Deep connections between the asymptotic growth of simple closed geodesics on hyperbolic surfaces and the geometry of moduli spaces established by Mirzakhani

February 17, 2025

1 Introduction

In this survey, we explore the fascinating interplay between number theory, geometry, and dynamical systems. To set the stage, we begin by recalling a classical result from analytic number theory—the Prime Number Theorem—which describes the asymptotic distribution of prime numbers. This theorem, a cornerstone of number theory, naturally leads us to examine analogous asymptotic counting problems in geometry, particularly the enumeration of closed geodesics on hyperbolic surfaces.

Several key works form the backbone of our approach. Mirzakhani's groundbreaking study [6] established deep connections between the asymptotic growth of simple closed geodesics on hyperbolic surfaces and the geometry of moduli spaces, while Arana-Herrera [1] provides a modern ergodic-theoretic perspective on counting problems that range from primitive integer points to simple closed curves. Foundational background on surface topology and mapping class groups is supplied by Farb and Margalit's *A Primer on Mapping Class Groups* [4] as well as Martelli's *An Introduction to Geometric Topology* [5]. In addition, comprehensive treatments of hyperbolic geometry and its spectral theory are available in Ratcliffe's *Foundations of Hyperbolic Manifolds* [7], Borthwick's *Spectral Theory of Infinite-Area Hyperbolic Surfaces* [2], and Dal'Bo's work on geodesic and horocyclic trajectories [3].

By synthesizing ideas from these sources, our survey aims to illustrate how techniques from number theory and dynamical systems can be adapted to study the geometry of hyperbolic surfaces and the asymptotic behavior of their closed geodesics.

The Prime Number Theorem and Asymptotic Notation

One of the crown jewels of analytic number theory is the *Prime Number Theorem.* Recall that if $\pi(N)$ denotes the number of prime numbers less than or equal to N, the Prime Number Theorem asserts that

$$\pi(N) \sim \frac{N}{\log(N)}$$

as $N \to \infty$. Here, the notation

$$f(N) \sim g(N)$$

means that

$$\lim_{N\to\infty} \frac{f(N)}{g(N)} = 1.$$

In other words, for large N the ratio of $\pi(N)$ to $N/\log(N)$ becomes arbitrarily close to 1. This theorem is a profound statement about the distribution of primes among the integers and is usually proved using techniques from complex analysis and analytic number theory.

1.1 Proofs of Quotient Properties of $\Gamma\backslash\mathbb{H}$

1.1.1 Statement 1: If Γ is torsion-free, then $\Gamma\backslash\mathbb{H}$ is smooth

Proof:

- **Free and Properly Discontinuous Action:**
  Recall that a group Γ acting on a manifold M (here, $M = \mathbb{H}$) is said to act *freely* if no nontrivial element of Γ fixes any point of M. If Γ is torsion-free, then for every nonidentity $\gamma \in \Gamma$, the equation

  $$\gamma(z) = z$$

  has no solution in $\mathbb{H}$. Therefore, the action of Γ on $\mathbb{H}$ is free.

- **Local Quotient Structure:**
  A discrete group of isometries (in our case, a Fuchsian group) acts *properly discontinuously* on $\mathbb{H}$. This means that for any point $z \in \mathbb{H}$, there exists a neighborhood U of z such that $\gamma(U) \cap U = \varnothing$ for all $\gamma \in \Gamma \setminus \{id\}$. Hence, the projection

  $$\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}$$

  is a covering map. Since the action is free, every point in the quotient has a neighborhood that is diffeomorphic to an open set in $\mathbb{R}^2$.

- **Conclusion:**
  Because the local charts are given by open subsets of $\mathbb{R}^2$ (with no "twisting" coming from nontrivial isotropy), the quotient $\Gamma\backslash\mathbb{H}$ is a smooth surface (i.e., a smooth 2-manifold).

1.1.2 Statement 2: If Γ has elliptic elements, then the quotient $\Gamma\backslash\mathbb{H}$ is an orbifold (with singular cone points)

Proof:

- **Elliptic Elements and Fixed Points:**
  An element $\gamma \in \Gamma$ is called *elliptic* if it has finite order. Such an element fixes some point $z_0 \in \mathbb{H}$. That is, there exists a nontrivial $\gamma \in \Gamma$ and a point z_0 such that

  $$\gamma(z_0) = z_0.$$

  Hence, the action of Γ is not free: the stabilizer (isotropy group) of z_0 is a nontrivial finite cyclic subgroup of Γ.

- **Local Model for the Quotient Near a Fixed Point:**
  By the general theory of group actions, if a finite group G acts linearly on $\mathbb{R}^2$, then a neighborhood of the origin in $\mathbb{R}^2$ quotiented by this action is homeomorphic to $\mathbb{R}^2/G$. In our case, near a fixed point z_0, the quotient $\Gamma\backslash\mathbb{H}$ is modeled on $\mathbb{R}^2/\mathbb{Z}_n$ (for some $n > 1$), which is not a smooth manifold—it is an orbifold chart with a cone point (a singular point with cone angle $2\pi/n$).

- **Conclusion:**
  Therefore, when Γ contains elliptic elements, the quotient $\Gamma\backslash\mathbb{H}$ is not a smooth manifold but rather an orbifold, with the images of the fixed points (under the projection π) appearing as singular cone points.

1.1.3 Statement 3: Hyperbolic elements produce closed geodesics via their invariant axes

Proof:

- **Properties of Hyperbolic Elements:**
  A hyperbolic element $\gamma \in \Gamma$ has the following properties:
  - It has two distinct fixed points on the boundary $\partial\mathbb{H}$.
  - It possesses a unique invariant geodesic A (called the *axis* of γ) in $\mathbb{H}$.
  - The action of γ on A is by translation: there exists a constant $L > 0$ (the translation length) such that for every point $z \in A$,

    $$d(z, \gamma(z)) = L.$$

- **Quotienting the Axis:**
  Consider the projection

  $$\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}.$$

  Since A is invariant under the cyclic subgroup $\langle\gamma\rangle$ generated by γ, the set of points $\{\gamma^n(z) : n \in \mathbb{Z}\}$ for any $z \in A$ is exactly the orbit of z under $\langle\gamma\rangle$. The geodesic A is "wrapped up" under the identification induced by $\langle\gamma\rangle$ and its projection $\pi(A)$ is a closed curve in $\Gamma\backslash\mathbb{H}$.

- **Closed Geodesic:**
  Because A is a geodesic in $\mathbb{H}$ and γ acts by a constant translation along A, the projection $\pi(A)$ is a closed geodesic in the quotient. In many cases (for example, if the geodesic does not self-intersect), this closed geodesic is simple.

- **Conclusion:**
  Thus, every hyperbolic element γ produces a closed geodesic in $\Gamma\backslash\mathbb{H}$ via its invariant axis A.

