid geometric structures on manifolds. An early example from the 1960s is Selberg–Calabi–Weil rigidity, which states that, except for clearly defined exceptions, irreducible lattices in semisimple Lie groups cannot be infinitesimally deformed in a way that retains their algebraic structure. Subsequently, many of these lattices were understood to be *globally rigid:* they cannot be deformed, and moreover their representations are unique (up to conjugacy in the Lie group). By contrast, the lattice $\mathbb{Z}^n$ in $\mathbb{R}^n$ (a *reducible* lattice when $n \geqslant 2$) can be deformed by tweaking the generators.

These algebraic facts can be reinterpreted as geometric ones in special cases: in dimension at least 3, a compact manifold of constant negative curvature carries no other constant curvature metric (Mostow rigidity), whereas on the torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ there is a continuous family of flat metrics (i.e., whose curvature tensor vanishes identically), no two of which are isometric.[2]

The history of smooth dynamics and geometry are intermingled, and rigidity questions also arise naturally in dynamics. Many of the motivating examples from the formative years of dynamical systems are highly structured, the dynamical counterparts of Platonic solids. These early examples are both geometric and algebraic in nature and are closely tied to the dynamics of the geodesic flow.

1.1.1. The Geodesic Flow. If M is a d-manifold endowed with a complete Riemannian metric $\langle\cdot,\cdot\rangle$, then the $2d$-dimensional tangent bundle TM carries an associated flow $\varphi_t : TM \to TM$, called the *geodesic flow*, which is defined as follows. To each $v \in TM$ is associated a unique locally length minimizing curve called a geodesic $\gamma_v : \mathbb{R} \to M$ with the property that $\dot\gamma_v(0) = v$. Given $t \in \mathbb{R}$, and $v \in TM$, we define

$$\varphi_t(v) := \dot\gamma_v(t),$$

as shown in Figure 1.1.

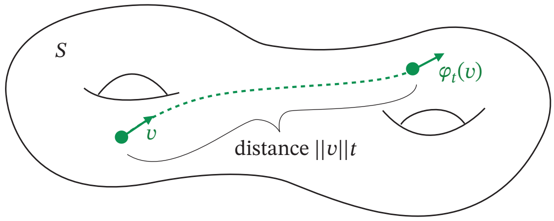


Figure 1.1. The geodesic flow for a surface S.

[2]Indeed, the flat metrics on the d-torus are parametrized by points in the double coset space $\mathrm{SL}(d,\mathbb{Z})\backslash \mathrm{SL}(d,\mathbb{R})/\mathrm{SO}(d)$.

Geodesics have constant speed, meaning that $||\varphi_t(v)||$ is constant for all $t \in \mathbb{R}$, equal to $||v||$. To understand the flow on the full tangent bundle TM, it is sufficient to restrict to the *unit tangent bundle*

$$T^1M := \{v \in TM : ||v|| = 1\},$$

a $(2d-1)$-manifold that is invariant under the flow (see Chapter 2, §2.1.4). The geodesic flow $\varphi_t : T^1M \to T^1M$ encodes dynamically all of the geodesics in M and their properties. For example, *periodic orbits* of the flow on T^1M project under $\pi : T^1M \to M$ to *closed geodesics* on M.

The geodesic flow can alternatively be defined as a type of flow called a *Hamiltonian flow*. Loosely speaking, a Hamiltonian flow on a symplectic manifold N is specified by the condition that it preserves both the symplectic form and a fixed "kinetic energy" function $H : N \to \mathbb{R}$ (see Chapter 2, §2.2). If $c \in \mathbb{R}$ is any regular value of H, then the smooth submanifold $H^{-1}(c)$ is invariant under the corresponding Hamiltonian flow. On $H^{-1}(c)$ there is a volume-type measure called the *Liouville measure* that is preserved by the flow.

The tangent bundle TM is a symplectic manifold (i.e., a manifold equipped with a symplectic form),[3] and the geodesic flow on TM is the Hamiltonian flow for the energy function $H : TM \to \mathbb{R}$ given by

$$H(v) = \frac{1}{2}||v||^2.$$

The unit tangent bundle $T^1M = H^{-1}(1/2)$ is a smooth level set for H, and the Liouville measure on T^1M preserved by the flow is locally just the product of volume on M with Lebesgue measure on the spherical fibers of T^1M. For compact manifolds (or even manifolds of finite volume), this Liouville measure can be normalized to be a probability measure on T^1M, which we will denote by m_{Liou}. The construction of this measure is described in detail in Chapter 2, §2.2.

Thus geodesic flows present a natural class of (*probability*) *measure-preserving flows*, which, along with their discrete time counterparts (measure-preserving transformations), are the central object of study in classical ergodic theory. The basics of ergodic theory are presented in Chapter 5. The Liouville measure m_{Liou} will play a central role in this book, and we will use methods from ergodic theory to deduce surprising rigidity results about Riemannian manifolds.

Research on geodesic flows began in earnest around the turn of the 20th century, when Jacques Hadamard [176] initiated a study of infinite geodesics in negative curvature. The subject attracted the interest of some of the leading figures of the time, among them Henri Poincaré, George Birkhoff, Marston Morse, and Eberhard Hopf. By the 1930s, the term "geodesic flow" had come into common usage, and the flow was intensively studied as a dynamical object in its own right [182].

1.1.2. The Case of Surfaces. Even in the simplest case of surfaces, the geodesic flow is a rich and fascinating subject from a dynamical point of view.

Let S be a Riemannian 2-manifold, and let $p \in S$. If the curvature $k(p)$ is nonzero, then p has a neighborhood that can be embedded isometrically as a submanifold

[3]The symplectic form on TM is the pullback of the canonical symplectic form on T^*M to TM under the identification given by the metric.

of Euclidean space $\mathbb{R}^3$; the sign of the curvature at p determines the shape of this embedding.[4]

Under a suitable choice of coordinates, a local surface Σ in $\mathbb{R}^3$ is just a graph of a smooth function $s: U \subset \mathbb{R}^2 \to \mathbb{R}$; if a point $p \in \Sigma$ is fixed, the coordinates can be chosen so that $p = (0, 0, s(0))$, and $(0, 0)$ is a critical point for s. The sign of the determinant of the Hessian

$$\mathrm{Hess}_{(0,0)}\, s := \left(\frac{\partial^2 s}{\partial x_i\, \partial x_j}(0,0)\right)_{i,j \in \{1,2\}}$$

determines the sign[5] of the curvature of Σ at p. Curvature is defined formally in Chapter 2, §2.3.1.

By inspecting the images in Figure 1.2, one can see that even on the local level, geodesics behave qualitatively differently according to the sign of the curvature at p. For closed surfaces of *constant* curvature these local differences are magnified to global ones, producing geodesic flows with distinctive dynamical fingerprints. Indeed, the geodesic flows for surfaces of constant curvature present three prototypical types of smooth, volume-preserving flows: elliptic, parabolic, and hyperbolic, as we now explain.

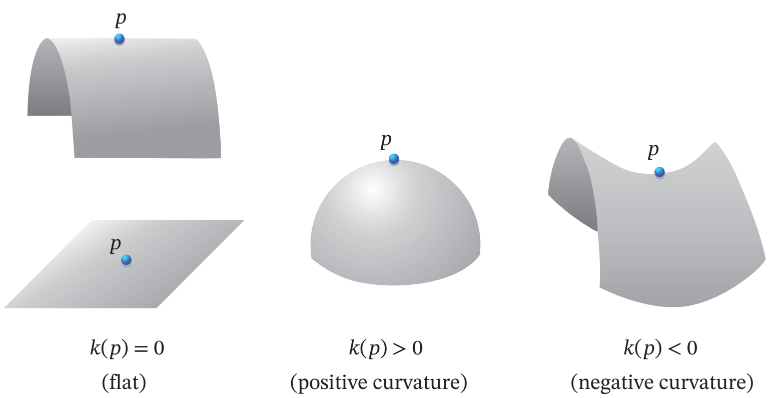


Figure 1.2. The sign of the curvature k determines three different local geometries on a surface. Depicted are surfaces of constant curvature locally embedded in $\mathbb{R}^3$.

Following the lead of Poincaré, for any Riemannian 2-manifold S, we can view the dynamics of the geodesic flow on the 3-manifold T^1S in cross-sections along an orbit. Between any two surfaces transverse to the orbit, there is a local diffeomorphism given by following the orbit of a point in the first surface until it hits the second surface, as depicted in Figure 1.3. For the geodesic flow, there is a natural sense in which these maps between cross-sections preserve area. Moreover, there is a geometrically natural identification of the normal bundle to the orbit with a trivial $\mathbb{R}^2$ bundle (we make precise in Proposition 2.2.8 of the next chapter a natural family of transverse coordinates along any orbit where this viewpoint can be made rigorous). Viewed infinitesimally down an orbit in these coordinates, the derivative of the flow then appears as an area-preserving linear flow on the plane.

[4]Whether *every* surface can be locally isometrically embedded in $\mathbb{R}^3$ is an open question; in particular, if $k(p) = 0$ it is not known whether there is a local isometric embedding of S near p, except in special cases [230]!

[5]The curvature itself is a constant multiple of this determinant.

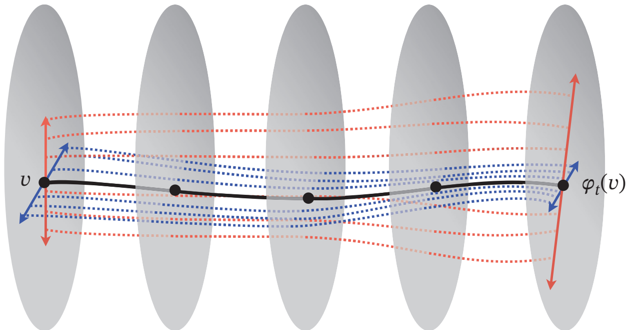
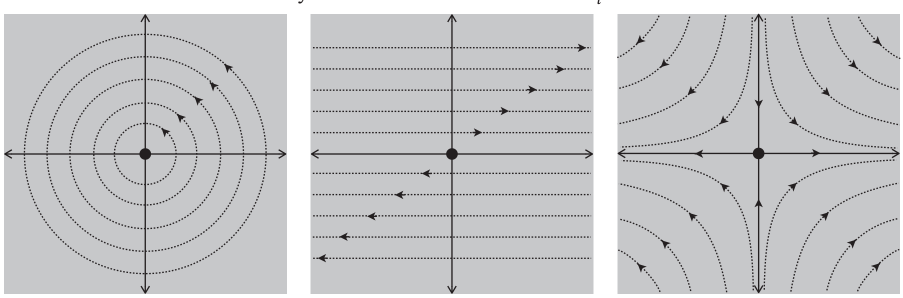


Figure 1.3. Viewing the geodesic flow on T^1S (infinitesimally) in cross-section down the orbit of v.

There are three types of area-preserving linear flows: elliptic, parabolic, and hyperbolic, as described in Figure 1.4. Imagine now constructing the "perfect" elliptic flow: down every orbit, in a naturally defined set of coordinates, the transverse derivative of φ_t is precisely the rotation through angle t. It turns out that such a perfect elliptic flow can be realized as the geodesic flow for the round sphere! Similarly, there are perfect parabolic and hyperbolic flows that are realized as geodesic flows in constant curvature.

Dynamics of the linear flow $x \mapsto A_t x$



$A_t = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$ elliptic

$A_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ parabolic

$A_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$ hyperbolic

Figure 1.4. The three types of dynamics in a linear flow $x \mapsto A_t x$, with $A_t \in \mathrm{SL}(2, \mathbb{R})$: elliptic, parabolic, and hyperbolic. The paradigmatic examples are illustrated; every nontrivial area-preserving linear flow is conjugate to one of these according to the signs of the eigenvalues of its infinitesimal generator.

Let S be a closed, orientable surface with a Riemannian metric with curvature $k: S \to \mathbb{R}$. The integral of k over the surface is determined by the Gauss–Bonnet formula

$$\int_S k\, dA = 2\pi\chi(S), \tag{1.1.1}$$

where dA is the area form determined by the metric, and $\chi(S)$ is the Euler characteristic of S (see Chapter 2, §2.6.3). In the case where $k \equiv k_0$ is constant, the formula simplifies to

$$k_0 = \frac{2\pi\chi(S)}{\text{area}(S)}.$$

Thus the three possibilities for the sign of the curvature are determined by the Euler characteristic, or equivalently by the genus $g(S)=1-\chi(S)/2$. They are the following:[6]

- **The round case: Constant positive curvature** occurs only in genus 0 (when $\chi(S)=2$), that is, on the sphere S^2. The dynamics of the geodesic flow on T^1S^2 are *elliptic:* orbits arrange themselves on invariant tori in a twisting fashion around a central orbit.
- **The flat case: Constant curvature** 0 occurs uniquely in genus 1 (when $\chi(S)=0$), that is, on the torus $\mathbb{T}^2$. The dynamics of the geodesic flow on $T^1\mathbb{T}^2$ are *parabolic:* again, orbits arrange themselves on invariant tori, but in a shearing fashion.
- **The hyperbolic case: Constant negative curvature** occurs on all surfaces of genus at least 2 (negative Euler characteristic). The dynamics of the geodesic flow on T^1S are *hyperbolic:* orbits diverge from each other exponentially fast in either forward or backward time.

We next present without proof some properties of these metrics, to give a feel for the landscape. Because these metrics are highly symmetric, they also admit algebraic descriptions. While there will be little discussion or use of Lie groups in the chapters that follow, we include here the algebraic constructions of the surfaces in question and of their geodesic flows. In the case of surfaces of constant negative curvature, this algebraic description will have further use in the following chapters. We recommend [124] for an introduction to Lie groups and homogeneous spaces in the context of dynamical systems; a quick review of the concepts needed here may be found in §1.5.2 of the Notes at the end of this chapter.

Finally, we discuss some properties of other metrics, of variable curvature, on these surfaces, both to contrast with the constant curvature case and to explore some of the richness of metric structures on surfaces.

The Round Case: Constant Curvature 1. The sphere

$$S^2=\{(x,y,z)\in\mathbb{R}^3 : x^2+y^2+z^2=1\},$$

endowed with the induced Euclidean metric from $\mathbb{R}^3$, has constant curvature 1, and this is the unique metric on the sphere of constant curvature 1. Geodesics in this *round metric* are great circles, obtained by intersecting the sphere with planes through the origin.

The (nonabelian) Lie group

$$\mathrm{SO}(3)=\{O\in M_{3\times 3}(\mathbb{R}) : O^tO=I,\ \det(O)=1\}$$

is the group of rotations of $\mathbb{R}^3$. Rotations preserve Euclidean length and hence preserve the sphere $S^2\subset\mathbb{R}^3$; moreover, the derivative of a rotation preserves the unit tangent bundle T^1S^2. Any vector in T^1S^2 can be taken to any other vector by the derivative of a rotation, and a rotation whose derivative fixes all unit tangent vectors is trivial. Thus the derivative action of the group of rotations on the unit tangent bundle T^1S^2 is transitive and free (see §1.5.2 in the Notes section for definitions), and so the unit tangent bundle to the sphere is diffeomorphic to SO(3). This identification between T^1S^2 and SO(3) is realized concretely by mapping the unit tangent vector $v\in T_xS^2$ to the matrix whose columns are x, v, and $x\times v$.

[6]There is a nice parallel in this terminology. In the terminology of partial differential equations, the associated PDE to the local embedding problem is elliptic in positive curvature and hyperbolic in negative curvature (and of mixed type otherwise).

The group SO(3) is diffeomorphic to T^1S^2 and is isomorphic to $\mathrm{PU}(2) = \mathrm{U}(2)/S^1$, where

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} \alpha & -\bar\beta \\ \beta & \bar\alpha \end{pmatrix} : \alpha, \beta \in \mathbb{C};\ |\alpha|^2 + |\beta|^2 = 1 \right\}.$$

One can see this by identifying $S^2 = \mathbb{C}P^1$ with $\mathbb{C} \cup \{\infty\}$ via stereographic projection. The round metric is transported to the metric

$$ds^2 = \left(\frac{2}{1+|z|^2} \right)^2 |dz|^2$$

on $\mathbb{C}$, and the action of SO(3) is transported to the action of PU(2) via Möbius transformations:

$$\begin{pmatrix} \alpha & -\bar\beta \\ \beta & \bar\alpha \end{pmatrix} : z \longmapsto \frac{\alpha z - \bar\beta}{\beta z + \bar\alpha}.$$

Note that SU(2) is the 3-sphere, and so $T^1S^2 = \mathrm{PU}(2) \cong \mathrm{SU}(2)/\{\pm I\}$ is double-covered by S^3.

Clearly, every orbit of the geodesic flow is closed and projects under $\pi : T^1S^2 \to S^2$ to a great circle. The geodesic flow is thus a free action of S^1 on T^1S^2; it is in fact the action on PU(2) of the 1-parameter subgroup

$$\mathscr{G}_{\mathbb{S}^2} = \left\{ \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} : t \in \mathbb{R} \right\}$$

by multiplication on the right. The orbits of this flow, when lifted to SU(2), form a foliation of the 3-sphere known as the Hopf fibration.[7] Notably, any two orbits of this flow are linked, and orbits arrange themselves on invariant tori surrounding any closed orbit. See Figure 1.5.

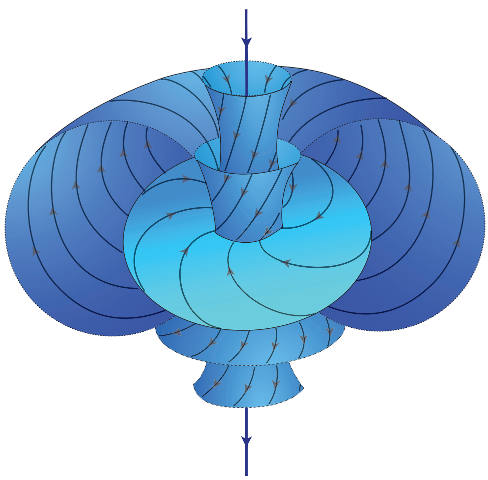


Figure 1.5. The Hopf fibration of S^3, which is depicted here as the 1-point compactification of $\mathbb{R}^3$ (under stereographic projection). The fibers are circles, geodesics in S^3. Circles are arranged as $(1,1)$ knots on tori around two central circles, the vertical one (through the point at infinity) and a meridional one (not pictured, but lying inside the light colored solid torus). In the projection of S^3 to T^1S^2 these two circles are identified.

[7]The Hopf fibration is named for Heinz Hopf. To avoid confusion, we will sometimes use first initials to distinguish between Eberhard and Heinz Hopf.

Now let's consider what happens to the geodesic flow when the metric is changed slightly, perturbed so that the curvature remains positive but no longer constant. The theory of Kolmogorov–Arnol'd–Moser (KAM theory, for short) implies that there are metrics arbitrarily close to the round metric whose geodesic flows have many (in the sense of volume) persistently quasiregular orbits, confined to invariant tori in T^1S^2, as in Figure 1.6. Experimentally, the flows for these metrics display a classic phenomenon: very complicated orbit structures, in which the stable invariant tori of orbits (from KAM theory) coexist with chaotically wandering orbits between these tori.[8]

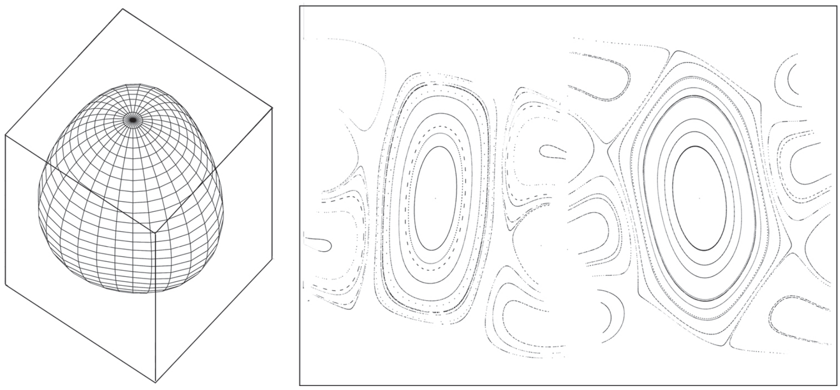


Figure 1.6. Orbits of the geodesic flow in positive curvature, seen at right in cross-section in T^1S^2. Images courtesy of Daniel Müller (see [264] for details).

When the metric is further perturbed beyond positive curvature, new possibilities for the dynamics of the geodesic flow come into relief. For example, there is a metric on the sphere (depicted in Figure 1.7) whose geodesic flow φ_t is *ergodic:* with respect to the Liouville measure, almost every φ_t-orbit is equidistributed in T^1S^2 [118]; formally, for every integrable function $\psi\colon\ T^1S^2 \to \mathbb{R}$ and almost every $v \in T^1S^2$ with respect to the Liouville probability measure m_{Liou} for this metric,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \psi(\varphi_t(v))\,dt = \int_{T^1S^2} \psi\,dm_{\mathrm{Liou}}.$$

Ergodicity of φ_t implies in particular that almost every geodesic in this metric is dense in S^2 (and even in T^1S^2!), in dramatic contrast to the round metric. It is an open question whether a geodesic flow in positive curvature can be ergodic; it is even possible that there is a metric in a neighborhood of the round metric whose geodesic flow is ergodic. The definition of ergodicity can be found in Chapter 5, §5.1.4.

While Hadamard found success characterizing very general properties of negatively curved manifolds, the subject of geodesic flows on the sphere remained relatively unexplored until Poincaré [285] commenced a study of closed geodesics in positively curved metrics. Poincaré's interest in spherical metrics derived from his work in understanding phenomena in celestial mechanics, and he saw important parallels between the two:

[8]Other rigorous mechanisms—such as Chirikov's resonance/overlap criterion, Arnol'd diffusion, Nekhoroshev long-time stability, and Aubry–Mather sets—can explain some of this observed coexistence, but the full phenomenon (of positive volume sets of "chaotic seas" between KAM tori) has yet to be explained theoretically.

> J'ai donc abordé l'étude des lignes géodésiques des surfaces convexes ; malheureusement le problème est beaucoup plus difficile que celui qui a été résolu par M. Hadamard. J'ai donc dû me borner à quelques résultats partiels, relatifs surtout aux géodésiques fermées qui jouent ici le rôle des solutions périodiques du problème des trois corps.[9] [285, p. 237]

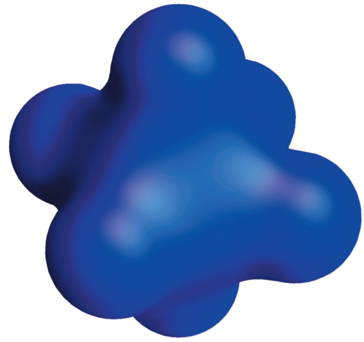


Figure 1.7. A sphere with ergodic geodesic flow. See Burns and Donnay [73] for mathematical details.

Poincaré's intuition was borne out in the following years, in his own work with Birkhoff, as well as the later work by Kolmogorov, Arnol'd, and Moser mentioned above. These developments are both analytic and topological and far too numerous and deep to describe here; restricting attention to the sphere itself, one notable result from the 1990s states that any Riemannian metric on S^2 has infinitely many closed geodesics, corresponding to infinitely many distinct periodic orbits for the geodesic flow[10] [33, 139].