1.1.4 Summary

1. **Torsion-Free Γ:** The action is free and properly discontinuous, so $\Gamma\backslash\mathbb{H}$ is a smooth manifold.

2. **Elliptic Elements:** Their fixed points lead to nonfree actions. Locally, the quotient looks like $\mathbb{R}^2/\mathbb{Z}_n$ (a cone point), so the quotient is an orbifold with singularities.

3. **Hyperbolic Elements:** They have invariant axes along which they translate by a fixed positive length, so the quotient of such an axis is a closed geodesic in $\Gamma\backslash\mathbb{H}$.

Geometric Analogues: Closed Geodesics on Hyperbolic Surfaces

It turns out that similar counting problems occur in a geometric setting. To explain this, we first introduce the concept of a *hyperbolic surface*. A hyperbolic surface is a two-dimensional Riemannian manifold whose geometry is modeled on the hyperbolic plane $\mathbb{H}^2$ (a non-Euclidean space of constant negative curvature, usually normalized to -1). One may think of $\mathbb{H}^2$ as the upper half-plane

$$\{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$$

with the metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

A *closed geodesic* on a hyperbolic surface X is a curve that is both geodesic (locally the shortest path between any two points on it) and closed (its endpoints coincide). Such geodesics can be *primitive* or non-primitive. A closed geodesic is said to be *primitive* if it is not obtained by repeatedly traversing a shorter geodesic. For example, if γ is a closed geodesic and γ^n denotes the curve that follows γ n times consecutively, then γ^n is *not* primitive (unless $n = 1$).

On Closed, Simple, and Primitive Geodesics and Their Group-Theoretic Interpretation

In our previous discussion, we considered a hyperbolic surface

$$X = \Gamma\backslash\mathbb{H}^2,$$

where Γ is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$. For simplicity, we often assume that Γ is torsion-free so that every nontrivial element is hyperbolic. In this setting, every nontrivial element of Γ acts by translating along a unique geodesic (its *axis*) in $\mathbb{H}^2$, and the displacement length (i.e., the distance moved along this axis) coincides with the length of the closed geodesic on X obtained by projecting the axis.

Before delving into the details, we now introduce and clarify some fundamental definitions.

Aspect	Prime Geodesic Theorem	Mirzakhani's Theorem
Objects Counted	Primitive closed geodesics on a hyperbolic surface (these may have self–intersections).	Simple closed geodesics (without self–intersections) on a hyperbolic surface.
Growth Rate	Exponential growth: roughly $\frac{e^T}{T}$ for geodesics of length $\leq T$.	Polynomial growth: approximately $c(S)\, L^{6g-6+2n}$ for a surface of genus g with n boundaries/cusps and length $\leq L$.
Asymptotic Formula	$\pi(T) \sim \frac{e^T}{T}$ as $T \to \infty$, with constants determined by the spectral data.	$N(L) \sim c(S)\, L^{6g-6+2n}$ as $L \to \infty$, where $c(S)$ depends on the geometry of the surface S.
Underlying Techniques	Spectral theory, the Selberg trace formula, and methods analogous to those in the Prime Number Theorem.	Ergodic theory, hyperbolic geometry, Weil–Petersson volume computations, and dynamical techniques (e.g., earthquake flow).
Motivational Analogy	Closed geodesics are viewed as the geometric analogues of prime numbers.	Although simple geodesics might be thought of as "primes" under topological constraints, their counting is much sparser.
Historical Context	Originated from mid–20th century work by Selberg, Margulis, and others.	Emerged in the early 2000s through Maryam Mirzakhani's work, which contributed significantly to our understanding of moduli spaces.
Scope & Applications	Fundamental in spectral geometry, quantum chaos, and the theory of automorphic forms.	Central to the study of moduli spaces of Riemann surfaces, mapping class groups, and combinatorial aspects of surface geometry.
Counting Domain	Applies to all primitive closed geodesics (up to conjugacy) on the hyperbolic surface.	Applies only to the subset of simple closed geodesics.

Table 1: Comparison of the Prime Geodesic Theorem and Mirzakhani's Result

Definitions

Closed Geodesic: A *geodesic* on a Riemannian manifold (or hyperbolic surface) X is a curve that locally minimizes distance. A geodesic is said to be *closed* if it is periodic—that is, if it is given by a continuous map

$$\gamma : \mathbb{R} \to X$$

satisfying $\gamma(t+T) = \gamma(t)$ for all t and some minimal period $T > 0$. The number T is then called the *length* of the closed geodesic.

Simple Geodesic: A closed geodesic is called *simple* if it does not intersect itself. Formally, if $\gamma(t_1) = \gamma(t_2)$ for some $t_1, t_2 \in \mathbb{R}$, then $t_1 \equiv t_2 \pmod{T}$.

Primitive Geodesic: A closed geodesic is *primitive* if it is not obtained by traversing a shorter closed geodesic repeatedly. In other words, if γ is a closed geodesic and γ^n denotes the curve that follows γ n times consecutively (with length nT), then γ^n is *not* primitive unless $n = 1$.

Group-Theoretic Interpretation

The theory of closed geodesics on a hyperbolic surface X is intimately connected with the algebraic properties of the fundamental group $\pi_1(X)$. When X is represented as

$$X = \Gamma \backslash \mathbb{H}^2,$$

the conjugacy classes in the discrete group Γ correspond bijectively to the free homotopy classes of closed curves on X.

Displacement Length and Closed Geodesics. For a hyperbolic element $\gamma \in \Gamma$, the *displacement length* is defined by

$$\ell(\gamma) \ := \ \inf_{z \in \mathbb{H}^2} d(z, \gamma z).$$

It is a standard fact that:

1. The function $z \mapsto d(z, \gamma z)$ attains its minimum on the unique geodesic (the *axis* of γ) that is invariant under γ.
2. The displacement length $\ell(\gamma)$ is positive and is exactly the distance by which γ translates points along its axis.

Thus, when we project the axis of γ to the surface X, we obtain a closed geodesic whose length is precisely $\ell(\gamma)$.

Primitive Elements. An element $\gamma \in \Gamma$ is called *primitive* if it is not a proper power of another element; that is, if

$$\gamma = \gamma_0^n \quad \text{for some } \gamma_0 \in \Gamma \text{ and } n \geq 1,$$

then necessarily $n = 1$. Equivalently, a closed geodesic on X is primitive if it is not the multiple traversal of a shorter closed geodesic. In the group picture, only the conjugacy classes of primitive elements correspond to primitive closed geodesics.

Discussion of Other Types of Elements:

While in many treatments we assume Γ is torsion-free so that every nontrivial element is hyperbolic, in a broader context one may encounter the following types of elements in $\mathrm{PSL}(2, \mathbb{R})$:

1. **Hyperbolic Elements:** These elements satisfy $|\mathrm{tr}(\gamma)| > 2$ and have two distinct fixed points on the boundary $\partial\mathbb{H}^2$. They act by translating along a unique geodesic (the axis), and their displacement length is positive and equals the length of the corresponding closed geodesic on X. Hyperbolic elements are the ones that give rise to closed geodesics, and the notion of primitivity applies naturally to them.