In contrast to the richness of metric structures, there are relatively few rigidity phenomena for metrics on the sphere.[11] On the one hand, there is the following topological type of rigidity that follows immediately from the Gauss–Bonnet formula (Theorem 2.6.9):

> any closed surface with everywhere positive curvature is the sphere.

The higher-dimensional analogue to this fact is the famous *sphere theorem*: any closed, orientable manifold carrying a metric whose curvature is positive and sufficiently close to constant (a condition known as "strictly 1/4-pinched positive curvature") is finitely covered by a sphere.

On the other hand, if we restrict to properties of metrics on the sphere itself, some of the most obvious rigidity conjectures fail to hold true. For example, the round metric has the special property that all geodesics are closed. One might ask whether this property is a defining one: Do there exist metrics on the sphere, not of constant curvature, in which all geodesics are closed? For such a metric, it is not hard to show that all geodesics must have the same length.

It might appear that this property is too special to exist outside of constant curvature, but in 1903, Otto Zoll [370] discovered an infinite-dimensional family of real analytic metrics with this property! To carry out a type of planar counterpart to this construction is an interesting exercise: there exist closed curves of constant

[9] "I have therefore begun to study the geodesic lines of convex surfaces; unfortunately, the problem is much more difficult than that which was solved by Mr. Hadamard. I must therefore confine myself to a few partial results, especially relating to closed geodesics, which here play the role of the periodic solutions to the problem of the three bodies."

[10] This result holds on all surfaces in fact, but due to simple connectivity, there are no simple topological arguments to produce these orbits on the sphere, which makes the result more interesting and much more difficult to prove in genus 0 than in higher genus. Morse's early efforts to find closed geodesics on the sphere led him to develop Morse theory.

[11] For some of them, see "Eleven properties of the sphere" in the classic Hilbert–Cohn-Vossen text *Geometry and the imagination* [186, §32]. Here's a sample: As an embedded submanifold of $\mathbb{R}^3$, the round unit sphere is the unique surface whose intersection with every plane is a circle of radius $\leqslant 1$.

width that are not the circle.[12] See Figure 1.8. We will discuss Zoll metrics further in Chapter 11.

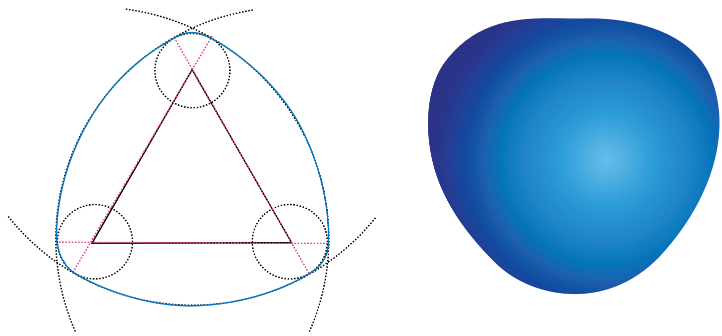


Figure 1.8. A C^1 closed curve of constant width (left) and a Zoll surface (right). The curve of constant width can roll freely between two tangent parallel lines. Every geodesic on the Zoll surface is closed, and all have the same length.

There remain interesting open rigidity questions for metrics on spheres, but we will not consider them here. Our focus will be on rigidity phenomena associated to metrics of a very different nature, those without conjugate points, a phenomenon that can occur in surfaces only of genus at least 1.

The Flat Case: Constant Curvature 0. While the sphere carries a unique Riemannian metric of constant curvature 1, there are infinitely many nonisometric *flat metrics*—that is, metrics of constant curvature 0—on the 2-dimensional torus.[13] Indeed, any lattice $\Lambda < \mathbb{R}^2$ gives rise to a flat torus $\mathbb{R}^2/\Lambda$ by projecting the Euclidean metric, which is preserved by translations. One can parametrize the space of such metrics with fixed area 1 using matrices in $\mathrm{SL}(2,\mathbb{R})$: the columns of such a matrix generate a lattice Λ with the area of $\mathbb{R}^2/\Lambda$ equal to 1.

There are two commuting operations on a matrix in $\mathrm{SL}(2,\mathbb{R})$ that produce lattices with isometric surfaces: multiplying on the left by an element of $\mathrm{SO}(2)$ and multiplying on the right by an element of $\mathrm{SL}(2,\mathbb{Z})$. The left multiplication rotates the lattice in the plane, and the right multiplication preserves the lattice, merely changing the two generators that form the columns of the matrix. The space of all such flat metrics of total area 1 on the torus is thus equal to the classical *moduli space* $\mathrm{SO}(2)\backslash\mathrm{SL}(2,\mathbb{R})/\mathrm{SL}(2,\mathbb{Z})$, which is a 2-dimensional orbifold, a punctured sphere with two cone points.

Since any lattice Λ is obtained by applying an element of $\mathrm{SL}(2,\mathbb{R})$ to the standard lattice $\mathbb{Z}^2$, and linear transformations send lines to lines, one can determine properties of geodesics in $\mathbb{R}^2/\Lambda$ from those in $\mathbb{R}^2/\mathbb{Z}^2$ by inverting a linear transformation.[14] Let us then fix the flat metric on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and describe its geodesic flow. Note that the unit tangent bundle $T^1\mathbb{T}^2$ is trivial, canonically isomorphic to $\mathbb{T}^2 \times S^1$.

Geodesics in Euclidean space are lines, and the geodesic flow on the Euclidean plane flows at unit speed along lines. The orbits of the geodesic flow for the torus

[12] The connection between Zoll surfaces and surfaces of constant width is just an analogy: whether there exists an embedded Zoll surface in $\mathbb{R}^3$ of constant width is actually an open question.

[13] Or, in the language of more than one 20th-century mathematician, on the *anchor ring* [52, 49].

[14] This linear transformation changes the lengths of all geodesics, so not all properties of the geodesic flow can be deduced directly from studying $\mathbb{R}^2/\mathbb{Z}^2$. See also Exercise 1.4.

are simply projected orbits of the flow for Euclidean space. Hence for $(x, v) \in T^1\mathbb{T}^2 = \mathbb{T}^2 \times S^1$,

$$\varphi_t(x, v) = (x + tv, v),$$

where the first coordinate is understood modulo $\mathbb{Z}^2$. See Figure 1.9.

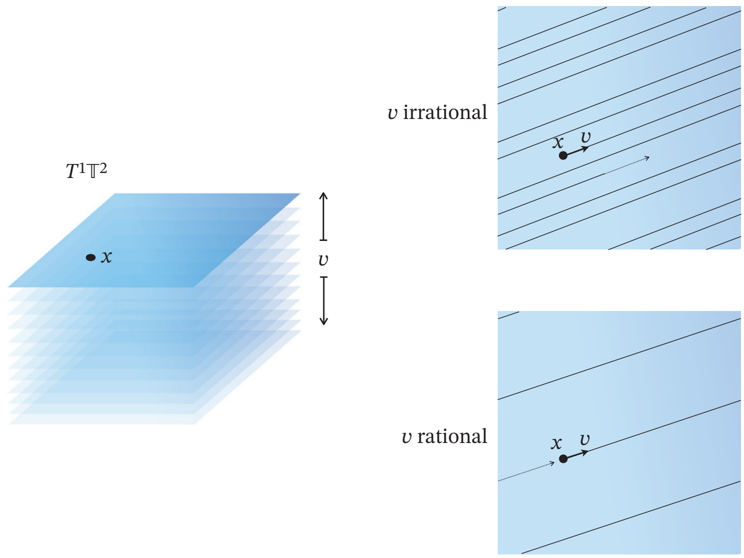


Figure 1.9. The geodesic flow on $T^1\mathbb{T}^2$: orbits are tangent to the tori $v \equiv$ constant and are either closed or wind densely in the 2-dimensional torus to which they are confined, depending on the rationality of the unit vector v.

A few observations about this flow:

- The orbits are confined to the leaves of the foliation

$$\{\mathbb{T}^2 \times \{v\} : v \in S^1\}$$

of $T^1\mathbb{T}^2$ by tori. Thus the flow φ_t is not ergodic: the orbits are not equidistributed, or even dense, in $T^1\mathbb{T}^2$.
- The dynamics of the restriction of the flow to an invariant leaf $\mathbb{T}^2 \times \{v\}$ depends dramatically on v. If $v = (p/\sqrt{p^2+q^2}, q/\sqrt{p^2+q^2})$, with $p, q \in \mathbb{Z}$, then all of the orbits on $\mathbb{T}^2 \times \{v\}$ are circles (closed (p, q)-knots, to be precise) in $\mathbb{T}^2 \times \{v\}$. Otherwise, v represents an *irrational* direction, and the orbits of φ_t wind densely about $\mathbb{T}^2 \times \{v\}$.

Thus, focusing on a single leaf $\mathbb{T}^2 \times \{v\}$ of this invariant foliation, the geodesic flow is the translation flow $\varphi_t : x \mapsto x + tv$ on the torus: that is, a flow along the 1-parameter subgroup $\mathbb{R}v$. The discrete time version of this flow—i.e., the time-t_0 map—is the translation $x \mapsto x + t_0 v$ by a fixed element $t_0 v$ in this subgroup.

Translations on the torus (in arbitrary dimension) have been studied as dynamical systems for more than 150 years. The assertion that every orbit of an irrational translation is dense was first proved by Kronecker in 1884 [221], who applied it to the problem of Diophantine approximation of real numbers by rationals (see Exercise 4.1). A study of the ergodic-theoretic properties of toral translations followed within 20 years in the work of Hermann Weyl, Wacław Sierpiński, and Piers Bohl.

Under the assumption that v is an irrational direction, Weyl et al. showed that for any point x and any continuous function ψ on the torus $\mathbb{T}^d$, the sequence $\{\psi(x+nv)\}_{n\in\mathbb{N}}$ satisfies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\psi(x+jv)=\int_{\mathbb{T}^d}\psi.$$

This is a property of the translation $x\mapsto x+v$ even stronger than ergodicity, which we now call *unique ergodicity*. This property is equivalent to the fact that Lebesgue measure is the unique invariant probability measure for an irrational translation, hence the name.

While irrational toral translations are ergodic, they are highly predictable from a statistical point of view: for any continuous ψ and any x, the values $\{\psi_n(x)\}_{n\leqslant 0}$ in the past completely determine the values $\{\psi_n(x)\}_{n\geqslant 0}$ in the future, a property of the translation known as *zero entropy*, a dynamical feature shared with the constant curvature geodesic flows on $T^1\mathbb{T}^2$ and T^1S^2. By contrast, the geodesic flows in constant negative curvature have positive entropy; the consequences of this will be explored in Chapter 10.

Although it is not commonly viewed this way, the geodesic flow for the torus can also be described algebraically, in a similar manner to the geodesic flow for the sphere. We describe this in detail in the Notes section at the end of this chapter.

By the Gauss–Bonnet theorem (Theorem 2.6.9), any metric on $\mathbb{T}^2$ of nonpositive curvature or of nonnegative curvature is flat. In particular, if the flat metric is perturbed so that it is no longer flat, regions of both positive and negative curvature must appear. If one continues to perturb the metric, one can produce localized sphere-like geometry by adding a bulb-like growth to the torus, as in Figure 1.10. In particular, on the torus (or indeed, on any surface), there is no obstruction to having a metric where, in the lift to the universal cover $\mathbb{R}^2$, two points are connected by more than one geodesic path. By contrast, the flat metric has the property that any two points in $\mathbb{R}^2$ are connected by a unique geodesic.