2. **Parabolic Elements:** Parabolic elements have $|\mathrm{tr}(\gamma)| = 2$ and possess a single fixed point on $\partial\mathbb{H}^2$. They act by "shearing" along horocycles centered at this fixed point and do not have an axis in the traditional sense. In a hyperbolic surface, parabolic elements are associated with cusps. Since they do not translate points along a geodesic by a fixed positive distance (i.e., they lack a finite displacement length), they do not yield closed geodesics in the usual sense.

3. **Elliptic Elements:** Elliptic elements satisfy $|\mathrm{tr}(\gamma)| < 2$ and have a fixed point in the interior of $\mathbb{H}^2$. They act as rotations about that fixed point. Elliptic elements occur when the quotient $X = \Gamma\backslash\mathbb{H}^2$ is an *orbifold* rather than a smooth manifold. Since elliptic elements rotate rather than translate, the notion of displacement length does not apply, and they do not produce closed geodesics in the standard sense.

4. **Loxodromic Elements:** In the context of Kleinian groups acting on hyperbolic 3-space $\mathbb{H}^3$ or higher-dimensional hyperbolic spaces, one distinguishes between hyperbolic and loxodromic elements. In two dimensions ($\mathbb{H}^2$), every hyperbolic element is also loxodromic. The term "loxodromic" is sometimes used to emphasize that the isometry combines translation with a rotational twist (a spiral motion). However, in $\mathbb{H}^2$ any hyperbolic element can be conjugated into a diagonal form (up to sign), so the distinction is not essential.

Summary: In our standard setting of a hyperbolic surface $X = \Gamma\backslash\mathbb{H}^2$ with Γ torsion-free:

- Every nontrivial element of Γ is hyperbolic.
- A closed geodesic on X is the projection of the axis of a hyperbolic element.
- The displacement length of a hyperbolic element equals the length of the corresponding closed geodesic.
- A closed geodesic is *primitive* if it is not a proper multiple of a shorter geodesic; equivalently, a hyperbolic element is primitive if it is not a proper power in Γ.

Parabolic elements (which arise in the presence of cusps) and elliptic elements (which occur when X is an orbifold) do not yield closed geodesics in the standard sense because they do not have a finite displacement length along a geodesic.

Displacement Length of Hyperbolic Elements

Let $\gamma \in \mathrm{SL}(2, \mathbb{R})$ be a hyperbolic element. This means that

$$|\operatorname{tr}(\gamma)| > 2.$$

Then γ has two distinct real eigenvalues, say λ and λ^{-1} (with $|\lambda| > 1$), and a unique invariant geodesic (its *axis*) in the hyperbolic plane $\mathbb{H}^2$. The *translation length* $\ell(\gamma)$ is defined as the distance by which γ translates any point on its axis:

$$\ell(\gamma) = \inf_{z \in \mathbb{H}^2} d(z, \gamma(z)).$$

It turns out that

$$\ell(\gamma) = 2 \cosh^{-1}\Bigl(\frac{|\operatorname{tr}(\gamma)|}{2}\Bigr).$$

Below we provide three proofs of this fact.

Proof 1: Using Eigenvalues

Since γ has eigenvalues λ and λ^{-1} with $|\lambda| > 1$, we have

$$\operatorname{tr}(\gamma) = \lambda + \lambda^{-1}.$$

It is well known that a hyperbolic isometry translates points along its axis by a distance

$$\ell(\gamma) = 2 \log |\lambda|.$$

Now, note that

$$e^{\ell(\gamma)/2} = |\lambda| \quad \text{and} \quad e^{-\ell(\gamma)/2} = |\lambda|^{-1}.$$

Thus,

$$e^{\ell(\gamma)/2} + e^{-\ell(\gamma)/2} = |\lambda| + |\lambda|^{-1} = \frac{|\operatorname{tr}(\gamma)|}{1} = \lambda + \lambda^{-1}.$$

But by definition,

$$e^{\ell(\gamma)/2} + e^{-\ell(\gamma)/2} = 2 \cosh\Bigl(\frac{\ell(\gamma)}{2}\Bigr).$$

Hence,

$$2 \cosh\Bigl(\frac{\ell(\gamma)}{2}\Bigr) = \lambda + \lambda^{-1} = \operatorname{tr}(\gamma),$$

so that

$$\cosh\Bigl(\frac{\ell(\gamma)}{2}\Bigr) = \frac{|\operatorname{tr}(\gamma)|}{2}.$$

Taking the inverse hyperbolic cosine yields

$$\frac{\ell(\gamma)}{2} = \cosh^{-1}\Bigl(\frac{|\operatorname{tr}(\gamma)|}{2}\Bigr) \quad \Longrightarrow \quad \ell(\gamma) = 2 \cosh^{-1}\Bigl(\frac{|\operatorname{tr}(\gamma)|}{2}\Bigr).$$

Proof 2: Conjugation to a Diagonal Form

Since γ is hyperbolic, there exists a matrix $A \in \mathrm{SL}(2,\mathbb{R})$ such that

$$\gamma = A \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix} A^{-1},$$

with $|\lambda| > 1$. Conjugating further if necessary, we may assume that γ fixes 0 and ∞; in this case, γ acts by

$$\gamma(z) = \lambda^2 z,$$

for $z \in \mathbb{H}^2$ (where we have written λ^2 instead of λ by a suitable normalization). The hyperbolic distance between a point $z > 0$ (on the positive real axis, which is a geodesic in $\mathbb{H}^2$) and its image $\gamma(z) = \lambda^2 z$ is

$$d(z, \gamma(z)) = \left| \log \left(\frac{\gamma(z)}{z} \right) \right| = \log(\lambda^2) = 2 \log |\lambda|.$$

Thus, the translation length is

$$\ell(\gamma) = 2 \log |\lambda|.$$

Now, as in Proof 1, note that

$$\mathrm{tr}(\gamma) = \lambda + \lambda^{-1} = 2 \cosh\Big(\frac{\ell(\gamma)}{2} \Big),$$

so that

$$\ell(\gamma) = 2 \cosh^{-1}\Big(\frac{|\,\mathrm{tr}(\gamma)|}{2} \Big).$$

Proof 3: Using the Cross-Ratio (Ahlfors' Approach)

The cross ratio is a projective invariant that can be used to express the translation length of a hyperbolic element. Let $\gamma \in \mathrm{SL}(2,\mathbb{R})$ be hyperbolic with fixed points ξ, η on $\partial\mathbb{H}^2$ (with, say, $\xi < \eta$ in a suitable coordinate system). For any point z on the geodesic connecting ξ and η, define the cross ratio

$$[z, \gamma(z); \xi, \eta] \;=\; \frac{(z - \xi)(\gamma(z) - \eta)}{(z - \eta)(\gamma(z) - \xi)}.$$

It can be shown (see, e.g., Ahlfors' *Complex Analysis*) that this cross ratio is independent of the choice of z on the geodesic and that

$$[z, \gamma(z); \xi, \eta] = e^{\ell(\gamma)},$$

where $\ell(\gamma)$ is the translation length of γ.