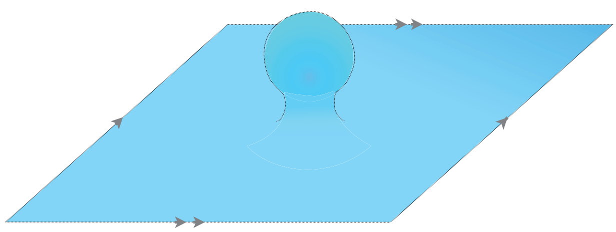


Figure 1.10. A spherical bulb growing out of a flat torus.

Eberhard Hopf [192] proved that the *only* metric on $\mathbb{T}^2$ with this "unique geodesic property" (known as *no conjugate points*) is the flat metric. In particular, this means that if one perturbs the flat metric even only slightly, the regions of positive curvature will conspire to bend geodesics together, forcing some of them to cross in more than one point.

Hopf's proof is perhaps the first explicit use of the geodesic flow to prove a geometric rigidity theorem. We present it in Chapter 2, §2.5.7. In Chapter 9, we will present a highly nontrivial generalization of this theorem to higher dimensions, due to Dmitri Burago and Sergei Ivanov [72].

The Hyperbolic Case: Constant Curvature –1. The early-19th-century discovery of the hyperbolic plane was an event in axiomatic geometry that spurred profound developments in mathematics as a whole. Unlike the round sphere, in which all geodesics intersect, or the Euclidean plane, in which nonintersecting lines lie at a constant distance from each other, the hyperbolic plane has nonintersecting lines that move arbitrarily far from each other: indeed, the distance between two unit-speed hyperbolic geodesics will always diverge exponentially fast in at least one direction of time. This instability leads to a great deal of complexity in the collection of all geodesics on a compact hyperbolic surface and in its geodesic flow.

The hyperbolic plane is the unique simply connected surface of constant negative curvature -1. There are two conformally equivalent models of the hyperbolic plane that are often used: the Poincaré disk $\mathbb{D}$ and the upper-half-plane model $\mathbb{H}$.

The *Poincaré disk* (or hyperbolic disk) is the domain $\mathbb{D} := \{z : |z| < 1\}$ with the metric

$$ds^2 = \frac{4|dz|^2}{(1-|z|^2)^2}.$$

The group of orientation-preserving isometries of $\mathbb{D}$ is

$$\mathrm{PSU}(1,1) = \left\{\left(\begin{smallmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{smallmatrix}\right) : \alpha, \beta \in \mathbb{C};\ |\alpha|^2 - |\beta|^2 = 1\right\}/\{\pm I\},$$

which acts by Möbius transformations:

$$\begin{pmatrix}\alpha & \beta\\ \bar\beta & \bar\alpha\end{pmatrix} : z \longmapsto \frac{\alpha z + \beta}{\bar\beta z + \bar\alpha}.$$

The hyperbolic disk is isometric via Möbius transformation to the upper-half-plane $\mathbb{H} := \{z : \Im(z) > 0\}$ with the metric

$$ds^2 = \frac{|dz|^2}{(\Im(z))^2}.$$

See Exercise 1.5. The isometry group of $\mathbb{H}$ is

$$\mathrm{PSL}(2,\mathbb{R}) = \left\{\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) : a, b, c, d \in \mathbb{R};\ ad - bc = 1\right\}/\{\pm I\},$$

also acting by Möbius transformations. The curvature of $\mathbb{H}$ is constant, equal to -1. We will refer to the $\mathbb{D}$ and $\mathbb{H}$ models (depicted in Figure 1.11) interchangeably. We will also refer to these models as $\mathbb{D}^2$ and $\mathbb{H}^2$, when the dimension becomes relevant.

Hyperbolic geodesics in $\mathbb{D}$ are Euclidean circular arcs, perpendicular to $\partial\mathbb{D} = \{z : |z| = 1\}$. In $\mathbb{H}$, hyperbolic geodesics are Euclidean (semi-) circular arcs, perpendicular to $\partial\mathbb{H} = \{z : \Im(z) = 0\}$, including the "degenerate circles" consisting of vertical rays perpendicular to $\partial\mathbb{H}$. To find all geodesics on $\mathbb{H}$, it suffices to find just one geodesic—for example the vertical ray in $\mathbb{H}$ tangent to the vector $\partial/\partial y \in T_i^1\mathbb{H}$—and then note that every other geodesic ray is the image of this geodesic by an element of $\mathrm{PSL}(2,\mathbb{R})$.

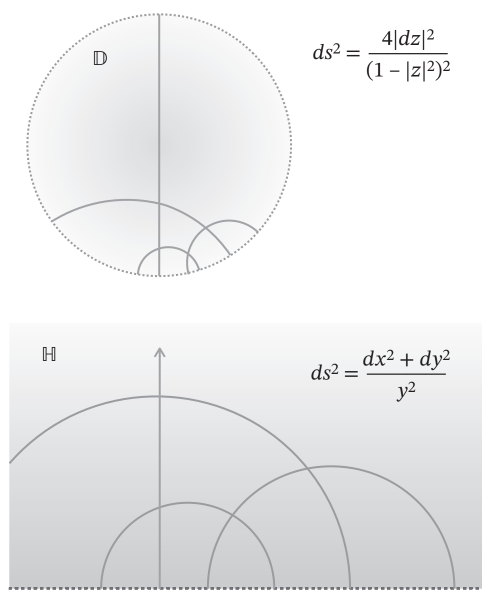


Figure 1.11. The disk (top) and upper-half-plane models of the hyperbolic plane, depicted with geodesics.

The stabilizer of a point under the action by $\mathrm{PSL}(2,\mathbb{R})$ is the compact subgroup $K = \mathrm{SO}(2)/\{\pm I\}$, which gives an identification of $\mathbb{H}$ with the coset space of K:

$$\mathbb{H} = \mathrm{PSL}(2,\mathbb{R})/K = \mathrm{SL}(2,\mathbb{R})/\,\mathrm{SO}(2).$$

The derivative action of $\mathrm{PSL}(2,\mathbb{R})$ on the unit tangent bundle $T^1\mathbb{H}$ is free and transitive; therefore, if we fix a vector $v \in T^1\mathbb{H}$ and map $g \in \mathrm{PSL}(2,\mathbb{R})$ to the image $g\cdot v$ under this action, we obtain an analytic diffeomorphism between $T^1\mathbb{H}$ and $\mathrm{PSL}(2,\mathbb{R})$. Under this identification the derivative action of $\mathrm{PSL}(2,\mathbb{R})$ on $T^1\mathbb{H}$ corresponds to the action on $\mathrm{PSL}(2,\mathbb{R})$ by left multiplication. We discuss this identification in more detail in Chapter 2, §2.4.1.

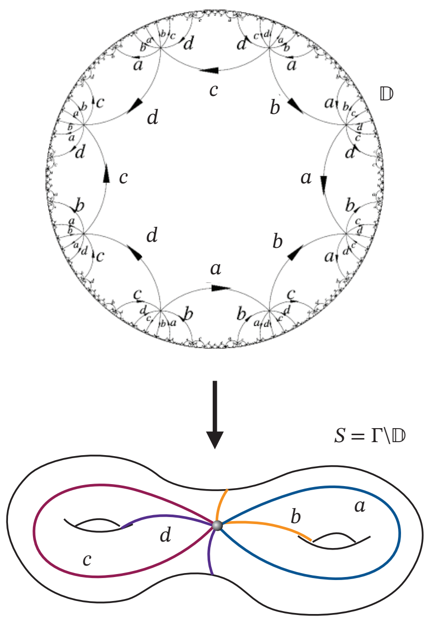


Figure 1.12. Constructing a hyperbolic surface of genus 2 from a regular hyperbolic octagon.

If S is a closed orientable surface with $\tilde{S} = \mathbb{H}$, then $\pi_1(S)$ acts by isometries on $\mathbb{H}$ and hence embeds as a discrete subgroup $\Gamma < \mathrm{PSL}(2,\mathbb{R})$. We thus obtain the following identifications:

$$S = \Gamma\backslash\mathbb{H} = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})/K$$

and

$$T^1S = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R}).$$

Any closed orientable surface of genus 2 or higher admits a metric of constant negative curvature. As with the flat torus $\mathbb{T}^2$, it is not possible to isometrically embed such a surface in $\mathbb{R}^3$. We can however rely on pictures of a fundamental domain in $\mathbb{H}$ or $\mathbb{D}$ with specified isometric gluing maps on the edges to picture such a surface.

Perhaps the simplest way to construct a genus g surface is to take a regular hyperbolic $4g$-gon with sum of vertex angles equal to 2π and use hyperbolic isometries to glue opposite sides, as in Figure 1.12. Much more generally, hyperbolic structures are constructed by gluing together $2g-2$ hyperbolic "pairs of pants" of varying cuff lengths. There are $6g-6$ degrees of freedom in this construction (cuff lengths of pants and twist parameters in gluing). The space of all such structures is a $(6g-6)$-dimensional space, diffeomorphic to a ball, called *Teichmüller space*, and the space of such structures modulo isometry is a $(6g-6)$-dimensional orbifold called *moduli space*, which is the hyperbolic version of the moduli space $\mathrm{SL}(2,\mathbb{Z})\backslash\,\mathrm{SL}(2,\mathbb{R})/\,\mathrm{SO}(2)$ of flat structures on the torus. We discuss the "pair of pants" construction in §1.5.4 of the Notes section; we recommend [130] for an introduction to moduli spaces of hyperbolic structures.

The action of the fundamental group $\pi_1(S)$ on the universal cover $\mathbb{H}$ by isometric deck transformations defines a discrete and faithful[15] representation $\rho:\ \pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$. Conversely, every such discrete and faithful representation defines a hyperbolic surface by

$$S = \Gamma\backslash\,\mathrm{PSL}(2,\mathbb{R})/K,$$

[15] Discrete means that the image is discrete in the compact-open topology on $\mathrm{PSL}(2,\mathbb{R})$; faithful is another way of saying injective.

with $\Gamma = \rho(\pi_1(S))$ and $K = \mathrm{SO}(2)$. We won't discuss this topic any further, but we remark that using algebraic methods, one can find special types of *arithmetic* representations, which give rise to *arithmetic surfaces* that have special rigidity properties of their own.

As with the round sphere and flat tori, the dynamics of the geodesic flow on a hyperbolic surface are described algebraically by translation in a 1-parameter subgroup. In the upper-half-plane model, the geodesic flow for $\Gamma\backslash \mathrm{PSL}(2,\mathbb{R})/K$ is given by the action of the 1-parameter subgroup

$$\mathscr{G}_{\mathbb{H}^2} := \left\{ \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} : t \in \mathbb{R} \right\}$$

on $\Gamma\backslash \mathrm{PSL}(2,\mathbb{R})$ under *right* multiplication. We will describe this action in greater detail in the next chapter.

The hyperbolic geodesic flow has a rich structure that we will discuss in depth in this text. In particular, for a compact hyperbolic surface S, closed orbits are dense in T^1S, a fact established by a series of authors, beginning with Morse in 1920 [260, 261, 18, 269, 217, 240], and the geodesic flow is ergodic, which was proved by Gustav Hedlund [181] in 1934, building on the earlier work of P. J. Myrberg [265].

Even earlier, properties of metrics of *nonconstant* negative curvature had begun to be understood. Hadamard [176] studied infinite geodesics in negative curvature and introduced the notion of asymptotic geodesics. His main insight was to organize the converging geodesics in negative curvature so that global properties of geodesics could be studied in a systematic way on a local level.

Building on Hadamard's insights, Hopf showed that the geodesic flow for *any* negatively curved closed surface is ergodic. The key ingredient in Hopf's proof—known as the *Hopf argument*—is one of the cornerstone arguments in smooth ergodic theory and has seen applications far beyond the context of geodesic flows: to Hamiltonian systems such as ideal gases, to differential equations modeling weather and population growth, and to homogeneous dynamical systems dictating number-theoretic phenomena. We discuss the Hopf argument in detail in Chapter 5, §5.2.

The works of Morse and Hadamard contain precursors of a central notion in dynamics called *structural stability.*[16] In this context, what structural stability means is that the geodesic flows for any two negatively curved metrics on a fixed surface "are topologically the same." Formally, the geodesic flows for the two metrics are *orbit equivalent:* there is a homeomorphism between the unit tangent bundles sending orbits of the first flow to orbits of the second.[17] Thus, in some crude sense, the dynamics of the geodesic flow in negative curvature are quite rigid: the orbits are forced into a fixed topological mold.

On the other hand, there is considerable flexibility in prescribing a negatively curved metric on a surface: for example, any smooth negative function on a closed Riemannian surface Σ of genus at least 2 is the curvature of a unique metric on Σ with the same conformal structure [46] (we discuss conformal metrics in Chapter 2, §2.6.1). The rigidity in the topological structure of geodesics combined with the flexibility in

[16] The deep connections between the work of Morse and dynamical systems should come as no surprise: Morse was a student of George Birkhoff.

[17] This homeomorphism will *not* in general preserve the parametrization by time. A time-preserving orbit equivalence is called a *conjugacy*. The existence of a conjugacy between geodesic flows on negatively curved surfaces is closely related to the existence of an isometry between the surfaces. We discuss this in detail in Chapter 4.

prescribing curvature gives a ripe setting for various rigidity questions in negative curvature. We explore two of them in this book: marked length spectrum rigidity and entropy rigidity.

On the boundary of negative curvature is nonpositive curvature. Some properties of negatively curved metrics extend readily to nonpositively curved metrics; for example, a nonpositively curved metric has no conjugate points (see Theorem 2.5.4), and there exist dense orbits for the geodesic flow for any nonpositively curved surface of genus $\geqslant 2$. Yet there remain fundamental open questions, among them, Is the geodesic flow ergodic for a nonpositively curved surface of genus $\geqslant 2$?

1.2. Three Rigidity Results

We have chosen to highlight three important theorems in the rigidity theory of Riemannian manifolds whose proofs make substantial use of the dynamics of geodesic flows.

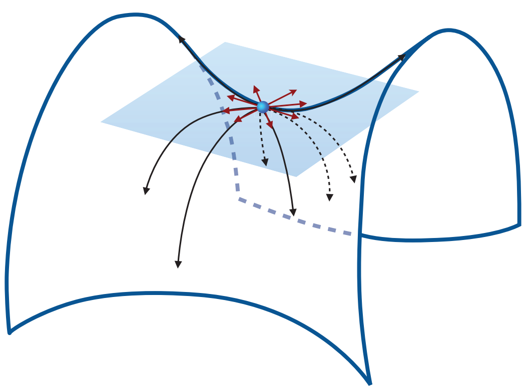


Figure 1.13. Flowing out a plane along geodesic rays (via the exponential map) to obtain a negatively curved surface.

We say that a Riemannian manifold is *negatively curved* if the sectional curvatures at every point are negative and *nonpositively curved* if the sectional curvatures are nonpositive. In the next chapter we define sectional curvature; informally, a manifold M is negatively curved if, for every $p \in M$, the surface obtained by flowing out an arbitrary plane in the tangent space T_pM along short geodesic rays (as in Figure 1.13) is negatively curved in the induced metric from M. Similarly, M is nonpositively curved if the resulting surface is negatively curved or flat at p, for every point p and every plane in T_pM.

In what follows, M will always denote a complete Riemannian manifold, and $\widetilde{M}$ will denote the universal cover of such a manifold. We will say that M is *closed* if it is compact, boundaryless, connected, and orientable.

1.2.1. Marked Length Spectrum. Let M be a closed, negatively curved manifold. A *free loop* in M is a continuous map $c : S^1 \to M$, up to reparametrization. The space of loops in M modulo homotopy is isomorphic to the set of conjugacy classes in $\pi_1(M)$, the fundamental group of M based at some (arbitrary) point $x_0 \in M$. Denote by $\mathscr{C}(M)$ this space of free loops modulo homotopy.

A (unit-speed) *closed geodesic* on a Riemannian manifold is a unit-speed curve $c : S^1 \to M$ that locally minimizes the Riemannian distance on M among all unit-speed curves. On any closed Riemannian manifold M there is a closed geodesic loop representing each free homotopy class of closed curves in M (see, e.g., [50, §11.7]). If M is negatively curved, then this geodesic representative is unique (see Corollary 3.3.5).

Given a closed, negatively curved M, we thus have a map $\ell = \ell_M : \mathscr{C}(M) \to \mathbb{R}_+$ called the *marked length spectrum for M* which assigns to $[c] \in \mathscr{C}(M)$ the length[18] of its

[18]See Chapter 2, §2.1.1 for the definitions of geodesics and length.

geodesic representative c. Observe that a homeomorphism $h: M \to M'$ induces a map $h_*: \mathscr{C}(M) \to \mathscr{C}(M')$. We say that M and M' have *the same marked length spectrum* if there exists a homeomorphism $h: M \to M'$ such that $\ell_{M'} \circ h_* = \ell_M$.

Burns and Katok asked whether the function ℓ_M determines M, up to isometry [75]. This is easily seen to be true if one restricts to hyperbolic surfaces: starting with an abstract pair of pants decomposition, one can find finitely many curves that determine both the twist and cuff parameters (see Exercise 1.6). In general, this question remains open but has been solved completely for surfaces by Otal [272] and independently slightly later, but in greater generality, by Croke [96].

THEOREM A (Otal, Croke, 1990 [272, 96]). *If two negatively curved surfaces S and S' have the same marked length spectrum then S and S' are isometric.*

Croke's generalization allows for one of S or S' to be nonpositively curved. The problem remains open in higher dimensions, but there are partial results by Hamenstädt [178].

We remark that the result fails without the marking given by the map ℓ_M; that is, the *values* of ℓ_M alone do not determine M. If we define the (unmarked) length spectrum of M to be the set of lengths $\{\ell_M([c]) : [c] \in \mathscr{C}(M)\}$, counted with multiplicity, then there exist nonisometric *hyperbolic* surfaces with the same length spectrum [345, 337].

We prove Theorem A in Chapter 8. The proof that we present is a hybrid of the techniques from various sources and exploits in full force the hyperbolicity of the geodesic flow in negative curvature. The Liouville probability measure m_{Liou} plays a key role in the proof, as does the fact that the geodesic flow is ergodic for negatively curved surfaces (Theorem 5.2.12).

1.2.2. Tori Without Conjugate Points.

Let M be a complete Riemannian manifold. Loosely speaking, conjugate points in M are pairs of points p and q such that initially divergent geodesic rays at p converge at q, on an infinitesimal level. While we defer the precise definition of conjugate points to the discussion of Jacobi fields in Chapter 2 (see Definition 2.5.3), an equivalent definition of "*no* conjugate points" can be made without reference to Jacobi fields.

DEFINITION 1.2.1. A Riemannian manifold M *has no conjugate points* if for every pair of points $p, q \in \widetilde{M}$ in the universal cover of M, there exists a unique geodesic in $\widetilde{M}$ from p to q, in the lifted Riemannian metric.

As discussed in the previous section, the round sphere has conjugate points (as depicted in Figure 1.14), whereas the flat torus does not, and one can easily construct a metric on the torus (or on any surface) with conjugate points by gluing in a sphere-like bulb.

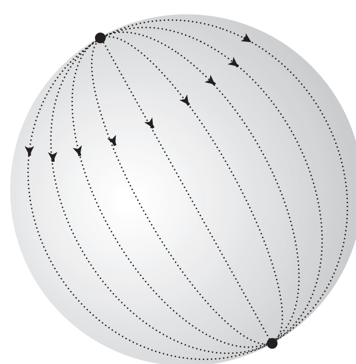


Figure 1.14. Conjugate points on the sphere.

More generally, no conjugate points on M implies that $\widetilde{M}$ is contractible (see Theorem 2.5.17), and so every metric on a sphere, in any dimension, has conjugate points. There is a close relationship between having no conjugate points and nonpositive curvature: by Theorem 2.5.4, nonpositive curvature implies no conjugate points. E. Hopf proved that the reverse implication holds for the 2-dimensional torus:

THEOREM B (E. Hopf [192]). *A metric on $\mathbb{T}^2$ without conjugate points is flat.*

We will prove this in Chapter 2. We remark that nothing of this form can hold on higher genus surfaces, as there exist metrics without conjugate points that also have regions of positive curvature: one can achieve this by adding a small "pimple" to a hyperbolic surface. Hopf's theorem left open the question on higher-dimensional tori, which came to be known as the "Hopf conjecture." The question was finally settled in the 1990s by Burago and Ivanov:

THEOREM C (Burago–Ivanov, 1994 [72]). *A metric on $\mathbb{T}^n$ without conjugate points is flat.*

The proof, which we present in Chapter 9, employs a lovely combination of large-scale geometry and convex geometric analysis and, like the proof of Theorems A and B, exploits in an essential way the fact that the geodesic flow preserves the Liouville probability measure m_{Liou} on $T^1\mathbb{T}^n$.

1.2.3. Entropy Rigidity. Let S be a closed, negatively curved surface of genus $g \geqslant 2$. A natural invariant of S is the exponential area growth of balls in the universal cover $\tilde{S}$. For $p \in \tilde{S}$ and $R > 0$, denote by $B(p,R)$ the geodesic ball of radius R in the lifted Riemannian metric. The following limit exists and is independent of p:

$$\lim_{R\to\infty} \frac{1}{R} \log\big(\mathrm{area}(B(p,R))\big),$$

where the area also is measured in the lifted Riemannian metric. We will see that the negative curvature of S implies that the limit is positive. Clearly this area growth rate is an isometry invariant of S.

Again due to the negative curvature of S, this quantity has two other equivalent definitions: as the growth in L of the number of closed geodesics c in S of length $\ell(c) \leqslant L$, and as a dynamical invariant of the geodesic flow on T^1S, known as the *topological entropy* $h_{\mathrm{top}}(S)$ [247]. The word "topological" has the potential to be confusing here: the entropy $h_{\mathrm{top}}(S)$ is *not* a topological invariant of S itself, but it is a topological invariant of the flow φ_t on T^1S, in the sense that if there is a topological conjugacy[19] between the geodesic flows of two negatively curved surfaces S and S', then $h_{\mathrm{top}}(S) = h_{\mathrm{top}}(S')$.

There is another, related, growth invariant that one can attach to S: the *Liouville entropy* $h_{\mathrm{Liou}}(S)$. Let $\varphi_t : T^1S \to T^1S$ be the geodesic flow for S. Because S is closed and negatively curved, one can show the following limit exists and is independent of v, for almost every $v \in T^1S$ with respect to volume:

$$h_{\mathrm{Liou}}(S) := \lim_{t\to\infty} \frac{1}{t} \log \|D_v\varphi_t\|, \tag{1.2.1}$$

where the norm $\|D\varphi_t\|$ of the linear map $D\varphi_t$ on $T(T^1(S))$ is measured with respect to any fixed Riemannian metric on T^1S (and is therefore independent of the choice of metric on T^1S, once the metric on S is fixed). A fundamental dynamical relationship (which we will establish in Chapter 5, §5.3.4) that holds in great generality is the *variational principle for entropy*, which in this context implies that for any S,

$$h_{\mathrm{Liou}}(S) \leqslant h_{\mathrm{top}}(S). \tag{1.2.2}$$

If S is a *hyperbolic* surface, then these quantities can be easily computed, and they are equal:

$$h_{\mathrm{Liou}}(S) = h_{\mathrm{top}}(S) = 1.$$

[19] That is, a homeomorphism $h : T^1S \to T^1S'$ with $\varphi_t^{S'} \circ h = h \circ \varphi_t^S$, for all $t \in \mathbb{R}$.