Taking the logarithm of both sides yields

$$\ell(\gamma) = \log[z, \gamma(z); \xi, \eta].$$

Now, if we conjugate γ so that its fixed points become 0 and ∞, then the cross ratio simplifies. In this conjugated form, the transformation acts by

$$\gamma(z) = \lambda^2 z,$$

and the fixed points are 0 and ∞. Choosing any $z > 0$ (which lies on the geodesic joining 0 and ∞), we have

$$[z, \gamma(z); 0, \infty] = \frac{(z - 0)(\gamma(z) - \infty)}{(z - \infty)(\gamma(z) - 0)}.$$

More rigorously, by a standard computation the cross ratio reduces to

$$[z, \gamma(z); 0, \infty] = \frac{\gamma(z)}{z} = \lambda^2.$$

Taking the logarithm, we obtain

$$\ell(\gamma) = \log(\lambda^2) = 2 \log |\lambda|.$$

As in the earlier proofs, since

$$\operatorname{tr}(\gamma) = \lambda + \lambda^{-1} = 2 \cosh\Big(\frac{\ell(\gamma)}{2}\Big),$$

we conclude that

$$\ell(\gamma) = 2 \cosh^{-1}\Big(\frac{|\operatorname{tr}(\gamma)|}{2}\Big).$$

Summary

We have provided three proofs:

1. Using eigenvalues directly.
2. Conjugating γ to a diagonal form.
3. Using the cross ratio as in Ahlfors' approach.

Each proof shows that for a hyperbolic element $\gamma \in \mathrm{SL}(2, \mathbb{R})$, the translation length is given by

$$\ell(\gamma) = 2 \cosh^{-1}\Big(\frac{|\operatorname{tr}(\gamma)|}{2}\Big).$$

Note that parabolic and elliptic elements do not have a finite translation length along a unique geodesic in $\mathbb{H}^2$ (parabolic elements fix a single point on the boundary and elliptic elements fix a point in $\mathbb{H}^2$), so this formula applies only to hyperbolic (or, in higher dimensions, loxodromic) elements.

Example: Consider the matrix

$$\gamma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$

A quick calculation shows that

$$\operatorname{tr}(\gamma) = 3 > 2,$$

so γ is hyperbolic. Its translation length is computed by

$$\ell(\gamma) = 2\cosh^{-1}\left(\frac{|\operatorname{tr}(\gamma)|}{2}\right) = 2\cosh^{-1}\left(\frac{3}{2}\right).$$

Thus, γ moves points along its unique invariant geodesic in $\mathbb{H}^2$ by a distance $2\cosh^{-1}(3/2)$. The closed geodesic on $X = \Gamma\backslash\mathbb{H}^2$ corresponding to γ has this length. Moreover, since γ cannot be written as a nontrivial power of another element in Γ, it is classified as *primitive*.

Displacement Length of Elliptic Elements

We want to show that if

$$\ell(T) = \inf_{z\in\mathbb{H}^2} d_{\mathbb{H}^2}(z, T(z))$$

is the minimal displacement (translation length) of a hyperbolic isometry T, then

$$\ell(T) = 2\operatorname{arccosh}\left(\frac{\operatorname{Tr}(T)}{2}\right).$$

In other words, the expression

$$2\operatorname{arccosh}\left(\frac{\operatorname{Tr}(T)}{2}\right)$$

comes from computing the minimal distance that T moves a point in $\mathbb{H}^2$.

In what follows we give a detailed proof of this fact.

—

Step 1. Reduction to a Canonical Form

A hyperbolic isometry T has two fixed points on the boundary of $\mathbb{H}^2$ and leaves invariant a unique geodesic (called the axis of T). It is a standard fact that any hyperbolic isometry in $\mathrm{PSL}(2,\mathbb{R})$ is conjugate to a dilation. That is, there exists a Möbius transformation M such that

$$M \circ T \circ M^{-1}(z) = \lambda^2 z,$$

with $\lambda > 1$. (Sometimes one writes the dilation as $z \mapsto e^{\ell/2}z$; we will see below that the translation length is then $\ell = 2\log\lambda$.)

Since the hyperbolic metric is invariant under Möbius transformations, the translation length of T is the same as that of its conjugate. Hence, without loss of generality we may assume that

$$T(z) = \lambda^2 z,$$

with $\lambda > 1$.

Step 2. Compute the Minimal Displacement for a Dilation
The Hyperbolic Distance

Recall that in the upper half-plane model the hyperbolic distance between z and w is given by

$$d_{\mathbb{H}^2}(z,w) = \operatorname{arccosh}\left(1 + \frac{|z-w|^2}{2\,\operatorname{Im}(z)\operatorname{Im}(w)}\right).$$

A key observation is that the dilation

$$T(z) = \lambda^2 z$$

has its invariant geodesic along the positive imaginary axis. (More generally, one can show that the axis of T is the unique geodesic joining 0 and ∞; in our case, that is the vertical line $\{x=0,\ y>0\}$.)

Thus, choose a point on the axis:

$$z = i\,y, \quad y > 0.$$

Then

$$T(i\,y) = \lambda^2(i\,y) = i\,\lambda^2 y.$$

Computing $d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y)$

For two points on the imaginary axis $z = i\,y$ and $w = i\,(\lambda^2 y)$, we have

- $|z-w| = |i\,y - i\,\lambda^2 y| = y|\lambda^2 - 1|$.
- $\operatorname{Im}(z) = y$ and $\operatorname{Im}(w) = \lambda^2 y$.

Thus,

$$d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y) = \operatorname{arccosh}\left(1 + \frac{(y|\lambda^2-1|)^2}{2\,y\,(\lambda^2 y)}\right) = \operatorname{arccosh}\left(1 + \frac{y^2(\lambda^2-1)^2}{2\lambda^2 y^2}\right).$$

Simplify the fraction:

$$\frac{y^2(\lambda^2-1)^2}{2\lambda^2 y^2} = \frac{(\lambda^2-1)^2}{2\lambda^2}.$$

Thus,

$$d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y) = \operatorname{arccosh}\left(1 + \frac{(\lambda^2-1)^2}{2\lambda^2}\right).$$

It turns out that this expression is independent of y (as it must be, since the translation along the axis is constant).

A Better Way: Using the Standard Formula

There is a well-known fact in hyperbolic geometry that when T is given by the dilation

$$T(z) = \lambda^2 z,$$

the translation length is

$$\ell(T) = 2\log\lambda.$$

We now verify that this agrees with the distance computed above.