Anatole Katok proved that equality of these entropies *characterizes* the hyperbolic surfaces:

THEOREM D (A. Katok 1982 [207]). *Let S be a closed, negatively curved surface. If the topological and Liouville entropies $h_{\mathrm{top}}(S)$ and $h_{\mathrm{Liou}}(S)$ of S coincide, then S has constant curvature.*

We prove Theorem D in Chapter 10. The higher-dimensional analogue of Theorem D remains open and is conjectured to hold, although there has been important progress on, and generalizations of, Theorem D [137, 43], which we also discuss. The basic phenomenon behind Theorem D, which is special to dimension 2, is that the quantity $h_{\mathrm{top}}(S)$ is minimized when S is hyperbolic, whereas $h_{\mathrm{Liou}}(S)$ is maximized for hyperbolic S.

The geodesic flows on the round sphere and flat tori have topological entropy (and hence, by (1.2.2), Liouville entropy) equal to 0. But metrics have been constructed on both the sphere and torus, arbitrarily close to the constant curvature metrics, with positive topological entropy [215]. It is still open whether a positively curved metric exists with positive Liouville entropy.

One might ask whether the property of having topological entropy 0 characterizes the constant curvature metrics on the sphere and torus. The answer is no: the Zoll surfaces, for example, have geodesic flows with entropy 0, as do the ellipsoids and the standard embedding of the torus in $\mathbb{R}^3$ as a surface of revolution.[20] On the other hand, it takes considerable work to construct metrics on these surfaces with zero entropy geodesic flows: the *generic* metric on the sphere has geodesic flow with positive entropy [216]!

1.3. Structure of the Book

Our main task is twofold: to explain the three rigidity results stated in the previous section and to give a flavor for the ideas and techniques behind them. We recognize the somewhat specialized nature of these results: the typical practitioner of dynamical systems has at most a passing familiarity with curvature and Jacobi fields, and the specialist in Riemannian geometry is not especially likely to know the statement of the ergodic theorem or the shadowing lemma. We therefore have included two background chapters to bring the reader up to speed. A benefit of this approach is that the text should also be accessible to the *nonspecialist* (in either dynamics or geometry) with a suitable background.

Chapter 2 is a crash course in Riemannian geometry and geodesic flows. Starting with the Levi-Civita connection on a Riemannian manifold M, we build the geodesic flow φ_t on TM and define various structures on TM that are invariant under this flow, in particular the symplectic form on TM, the contact form on T^1M, and the Liouville measure m_{Liou}.

[20] All of these are *integrable* geodesic flows, meaning that there is a smooth, flow-invariant function. One might ask whether on surfaces, zero entropy for the geodesic flow implies integrability of the flow. On T^2, $h_{\mathrm{top}} = 0$ imposes strong "integrable-like" features (a continuous first integral and strip-like behavior of lifted geodesics), but does not by itself yield smooth integrability [145]. For S^2 in the Riemannian category, as far as we know the question of whether $h_{\mathrm{top}} = 0$ forces Liouville integrability is open. (For genus $\geqslant 2$, one always has $h_{\mathrm{top}} > 0$ for any metric, so the question does not arise.)

We then turn to infinitesimal properties of this flow, that is, the properties of the derivative flow $D\varphi_t$ on the double tangent bundle TT^1M. To this end, we introduce the Sasaki metric on the double tangent bundle and define variations of geodesics and Jacobi fields. The derivative flow is then analyzed using the Jacobi and Riccati equations. We define conjugate points and explore the consequences of nonpositive curvature.

Finally, we focus on negatively curved manifolds and establish that the geodesic flow is *Anosov*, a global hyperbolicity condition on the derivative $D\varphi_t$. This property was initially established by Hadamard and was used by E. Hopf to prove ergodicity of geodesic flows for closed, negatively curved surfaces.

Next we turn in **Chapter 3** to the large-scale geometry of simply connected, nonpositively curved manifolds (and manifolds without conjugate points). After introducing some basic concepts from the theory of convex functions and sets, we describe convexity properties of the metric and define Busemann functions and horospheres in nonpositive curvature. Some boundary theory is developed, including an in-depth introduction to cross-ratios.

Chapter 4 is a crash course in hyperbolic dynamics. We begin with the basic notions of recurrence in dynamics, giving a criterion for existence of a dense orbit (topological transitivity). We then move to the central examples from hyperbolic dynamics, which can be viewed as toy models of the geodesic flow in negative curvature: expanding maps and Anosov diffeomorphisms.

We establish quickly the essential chaotic topological features of these hyperbolic systems: transitivity, density of periodic orbits, and structural stability. The central mechanism behind these properties is *shadowing*, and we devote some space to developing the shadowing property in both dynamical and geometric contexts.

In **Chapter 5** we turn to the ergodic theory of smooth systems. The central object of study here is a *conservative* dynamical system, that is, one that preserves a volume-type probability measure. We explore a zoo of examples. We then discuss ergodicity and prove the most basic result in ergodic theory (after Poincaré recurrence): the mean (Carleman–von Neumann) ergodic theorem.

We develop arguments for proving ergodicity of smooth dynamical systems, including density points and the Hopf argument. We illustrate these methods for the zoo of examples, starting with linear systems and moving to nonlinear systems, where distortion estimates become a key tool. This culminates in a proof of the ergodicity of geodesic flows in negative curvature, a celebrated result of Anosov.

Finally, we introduce the concept of entropy as a measure of information generated by a dynamical system, and define two types of entropy: topological and metric (measure theoretic). We show how to compute these entropies for several systems of interest and discuss a relationship between topological and metric entropies known as the variational principle of entropy. This relationship between metric and topological entropies is centered in a conjecture by Burns and A. Katok, which we prove in Chapter 10 for surfaces.

Chapter 6 is an appendix to Part II, in which we provide unified proofs of the Lebesgue density theorem, the pointwise ergodic theorem, and the martingale convergence theorem using maximal inequalities.

Chapter 7 is an introduction to dynamical rigidity, presented in the simplest case of expanding maps on the circle. For these dynamical systems, there are direct analogues of marked length spectrum rigidity and entropy rigidity that utilize some of the techniques in the proofs of their geometric counterparts.

Key among these techniques is the Livšic theory for hyperbolic systems, which is the study of solutions to dynamical cohomological equations. Livšic theory is an essential tool in constructing global objects in dynamical rigidity theory, including topological conjugacies, and it will be used in Chapter 8 to prove rigidity of the marked length spectrum. Using Livšic theory, we give as a warm-up a type of marked length spectrum rigidity for expanding maps on the circle.

We also establish some dynamical rigidity properties of Anosov diffeomorphisms and flows, discussing two major tools: Livšic theory and Journé's theorem.

Chapters 8, 9, and 10 treat in turn the proofs of Theorems A, C, and D. All three chapters contain more advanced material as well, giving alternate proofs and treating higher-dimensional cases.

Chapter 11, an appendix to Part III, develops a microlocal (Fourier integral) approach to spectral rigidity. We first discuss the proof of deformation rigidity of the marked length spectrum of Victor Guillemin and David Kazhdan in §11.1. In §11.3 we revisit the classic Zoll rigidity problem. Finally, in §11.4 we outline the recent Guillarmou–Lefeuvre proof of marked length spectrum rigidity in negative curvature, which combines microlocal methods with dynamical cohomology arguments.

1.4. What Is Not Covered in This Book

Dynamical and geometric rigidity go far beyond what is in this book. Each of these topics is fascinating and is itself the subject of many books. We give here some references for those interested in exploring these. An excellent introduction to some of these topics is Ralf Spatzier's survey [332].

- **Rigidity in positive curvature:** We discuss Zoll manifolds in Chapter 11; further reading on this and related topics is in Arthur Besse's text [41]. Wolfgang Ziller's text [369] treats positively curved manifolds in greater depth. For other positive curvature rigidity results (and more), see Marcel Berger's survey [38].
- **Einstein, Kähler–Einstein, and Calabi–Yau metrics:** The classic text on Kähler manifolds is Besse's book [42]. Thierry Aubin and Shing-Tung Yau's solution [22, 362, 363] of Eugenio Calabi's conjecture in complex geometry is also presented in Gang Tian's monograph [341].
- **Rigidity in representation theory:** We prove Mostow rigidity in Chapter 10; for a comprehensive discussion of other rigidity phenomena (such as Margulis superrigidity), see the text by Grigory Margulis [250].
- **Harmonic maps:** See the survey of James Eells and Luc Lemaire [122], the paper of Jürgen Jost and Yau [202], and the work of Mikhail Gromov and Richard Schoen [162].
- **Rigidity in conformal dynamics:** Dennis Sullivan has written many papers on the subject, some of them unpublished. On this topic, we recommend [336] and the monograph of Curtis McMullen [253].
- **Rigidity in arithmetic and algebraic dynamics:** Joseph Silverman's book [322] introduces the subject of arithmetic dynamics and discusses rigidity among Lattès maps. Laura DeMarco discusses rigidity phenomena around the distribution of periodic orbits in her survey [113]. In [88], Serge Cantat and Romain Dujardin survey their work on rigidity phenomena in the dynamics of automorphism groups of compact complex surfaces.

- **Measure rigidity in homogeneous dynamics and applications:** Marina Ratner's theorem is presented in elementary terms by David Witte Morris in [259]. Manfred Einsiedler and Elon Lindenstrauss survey measure rigidity (in particular, entropy methods) and its applications in [123]. Finally, Yves Benoist and Jean-François Quint present their measure rigidity results in [36].
- **Measure rigidity in Teichmüller and smooth dynamics:** The measure rigidity results [126, 127] of Alex Eskin, Maryam Mirzakhani, and Amir Mohammadi in Teichmüller dynamics are discussed in Alex Wright's beautiful essays [359, 360]. The Katok–Spatzier paper [210] lays out rigidity properties of smooth actions of discrete groups. We also recommend the book by Viorel Niţică and A. Katok [209] on the subject.
- **Quantum unique ergodicity (QUE):** See E. Lindenstrauss's survey [232] for quantum unique ergodicity for arithmetic hyperbolic surfaces and Steve Zelditch's survey [367] discussing microlocal techniques in QUE. Nalini Anantharaman's survey [9] reviews some of her foundational work on the subject. Peter Sarnak discusses some of the state of the art (in 2011) from an algebraic perspective in the paper [312].
- **Rank rigidity in nonpositive curvature:** Werner Ballmann text [30] is an in-depth presentation of the rigidity of geometric higher-rank manifolds established by Ballmann, Keith Burns, and Ralf Spatzier.
- **Rigidity of homeomorphism and diffeomorphism groups:** The Zimmer program (named for Robert Zimmer) is presented by David Fisher in [135] and Aaron Brown in [67]. These treatments discuss the history of the subject, leading up to the exciting recent advance by Brown, Fisher, and Sebastian Hurtado [68].

  For an introduction to groups acting on the circle, a must-read is Étienne Ghys's article [144], and Andrés Navas's book [267] is essential reading on smooth actions on the circle. Augustine Banyaga's book [34] connects the algebraic properties of symplectomorphism (and other diffeomorphism) groups to symplectic (and smooth) structures.
- **The Borel conjecture:** Chapter 10 contains proofs that, under the right (special) circumstances, a homotopy equivalence is homotopic to an isometry. A related topological phenomenon is addressed by the *Borel conjecture*, which posits that homotopy equivalence implies homeomorphism for any pair of aspherical manifolds; it is open in dimensions 4 and higher. We recommend Shmuel Weinberger's beautiful book on the subject [348].