Since

$$\cosh(2\log\lambda) = \frac{e^{2\log\lambda} + e^{-2\log\lambda}}{2} = \frac{\lambda^2 + \lambda^{-2}}{2},$$

we can write

$$2\log\lambda = 2\operatorname{arccosh}\left(\frac{\lambda^2 + \lambda^{-2}}{2}\right).$$

Step 3. Relating the Trace to λ and $\ell(T)$

Any hyperbolic isometry T (when represented in $\mathrm{PSL}(2,\mathbb{R})$) has a matrix representative conjugate to

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Its trace is

$$\mathrm{Tr}(T) = \lambda + \lambda^{-1}.$$

Observe that

$$\cosh\left(\frac{\ell(T)}{2}\right) = \cosh(\log\lambda) = \frac{\lambda + \lambda^{-1}}{2} = \frac{\mathrm{Tr}(T)}{2}.$$

Taking the inverse hyperbolic cosine of both sides, we have

$$\frac{\ell(T)}{2} = \operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

Multiplying by 2 gives

$$\ell(T) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

Step 4. Conclusion

Thus, by choosing a point on the axis of T (where the displacement is minimized) and computing the hyperbolic distance, we find that the minimal displacement of a hyperbolic isometry T is exactly

$$\ell(T) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

This shows that the formula for $\ell(T)$ indeed follows by computing the minimum displacement (i.e., the translation length) using the hyperbolic distance formula.

Summary

1. **Conjugation to a Dilation:** Every hyperbolic isometry is conjugate to a dilation $z \mapsto \lambda^2 z$.

2. **Minimal Displacement:** On the axis (the vertical line in the model), the displacement is constant and equals $2\log\lambda$.

3. **Trace Relation:** The matrix representing the dilation has trace $\lambda + \lambda^{-1}$, and one shows that

$$2\log\lambda = 2\operatorname{arccosh}\Big(\frac{\lambda + \lambda^{-1}}{2}\Big) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

4. **Conclusion:** Hence, the minimal displacement (translation length) is given by

$$\ell(T) = 2\,\text{arccosh}\Big(\frac{\text{Tr}(T)}{2}\Big).$$

This completes the proof.

We want to show that if

$$\ell(T) = \inf_{z\in\mathbb{H}^2} d_{\mathbb{H}^2}(z, T(z))$$

is the minimal displacement (translation length) of a hyperbolic isometry T, then

$$\ell(T) = 2\,\text{arccosh}\Big(\frac{\text{Tr}(T)}{2}\Big).$$

In other words, the expression

$$2\,\text{arccosh}\Big(\frac{\text{Tr}(T)}{2}\Big)$$

comes from computing the minimal distance that T moves a point in $\mathbb{H}^2$.

In what follows we give a detailed proof of this fact.

—

Step 1. Reduction to a Canonical Form

A hyperbolic isometry T has two fixed points on the boundary of $\mathbb{H}^2$ and leaves invariant a unique geodesic (called the axis of T). It is a standard fact that any hyperbolic isometry in $\text{PSL}(2,\mathbb{R})$ is conjugate to a dilation. That is, there exists a Möbius transformation M such that

$$M \circ T \circ M^{-1}(z) = \lambda^2\, z,$$

with $\lambda > 1$. (Sometimes one writes the dilation as $z \mapsto e^{\ell/2}z$; we will see below that the translation length is then $\ell = 2\log\lambda$.)

Since the hyperbolic metric is invariant under Möbius transformations, the translation length of T is the same as that of its conjugate. Hence, without loss of generality we may assume that

$$T(z) = \lambda^2 z,$$

with $\lambda > 1$.

Step 2. Compute the Minimal Displacement for a Dilation

The Hyperbolic Distance

Recall that in the upper half-plane model the hyperbolic distance between z and w is given by

$$d_{\mathbb{H}^2}(z, w) = \text{arccosh}\left(1 + \frac{|z-w|^2}{2\,\text{Im}(z)\,\text{Im}(w)}\right).$$

A key observation is that the dilation

$$T(z) = \lambda^2 z$$

has its invariant geodesic along the positive imaginary axis. (More generally, one can show that the axis of T is the unique geodesic joining 0 and ∞; in our case, that is the vertical line $\{x = 0,\ y > 0\}$.)

Thus, choose a point on the axis:

$$z = i\,y, \quad y > 0.$$

Then

$$T(i\,y) = \lambda^2(i\,y) = i\,\lambda^2 y.$$

Computing $d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y)$

For two points on the imaginary axis $z = i\,y$ and $w = i\,(\lambda^2 y)$, we have

- $|z - w| = |i\,y - i\,\lambda^2 y| = y|\lambda^2 - 1|$.
- $\mathrm{Im}(z) = y$ and $\mathrm{Im}(w) = \lambda^2 y$.

Thus,

$$d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y) = \operatorname{arccosh}\left(1 + \frac{(y|\lambda^2 - 1|)^2}{2\,y\,(\lambda^2 y)}\right) = \operatorname{arccosh}\left(1 + \frac{y^2(\lambda^2 - 1)^2}{2\lambda^2 y^2}\right).$$

Simplify the fraction:

$$\frac{y^2(\lambda^2 - 1)^2}{2\lambda^2 y^2} = \frac{(\lambda^2 - 1)^2}{2\lambda^2}.$$

Thus,

$$d_{\mathbb{H}^2}(i\,y,\, i\,\lambda^2 y) = \operatorname{arccosh}\left(1 + \frac{(\lambda^2 - 1)^2}{2\lambda^2}\right).$$

It turns out that this expression is independent of y (as it must be, since the translation along the axis is constant).

A Better Way: Using the Standard Formula

There is a well-known fact in hyperbolic geometry that when T is given by the dilation

$$T(z) = \lambda^2 z,$$

the translation length is

$$\ell(T) = 2\log\lambda.$$

We now verify that this agrees with the distance computed above.

Since

$$\cosh(2\log\lambda) = \frac{e^{2\log\lambda} + e^{-2\log\lambda}}{2} = \frac{\lambda^2 + \lambda^{-2}}{2},$$

we can write

$$2\log\lambda = 2\operatorname{arccosh}\left(\frac{\lambda^2 + \lambda^{-2}}{2}\right).$$

Step 3. Relating the Trace to λ and $\ell(T)$

Any hyperbolic isometry T (when represented in $\mathrm{PSL}(2,\mathbb{R})$) has a matrix representative conjugate to

$$\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}.$$

Its trace is

$$\mathrm{Tr}(T) = \lambda + \lambda^{-1}.$$

Observe that

$$\cosh\Big(\frac{\ell(T)}{2}\Big) = \cosh(\log\lambda) = \frac{\lambda + \lambda^{-1}}{2} = \frac{\mathrm{Tr}(T)}{2}.$$

Taking the inverse hyperbolic cosine of both sides, we have

$$\frac{\ell(T)}{2} = \operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

Multiplying by 2 gives

$$\ell(T) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

Step 4. Conclusion

Thus, by choosing a point on the axis of T (where the displacement is minimized) and computing the hyperbolic distance, we find that the minimal displacement of a hyperbolic isometry T is exactly

$$\ell(T) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

This shows that the formula for $\ell(T)$ indeed follows by computing the minimum displacement (i.e., the translation length) using the hyperbolic distance formula.

Summary

1. **Conjugation to a Dilation:** Every hyperbolic isometry is conjugate to a dilation $z \mapsto \lambda^2 z$.

2. **Minimal Displacement:** On the axis (the vertical line in the model), the displacement is constant and equals $2\log\lambda$.

3. **Trace Relation:** The matrix representing the dilation has trace $\lambda + \lambda^{-1}$, and one shows that

$$2\log\lambda = 2\operatorname{arccosh}\Big(\frac{\lambda+\lambda^{-1}}{2}\Big) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

4. **Conclusion:** Hence, the minimal displacement (translation length) is given by

$$\ell(T) = 2\operatorname{arccosh}\Big(\frac{\mathrm{Tr}(T)}{2}\Big).$$

This completes the proof.

Displacement Length of Elliptic Elements

In hyperbolic geometry, the displacement (or translational) length of an isometry T is defined as

$$\ell(T) = \inf_{z \in \mathbb{H}} d(z, T(z)).$$

If T is an elliptic isometry, it has a fixed point $z_0 \in \mathbb{H}$ such that $T(z_0) = z_0$. Consequently,

$$d(z_0, T(z_0)) = d(z_0, z_0) = 0,$$

which implies that

$$\ell(T) = 0.$$

It's also worth noting that while elliptic (and parabolic) isometries have $\ell(T) = 0$, hyperbolic isometries have a positive translation length, with the infimum being achieved on a unique geodesic called the axis of the isometry.

Furthermore, we can prove that the expression

$$2 \operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right)$$

is purely imaginary when T is **elliptic** and is related to the rotation angle of T.

Step 1: Understanding the Trace Condition for Elliptic Isometries An isometry T of the hyperbolic plane $\mathbb{H}$ in $\mathrm{PSL}(2, \mathbb{R})$ is elliptic if and only if

$$|\operatorname{Tr}(T)| < 2.$$

Since T is an element of $\mathrm{PSL}(2, \mathbb{R})$, it can be written in matrix form as

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{with} \quad ad - bc = 1.$$

Yes, that is correct. When working in $\mathrm{PSL}(2, \mathbb{R})$** rather than **$\mathrm{SL}(2, \mathbb{R})$**, the absolute value around $\operatorname{Tr}(T)$ is not necessary.

Why?

1. In $\mathrm{SL}(2, \mathbb{R})$, matrices have determinant 1, but we can have **both** T and $-T$ representing the same Möbius transformation. That is, both:

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and

$$-T = \begin{bmatrix} -a & -b \\ -c & -d \end{bmatrix}$$

induce the same transformation on $\mathbb{H}^2$. However, their traces satisfy:

$$\operatorname{Tr}(-T) = -\operatorname{Tr}(T).$$

Since both matrices define the same transformation, we need to ensure that the trace does not change sign when switching representations. Therefore, in $\mathrm{SL}(2,\mathbb{R})$, we use:

$$|\operatorname{Tr}(T)|$$

to account for this ambiguity.

2. In $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm I\}$, we identify T and $-T$ as the same element. This means the trace ambiguity is naturally resolved, and we can simply write:

$$\ell(T) = 2 \operatorname{arccosh}\left(\frac{\operatorname{Tr}(T)}{2}\right).$$

without needing the absolute value.

- **In** $\mathrm{SL}(2,\mathbb{R})$**, we use** $|\operatorname{Tr}(T)|$ **because** T **and** $-T$ **have opposite traces.**
- **In** $\mathrm{PSL}(2,\mathbb{R})$**, we do not need absolute values, since the trace is well-defined up to sign.**
- Thus, when working purely in $\mathrm{PSL}(2,\mathbb{R})$, you can safely use:

$$\ell(T) = 2 \operatorname{arccosh}\left(\frac{\operatorname{Tr}(T)}{2}\right).$$

The trace is given by:

$$\operatorname{Tr}(T) = a + d.$$

If T is elliptic, its action on $\mathbb{H}$ is a **rotation** about a fixed point in $\mathbb{H}$. The characteristic equation for T satisfies:

$$\lambda^2 - \operatorname{Tr}(T)\lambda + 1 = 0.$$

Solving for λ, the eigenvalues of T are:

$$\lambda = \frac{\operatorname{Tr}(T) \pm \sqrt{\operatorname{Tr}(T)^2 - 4}}{2}.$$

Since $|\operatorname{Tr}(T)| < 2$, the quantity under the square root is negative:

$$\operatorname{Tr}(T)^2 - 4 < 0.$$

Thus, the eigenvalues of T are **complex** and can be written as:

$$\lambda = e^{\pm i\theta},$$

for some real θ, where θ is the **rotation angle** of T.

Step 2: Evaluating arccosh By definition, the inverse hyperbolic cosine is:

$$\operatorname{arccosh}(x) = \ln\left(x + \sqrt{x^2 - 1}\right), \quad \text{for } x \geq 1.$$

Applying this to our case, we set:

$$x = \frac{|\operatorname{Tr}(T)|}{2}.$$

Since $|\operatorname{Tr}(T)| < 2$, we have:

$$-1 < \frac{\operatorname{Tr}(T)}{2} < 1.$$

Now, recall the identity:

$$\operatorname{arccosh}(x) = i \arccos(x), \quad \text{for } -1 \leq x \leq 1.$$

Thus, since $-1 < \frac{\operatorname{Tr}(T)}{2} < 1$, we obtain:

$$\operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right) = i \arccos\left(\frac{|\operatorname{Tr}(T)|}{2}\right).$$

Multiplying by 2,

$$2 \operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right) = 2i \arccos\left(\frac{|\operatorname{Tr}(T)|}{2}\right).$$

Step 3: Connection to the Rotation Angle From the eigenvalues of T, we identified the rotation angle θ as:

$$\theta = \arccos\left(\frac{\operatorname{Tr}(T)}{2}\right).$$

Comparing with our previous result,

$$2 \operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right) = 2i\theta.$$

Thus, the expression is **purely imaginary**, and the imaginary part corresponds to **twice the rotation angle** of T.

Conclusion For an **elliptic** isometry T:

$$2 \operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right) = 2i\theta,$$

where θ is the rotation angle of T. This confirms that the formula, when applied to elliptic isometries, yields an **imaginary** value rather than a real translation length, and its imaginary component is directly related to the angle of rotation.

1.2 Elliptic Elements in Fuchsian Groups are Not Primitive

Below is a concise, self-contained proof that *elliptic elements* in a Fuchsian group (or more generally, in $\mathrm{PSL}(2, \mathbb{R})$) cannot be primitive, under the usual definition of "primitive element."