1.5. Notes

1.5.1. More History The geometry of the hyperbolic plane was first discovered and developed around 1830 by János Bolyai, Nikolai Ivanovich Lobachevsky, and Carl Friedrich Gauss. Bernhard Riemann founded the field of Riemannian geometry in 1854, introducing manifolds, Riemannian metrics, and curvature. The term "hyperbolic geometry" was introduced by Felix Klein in 1871. Klein also introduced the Erlangen Program in 1872, in which he formalized the notion of a *geometry* as a simply connected space equipped with a transitive group of isometries. The study of hyperbolic surfaces began in earnest with the work of Poincaré (1882) in which he introduced the general notion of discrete groups of hyperbolic isometries, which he called Fuchsian groups, after Lazarus Fuchs. For a great (if slightly outdated) list of open problems in differential geometry, see S.-T. Yau's problem section [364].

1.5.2. Groups and Their Actions A (*left*) *action* of a group G on a space X is a map $\rho: G\times X \to X$ with the properties

(1) $\rho(e,x)=x$, for all $x\in X$,
(2) $\rho(g_1g_2,x)=\rho(g_1,\rho(g_2,x))$.

If G is a group of $n\times n$ real matrices, then the linear map $(A,x)\mapsto Ax$ on $\mathbb{R}^n$ is an example of a left action.

A *right action* $\hat{\rho}: G\times X\to X$ is very similar to a left action, except it reverses the order of composition: it satisfies

(1) $\hat{\rho}(e,x)=x$, for all $x\in X$,
(2) $\hat{\rho}(g_1g_2,x)=\hat{\rho}(g_2,\hat{\rho}(g_1,x))$.

Any left action ρ induces a right action $\hat{\rho}$ by setting $\hat{\rho}(g,x)=\rho(g^{-1},x)$ (and vice versa).

A group G acts on itself by the left action $g\mapsto hg$ and the right action $g\mapsto gh$. The quotient of G by the left action of a subgroup H is denoted $H\backslash G$, while the quotient by the right action is denoted G/H.

The action of a group G on a space X is *transitive* if for any two points $x,y\in X$ there exists $g\in G$ with $\rho(g,x)=y$. The action is *free* if the only element of G fixing any point in X is the identity e. If G acts freely and transitively on X, then fixing a point $x_0\in X$ gives a natural isomorphism between G and X via $g\mapsto\rho(g,x_0)$. More generally, if G acts transitively on X, then for any $x\in X$ the stabilizer subgroup

$$H=\{g\in G : \rho(g,x)=x\}$$

satisfies $X\cong G/H$.

When G is a topological group (i.e., equipped with a topology with respect to which the group operations are continuous) one can quantify the action topologically. We say that the action of G on a topological space X is

- *proper* if the map $\rho: G\times X\to X\times X$ defined by $\rho(g,x)=(x,gx)$ (or $\rho(g,x)=(x,xg)$ in the case of a right action) is a proper map (i.e., the preimage of every compact subset of $X\times X$ is compact in $G\times X$);
- *properly discontinuous* if for every compact subset $K\subset X$, the set $\{g\in G: gK\cap K\neq\emptyset\}$ is finite;
- *cocompact* if there exists a compact subset $K\subset X$ such that $G\cdot K=X$ (i.e., the quotient space X/G is compact).

(One makes similar definitions for right actions.) A topological group Γ is *discrete* if it is topologically discrete. For discrete groups, every properly discontinuous action is proper, and vice versa.

If X is a metric space and a discrete group G acts properly discontinuously on X, then the quotient space X/G is Hausdorff; moreover, if the action is free, then the projection $X\to X/G$ is a covering map. If X is a manifold and a discrete group G acts freely and properly discontinuously on X, then the quotient X/G is also a manifold.

If X is a smooth manifold and G is a Lie group (i.e., equipped with a smooth structure with respect to which the group actions are smooth, which includes the trivial case where G is discrete), we say the action of G on X is *smooth* if it is smooth when considered as a map $\rho: G\times X\to X$. If the action of G on X is smooth, proper, and free, then the quotient X/G inherits a unique smooth structure for which the natural projection $X\to X/G$ is a submersion (and a local diffeomorphism when G is discrete).

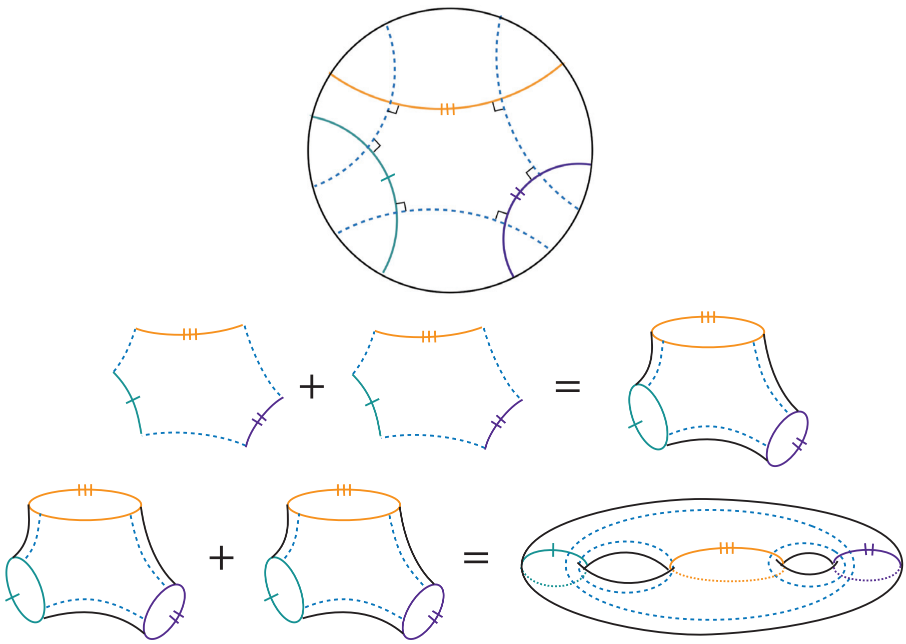


Figure 1.15. Constructing a hyperbolic genus 2 surface using two (identical) pairs of pants (with trivial twist parameters).

When G is a Lie group, its Lie algebra $\mathfrak{g}$ is defined as the tangent space at the identity, equipped with the Lie bracket induced by the commutator of left-invariant vector fields. In addition, every Lie group admits a left-invariant measure, the Haar measure, which is unique up to scaling. Similarly, every Lie group G admits a right-invariant Haar measure, unique up to scaling; the left and right Haar measures agree if and only if G is *unimodular* (which holds for many examples we will consider, such as when G is compact, nilpotent, or semisimple). Haar measure is indispensable in averaging arguments and plays a central role in the study of dynamical systems and representation theory.

A subgroup $\Gamma < G$ of a Lie group G is called a *lattice* if Γ is a discrete subgroup and the quotient space G/Γ has finite volume with respect to the Haar measure. We say a lattice is *uniform* or *cocompact* if the quotient space G/Γ is compact.

1.5.3. The Geodesic Flow on the Torus, Seen Algebraically. Because it is simpler, we describe the geodesic flow on the tangent bundle $T\mathbb{T}^2$ and leave the flow on $T^1\mathbb{T}^2$ to the reader. For $t \in \mathbb{R}$, let A_t be the 4×4 matrix

$$A_t := \begin{pmatrix} 1 & 0 & t & 0 \\ 0 & 1 & 0 & t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and let

$$G = \mathbb{R}^4 \rtimes_{A_t} \mathbb{R}$$

be the nilpotent Lie group of matrices of the form

$$\left\{ \left(\begin{smallmatrix} A_t & \underline{u} \\ 0 & 1 \end{smallmatrix} \right) : \underline{u} \in \mathbb{R}^4,\ t \in \mathbb{R} \right\} \subset \mathrm{SL}(5, \mathbb{R}),$$

under matrix multiplication. The closed (nondiscrete) subgroup

$$\Gamma := \left\{ \begin{pmatrix} A_s & \underline{z} \\ 0 & 1 \end{pmatrix} : \underline{z} = \begin{pmatrix} m \\ n \\ 0 \\ 0 \end{pmatrix},\ m, n \in \mathbb{Z},\ s \in \mathbb{R} \right\}$$

acts on G on the left, and $\Gamma \backslash G$ is diffeomorphic to $T\mathbb{T}^2$. The geodesic flow on $T\mathbb{T}^2$ is given by the right action of the 1-parameter subgroup

$$\mathscr{G}_{\mathbb{E}^2} := \left\{ \begin{pmatrix} A_t & 0 \\ 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\}.$$

Thus, although the torus itself is an abelian group, its geodesic flow can be realized as a unipotent flow on a homogeneous space of a (nonabelian) nilpotent Lie group.

1.5.4. Constructing Hyperbolic Surfaces Using a "Pair of Pants" Decomposition. Any closed hyperbolic surface of genus $g \geqslant 2$ can be decomposed into $2g-2$ pairs of pants (three-holed spheres). The hyperbolic structure on a pair of pants is uniquely determined by the lengths of its three boundary geodesics. After fixing these lengths, one glues the pairs of pants together along their boundaries using twist parameters; these twists and the cuff lengths together form the Fenchel–Nielsen coordinates (see Figure 1.15). In this way, the Teichmüller space of hyperbolic metrics on a surface of genus g is parametrized by $6g-6$ real parameters. For further details, see [78, §3.1] and [130, §8.3.1].

1.5.5. $\Gamma \backslash G/K$ These crystals of the dynamics world—the perfect elliptic, parabolic, and hyperbolic flows—are geodesic flows on the most symmetric Riemannian manifolds, the locally symmetric manifolds. A locally symmetric manifold is one for which reversing the direction of geodesics emanating from a point produces a local isometry (see Figure 1.16). Such manifolds are locally modeled on symmetric spaces[21]—simply connected manifolds in which the geodesic reflection at any point is a global isometry. While locally symmetric surfaces have constant curvature, in higher dimensions the curvature need not be constant (though the curvature tensor is parallel).

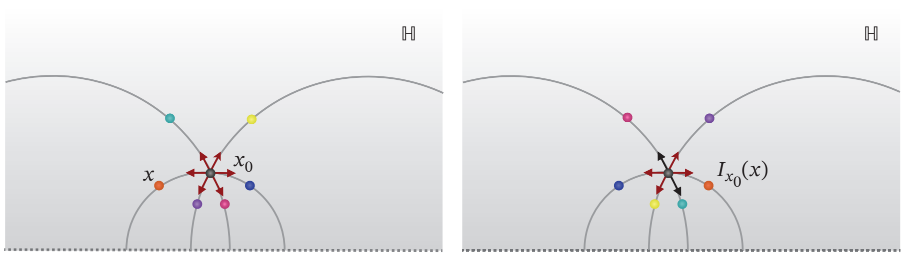


Figure 1.16. $\mathbb{H} = \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$ is a symmetric space: the time-reversing involution $I_{x_0} : \mathbb{H} \to \mathbb{H}$ through a point $x_0 \in \mathbb{H}$ is an isometry.

If M is a symmetric space, then the identity component of its isometry group G acts transitively on M. The stabilizer of a point is a compact subgroup K of G, so that $M \cong G/K$. The pointwise involution about the identity corresponds to an involutive automorphism of G fixing K, known as the *Cartan involution*. Every symmetric space

[21] These are also called *Riemannian* symmetric spaces, to distinguish them from pseudo-Riemannian symmetric spaces.