1.2.1 Definitions and Setup

- **Elliptic Element**: In $\mathrm{PSL}(2,\mathbb{R})$, an element γ is called *elliptic* if it has finite order $n > 1$. Equivalently, γ fixes a point in $\mathbb{H}^2$ and acts as a rotation of finite angle around that point. Concretely, $\gamma^n = \mathrm{Id}$ for some integer $n > 1$.
- **Primitive Element**: An element $\gamma \in \Gamma$ (where $\Gamma \subseteq \mathrm{PSL}(2,\mathbb{R})$) is said to be *primitive* if it is *not* a proper power of any other element in Γ. Formally, γ is primitive if there do not exist $\gamma_0 \in \Gamma$ and an integer $k > 1$ such that

$$\gamma = \gamma_0^k.$$

Geometric Interpretation:

- **Hyperbolic elements** in $\mathrm{PSL}(2,\mathbb{R})$ have infinite order and correspond to closed geodesics in the quotient manifold $\Gamma\backslash\mathbb{H}^2$. A hyperbolic element γ is *primitive* if the corresponding closed geodesic is not just a multiple cover of a shorter geodesic.
- **Elliptic elements**, on the other hand, have finite order and do not produce a "usual" closed geodesic in $\Gamma\backslash\mathbb{H}^2$. Instead, they correspond to orbifold points (cone points) or rotational symmetries.

1.2.2 Statement

Claim: Elliptic elements are not primitive.

In other words, if γ is elliptic of finite order $n > 1$, then γ can be expressed as a nontrivial power of another element in the group. Hence, by definition, γ is not primitive.

1.2.3 Proof

1. **Finite Order:** Since γ is elliptic, there is an integer $n > 1$ such that

$$\gamma^n = \mathrm{Id}.$$

   That is, γ has order n.

2. **Constructing a Root:** We claim that γ is itself a power of some element $\delta \in \Gamma$. Indeed, pick m such that $\gcd(m,n) = 1$. By Bézout's identity, there exist integers x, y such that

$$mx + ny = 1.$$

   Define $\delta := \gamma^m$. Then

$$\delta^x = (\gamma^m)^x = \gamma^{mx}, \quad \gamma^{ny} = \mathrm{Id}^y = \mathrm{Id}.$$

   Multiplying, we get

$$\delta^x \gamma^{ny} = \gamma^{mx+ny} = \gamma.$$

   But $\gamma^{ny} = \mathrm{Id}$, so

$$\delta^x = \gamma.$$

   Since $x \neq \pm 1$ (indeed, $|x| \geq 2$ or $|y| \geq 1$ in typical cases where $n > 1$), we see that γ is expressed as a nontrivial power δ^x. Hence, γ is *not* primitive.

3. **Conclusion:** Because every elliptic element γ can be written as a proper power δ^k for some $k > 1$, it follows that no elliptic element is primitive. Thus, elliptic elements fail the "no proper power" condition that defines primitiveness.

$$\boxed{\text{Therefore, elliptic elements are not primitive.}}$$

1.2.4 Geometric Interpretation

From a geometric perspective, an elliptic element γ means it acts on the hyperbolic plane $\mathbb{H}^2$ by rotation around a fixed point. This element does *not* give rise to a closed geodesic of positive length in the quotient manifold, unlike a hyperbolic element that has infinite order and translates along an axis. The concept of "primitive" is primarily meaningful for hyperbolic elements with infinite order. Elliptic elements—having finite order—never qualify as primitive because:

- They do not produce "genuine" closed geodesics in the usual sense.
- Algebraically, they *are* finite-order elements that can be expressed as nontrivial powers of other elements (as shown in the proof).

Hence, "primitive" in geometric group theory typically refers to infinite-order elements that cannot be expressed as a power of another infinite-order element, something that fails automatically for finite-order (elliptic) elements.

$$\boxed{\text{Elliptic elements are finite-order, thus always expressible as proper powers, so they're not primitive.}}$$

Displacement Length of Parabolic Elements

If T is a parabolic isometry of the hyperbolic plane $\mathbb{H}$, then its displacement length $\ell(T)$ is also 0.

Explanation: By definition, the displacement length of an isometry T is given by:

$$\ell(T) = \inf_{z \in \mathbb{H}} d(z, T(z)).$$

A parabolic isometry has exactly one fixed point on the boundary $\partial\mathbb{H}$ at infinity and does not have any fixed points in $\mathbb{H}$. It acts as a "shear" along horocycles centered at the fixed point.

For any point $z \in \mathbb{H}$, the hyperbolic distance $d(z, T(z))$ is strictly positive, but there exist sequences z_n in $\mathbb{H}$ that approach the parabolic fixed point at infinity for which $d(z_n, T(z_n)) \to 0$. This ensures that the infimum in the definition of $\ell(T)$ is attained at 0.

Alternatively, we can still apply the formula

$$2 \operatorname{arccosh}\left(\frac{|\operatorname{Tr}(T)|}{2}\right)$$

even when T is **parabolic**. However, we must carefully consider its evaluation.

Case: T is Parabolic ($|\operatorname{Tr}(T)| = 2$) By definition, a parabolic isometry satisfies:

$$|\operatorname{Tr}(T)| = 2.$$

Substituting this into the formula,

$$2 \operatorname{arccosh}\left(\frac{2}{2}\right) = 2 \operatorname{arccosh}(1).$$

Since we know that

$$\operatorname{arccosh}(1) = 0,$$

we obtain:

$$2 \operatorname{arccosh}(1) = 2 \times 0 = 0.$$

Interpretation - The formula correctly gives 0, which aligns with the fact that **parabolic isometries have zero translation length $\ell(T) = 0$**.

- Thus, even though parabolic isometries behave differently from hyperbolic ones, the formula still holds in the limiting sense.

- For hyperbolic isometries, $|\operatorname{Tr}(T)| > 2$, and the formula gives a **positive** translation length.

- For elliptic isometries, $|\operatorname{Tr}(T)| < 2$, the argument of arccosh is **less than 1**, which is undefined in the real domain, indicating a purely imaginary result instead of a real translation length.

Conclusion: Thus, for a parabolic isometry T, we conclude that:

$$\ell(T) = 0.$$

This result aligns with the classification of isometries in hyperbolic geometry:

- **Elliptic**: $\ell(T) = 0$ (has a fixed point in $\mathbb{H}$).
- **Parabolic**: $\ell(T) = 0$ (has exactly one fixed point at infinity).
- **Hyperbolic**: $\ell(T) > 0$ (has an invariant geodesic and translates points along it).

Conclusion: To summarize, we define:

- A *closed geodesic* on X is a geodesic that is periodic.
- It is *simple* if it has no self-intersections.
- It is *primitive* if it is not obtained by iterating a shorter closed geodesic. Equivalently, in the group Γ, an element is primitive if it is not a proper power.