$\widetilde{M}$ can be uniquely decomposed as a product of irreducible symmetric spaces, and in the case of strictly positive or negative curvature the space is irreducible (though the converse is not true). Thus any locally symmetric manifold is of the form

$$M = \Gamma \backslash G/K,$$

where $\Gamma < G$ is (isomorphic to) the fundamental group of M. In our context, the locally symmetric surfaces are the flat tori, the 2-sphere $S^2 \cong \mathrm{SO}(3)/\mathrm{SO}(2)$, and the hyperbolic surfaces $\Gamma \backslash \mathrm{SL}(2,\mathbb{R})/\mathrm{SO}(2)$. The orientation-preserving isometry group of S^2 is $\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{\pm I\}$.

In general, the irreducible symmetric spaces were classified by Cartan and include the higher-dimensional analogues of the round sphere and hyperbolic plane. Those of nonpositive curvature include higher-rank symmetric spaces such as $\mathrm{SL}(n,\mathbb{R})/\mathrm{SO}(n)$ for $n \geqslant 3$, which are central to higher-rank rigidity phenomena. We discuss symmetric spaces in further detail at the end of the next chapter (Chapter 2, §2.7.2).

1.5.6. Local Versus Global Rigidity. Guillarmou and Lefeuvre proved that for any closed Riemannian manifold with Anosov geodesic flow (in particular, any manifold with negative sectional curvatures), the marked length spectrum locally determines the metric. That is, if two metrics are sufficiently C^N close (for N large enough) and have the same marked length spectrum, then they are isometric [166]. Their proof employs sophisticated microlocal analysis techniques (see Chapter 11, §11.4). This result is an example of *local rigidity*.

However, local rigidity results do not always extend globally. For instance, consider isoperimetric inequalities on the 2-sphere. Let $\underline{\ell}(g)$ denote the length of the shortest closed geodesic of a metric g on S^2. For the round metric g_0 on the unit sphere, we have $\underline{\ell}(g_0) = 2\pi$, and the same holds for Zoll metrics of area 4π. Balacheff showed that the second variation of $\underline{\ell}$ at g_0 is nonpositive and conjectured that the round sphere is a local maximum [27]. This conjecture was proved in 2017 by Alberto Abbondandolo, Barney Bramham, Umberto Hryniewicz, and Pedro Salomão, who demonstrated that for all metrics of sufficiently pinched positive curvature and fixed area, $\underline{\ell}$ is locally maximized precisely at the round sphere and Zoll surfaces [1].

Nonetheless, there exist other metrics on the sphere of area 4π for which the shortest closed geodesic is longer than 2π. For example, one may consider the "Calabi–Croke sphere," obtained by gluing two equilateral Euclidean triangles along their boundary and then smoothing the resulting conical metric. After rescaling to area 4π, the length of the shortest closed geodesic can be made arbitrarily close to $2(12)^{1/4}\sqrt{\pi} \approx 6.6$ (see Figure 1.17). Calabi conjectured that this value is the supremum over all such metrics; his conjecture remains open.

We thank André Neves for informing us about this work.

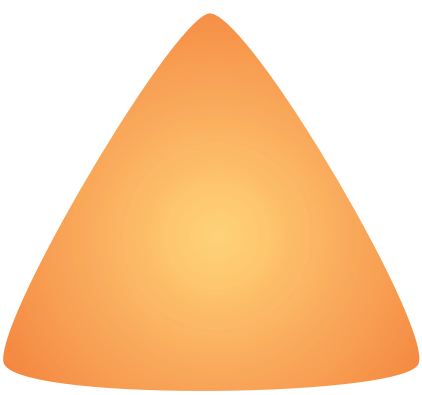


Figure 1.17. The Calabi–Croke pillow.

1.5.7. More on Zoll Surfaces and Besse Manifolds. A Riemannian manifold in which all geodesics are closed is known as a *Besse manifold*, and a *Zoll manifold* is a Besse manifold in which all unit-speed geodesics have the same minimal period (so that the lengths of all geodesics are integer multiples of this period).[22]

Victor Guillemin characterized the Zoll metrics that are deformations of the round metric on the sphere [169]. We discuss Guillemin's method further in Chapter 11, §11.4.

Marco Mazzucchelli and Stefan Suhr [252] proved a form of length spectrum rigidity for the 2-sphere: if all *simple* closed geodesics of a metric g on S^2 have length 2π, then all geodesics are simple and closed, of length 2π, i.e., the metric is Zoll (this result was earlier claimed by Lusternik, albeit in a weaker form). Their proof employs Lusternik–Schnirelmann theory, a variational method for finding closed geodesics on spheres.

Detlef Gromoll and Karsten Grove [153] showed that for any Riemannian metric on S^2, if all geodesics are closed, then they are all simple and have the same length (i.e., there are no non-Zoll Besse metrics on S^2). Marcel Berger conjectured that this property holds for any topological n-sphere, and this was proved for $n > 3$ by Marco Radeschi and Burkhard Wilking [293].

We discuss Zoll metrics in depth in Chapter 11, §11.3.

Exercises

EXERCISE 1.1 (Combinatorial Gauss–Bonnet). Let S be a closed, connected, orientable surface equipped with a triangulation $\mathcal{T}$. Denote by V, E, and F the sets of vertices, edges, and triangles, respectively, and write $\chi(S) = \#V - \#E + \#F$ for the Euler characteristic. Give every triangle the *flat* Euclidean metric in which its three edge lengths equal 1; each face is then an equilateral triangle with interior angles $\pi/3$.

(1) For a vertex $v \in V$ let
$$\kappa(v) = 2\pi - \sum_{\triangle \ni v} \measuredangle_v(\triangle)$$
be the angle deficit at v. Show that
$$\sum_{v \in V} \kappa(v) = 2\pi\chi(S).$$
Hint: Each triangle contributes π to the total of the angles appearing in the sum over v, because $\measuredangle_1 + \measuredangle_2 + \measuredangle_3 = \pi$ in Euclidean geometry. Count triangles, edges, and vertices separately and combine with Euler's formula.

(2) Interpret $\kappa(v)$ as a discrete (combinatorial) curvature and conclude that the classical Gauss–Bonnet formula $\int_S K\,dA = 2\pi\chi(S)$ is already encoded in the combinatorics of any equilateral triangulation.

EXERCISE 1.2. *Figure 1.8 depicts a C^1 curve of constant width. Construct a C^∞ curve of constant width. Is there an analytic one? See Figure 1.18 for other curves of constant width.

EXERCISE 1.3. The quaternion $\mathbf{H}$ is the set of $q = x_1 + \mathbf{i}x_2 + \mathbf{j}x_3 + \mathbf{k}x_4$, with $(x_1, x_2, x_3, x_4) \in \mathbb{R}^4$. The quaternions are a division algebra over $\mathbb{R}$ with a product defined by $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$.

[22]Sometimes the requirement that all geodesics be simple is added to the definition of "Zoll metric." In dimension 2, this is no extra imposition [153], but in higher dimensions, this further restricts the definition.

Figure 1.18. Coins of constant width, from the UK: two of them are not round, but they all work in vending machines. Photo credit: iStock.com/pamela_d_mcadams.

For $q \in \mathbf{H}$, set $q^* = x_1 - \mathbf{i}x_2 - \mathbf{j}x_3 - \mathbf{k}x_4$. Note that $qq^* = |q|^2$, where $|q|$ is the ordinary Euclidean norm induced by the identification of $\mathbf{H}$ with $\mathbb{R}^4$. Note that $\mathbf{H}^1 = \{q \in \mathbf{H} : qq^* = 1\}$ is the 3-sphere.

(1) Let $\Im(\mathbf{H}) = \{q = \mathbf{i}x_2 + \mathbf{j}x_3 + \mathbf{k}x_4\}$ be the set of imaginary quaternions, which we identify with $\mathbb{R}^3$ in the natural way, and let $\Im(\mathbf{H})^1 = \{u \in \Im(\mathbf{H}) : |u| = 1\}$ be the set of unit imaginary quaternions. Note that $\Im(\mathbf{H})^1$ is the 2-sphere (embedded in $\mathbb{R}^3$). For $u \in \Im(\mathbf{H})^1$ and $\theta \in \mathbb{R}$ we define $e^{u\theta} := \cos\theta + u\sin\theta$. Show that every $v \in \mathbf{H}^1$ can be written in the form $v = e^{u\theta}$, with $\theta \in [0, 2\pi]$ and $u \in \Im(\mathbf{H})^1$. Observe that $(e^{u\theta})^* = e^{-u\theta}$.

(2) Fix $v \in \mathbf{H}^1$ and consider the map $R_v : \Im(\mathbf{H}) \to \Im(\mathbf{H})$ given by $R_v(q) = vqv^*$. Show that R_v is a rigid rotation of $\mathbb{R}^3$. What are the axis and angle of this rotation?

(3) Fix $u \in \Im(\mathbf{H})^1$ and consider the map $p_u : \mathbf{H}^1 \to \Im(\mathbf{H})^1$ given by $p_u(q) = quq^*$. Prove that this map is a surjective, proper submersion. Show that this is the Hopf fibration.

EXERCISE 1.4 (See [96]). Let Λ_1 and Λ_2 be lattices in $\mathbb{R}^2$.

(1) Prove that the geodesic flows for $\mathbb{R}^2/\Lambda_1$ and $\mathbb{R}^2/\Lambda_2$ are *smoothly orbit equivalent:* there exists a diffeomorphism $h : T^1\mathbb{R}^2/\Lambda_1 \to T^1\mathbb{R}^2/\Lambda_2$ sending orbits of the geodesic flow φ_t^1 on $T^1\mathbb{R}^2/\Lambda_1$ to orbits of the geodesic flow φ_t^2 on $T^1\mathbb{R}^2/\Lambda_2$: for every $(x, v) \in T^1\mathbb{R}^2/\Lambda_1$,

$$h(\varphi_t^1(x, v)) = \varphi^2_{\tau_{(x,v)}(t)}(h(x, v)), \quad t \in \mathbb{R},$$

Hint: Choose an invertible linear map $A \in \mathrm{GL}_2(\mathbb{R})$ with $A(\Lambda_1) = \Lambda_2$.

(2) Suppose that there exists a *smooth conjugacy* between the flows, that is, a diffeomorphism $h : T^1\mathbb{R}^2/\Lambda_1 \to T^1\mathbb{R}^2/\Lambda_2$ such that, for every $(x, v) \in T^1\mathbb{R}^2/\Lambda_1$ and every $t \in \mathbb{R}$,

$$h \circ \varphi_t^1(x, v) = \varphi_t^2 \circ h(x, v).$$

Prove that $\mathbb{R}^2/\Lambda_1$ and $\mathbb{R}^2/\Lambda_2$ are isometric. **Hint:** Argue that an orbit equivalence must be affine, with linear part $A \in \mathrm{GL}(2, \mathbb{R})$, and show that the conjugacy condition forces A to be orthogonal (up to the natural identification of the flat structures) and to satisfy $A(\Lambda_1) = \Lambda_2$.

EXERCISE 1.5. Find a Möbius transformation taking the disk to the upper half-plane and verify that the hyperbolic metric on $\mathbb{H}$ pulls back to the metric on $\mathbb{D}$. Prove that the geodesics in $\mathbb{H}$ and $\mathbb{D}$ are precisely the circular arcs orthogonal to the boundary.

EXERCISE 1.6. *See [130] for a description of the pair of pants decomposition for hyperbolic surfaces. Using this decomposition, prove that any two hyperbolic metrics on a fixed surface are isometric if they have the same marked length spectrum. How many values of ℓ are needed to determine the metric?

INDEX