In the standard case where Γ is torsion-free, every nontrivial element is hyperbolic, so every closed geodesic arises from a hyperbolic element whose translation length equals the length of the geodesic. Parabolic and elliptic elements, by contrast, do not yield closed geodesics in the usual sense because they lack a finite translation length along a unique invariant geodesic.

1.3 Parabolic Elements in a Fuchsian Group Do Not Produce Closed Geodesics and Are Not Primitive

Below is a concise, self-contained explanation of why parabolic elements in a Fuchsian group (or, more generally, in $\mathrm{PSL}(2,\mathbb{R})$) do *not* produce simple closed geodesics in the quotient manifold $\Gamma\backslash\mathbb{H}^2$. In short, parabolic transformations act by "sliding" points along horocycles around a single boundary fixed point, and thus they do *not* have an axis in $\mathbb{H}^2$. Consequently, they cannot produce closed geodesics of positive length [7, Section 7.2, 12.1].

1.3.1 Definitions and Setup

1. **Fuchsian Group**: A discrete subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ acting on the hyperbolic plane $\mathbb{H}^2$. The quotient
$$X := \Gamma\backslash\mathbb{H}^2$$
(or orbifold) is typically a hyperbolic surface (or orbifold) of finite volume if Γ is cofinite [7, Chapter 7].

2. **Parabolic Element**: In $\mathrm{PSL}(2,\mathbb{R})$, an element γ is called *parabolic* if it has infinite order but *exactly one* fixed point on the boundary $\partial\mathbb{H}^2$. Concretely, γ can be conjugated into a matrix of the form
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$
which fixes the boundary point ∞ [7, Section 7.2].

3. **Closed Geodesic in $\Gamma\backslash\mathbb{H}^2$**: A *closed geodesic* is a loop that is geodesic in each homotopy class. It arises precisely from a *hyperbolic* element of Γ, which has two boundary fixed points and an *axis* in $\mathbb{H}^2$. This axis projects to a loop of positive length in X [7, Section 12.1].

Key Point : Parabolic elements have no axis in $\mathbb{H}^2$ and therefore cannot correspond to closed geodesics in the quotient manifold.

1.3.2 Geometric Reason: Single Boundary Fixed Point, No Axis

1. **Single Fixed Point on Boundary**: Since γ is parabolic, it fixes exactly one point $\xi \in \partial\mathbb{H}^2$. For instance, in the upper half-plane model, one can assume $\gamma(\infty) = \infty$.

2. **Horocycles Instead of Axes**: The isometry γ does *not* have a geodesic in the interior of $\mathbb{H}^2$ that is stabilized by γ. Instead, it slides points along a family of **horocycles** centered at the boundary point ξ. In the upper half-plane model, with $\xi = \infty$, these horocycles look like horizontal lines $y = \text{constant}$.

3. **No Closed Loop**: Because γ merely translates along these horocycles, it does not create a loop in $\Gamma\backslash\mathbb{H}^2$. The corresponding region in the quotient is a *cusp*, typically represented by an end of the surface that is topologically an infinite funnel or a cylinder of zero length in the limit. Thus, γ yields a *cusp* in the quotient, *not* a closed geodesic [7, Section 12.1].

1.3.3 Parabolic Elements Are Not Primitive

A parabolic element γ is *not primitive* because it can be expressed as a nontrivial power of another parabolic element [7, Section 7.2]. Explicitly, if γ is conjugate to

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

then it can be written as

$$\gamma = \delta^n, \quad \text{where } \delta = \begin{pmatrix} 1 & 1/n \\ 0 & 1 \end{pmatrix},$$

for some integer $n > 1$. Since γ is a proper power of δ, it fails the definition of primitiveness.

Parabolic elements are not primitive because they are always expressible as proper powers.

1.4 Counting Closed Geodesics on Hyperbolic Surfaces

In number theory, the Prime Number Theorem tells us that the number of prime numbers up to N, denoted by $\pi(N)$, satisfies

$$\pi(N) \sim \frac{N}{\log N} \quad \text{as } N \to \infty.$$

In analogy, in the geometric setting of hyperbolic surfaces we consider a counting problem for closed geodesics.

Primitive Closed Geodesics and Translation Length

Let X be a closed, connected, oriented hyperbolic surface. (Recall that a hyperbolic surface is a two-dimensional Riemannian manifold whose metric has constant curvature -1.) Every closed geodesic on X can be viewed as the projection to X of the axis of a hyperbolic element in the fundamental group Γ of X.

More precisely, one may represent X as

$$X = \Gamma \backslash \mathbb{H}^2,$$

where Γ is a discrete, torsion-free subgroup of $\mathrm{PSL}(2, \mathbb{R})$. A nontrivial element $\gamma \in \Gamma$ is *hyperbolic* if

$$|\operatorname{tr}(\gamma)| > 2.$$

For such a hyperbolic element, one can define its *translation length* by

$$\ell(\gamma) = \inf_{z \in \mathbb{H}^2} d(z, \gamma z).$$

It can be shown (by various methods, e.g., via eigenvalues or the cross-ratio) that

$$\ell(\gamma) = 2 \cosh^{-1}\Big(\frac{|\operatorname{tr}(\gamma)|}{2}\Big).$$

This translation length is exactly the length of the closed geodesic on X corresponding to γ.

A closed geodesic is called *primitive* if it is not obtained by repeatedly traversing a shorter geodesic. Equivalently, a hyperbolic element $\gamma \in \Gamma$ is called primitive if it is not a proper power (i.e. there is no $\gamma_0 \in \Gamma$ and an integer $n \geq 2$ with $\gamma = \gamma_0^n$). Thus, the length L in our counting problem is exactly the upper bound on the translation lengths of the primitive hyperbolic elements of Γ.

The Prime Geodesic Theorem

Given a closed, oriented hyperbolic surface X and a positive real number L, let

$$c(X, L)$$

denote the number of (non-oriented) *primitive* closed geodesics on X whose lengths are at most L. The *Prime Geodesic Theorem* is the geometric analogue of the Prime Number Theorem. It states that, for any such surface X,

$$c(X, L) \sim \frac{e^L}{2L} \quad \text{as } L \to \infty.$$

That is, the number of primitive closed geodesics grows exponentially with rate e^L (and a logarithmic correction factor $1/(2L)$), and remarkably, the asymptotic behavior is universal—it does not depend on the finer geometric or topological details of X.

How L Relates to Translation Length: Recall that for a hyperbolic element $\gamma \in \Gamma$, the translation length $\ell(\gamma)$ is the distance by which γ moves any point on its axis in $\mathbb{H}^2$. The closed geodesic on X corresponding to γ has length exactly $\ell(\gamma)$. In our counting function $c(X, L)$, we count only those primitive elements for which $\ell(\gamma) \leq L$. Thus, the parameter L directly represents an upper bound on the length (or translation length) of the geodesics we are counting.

A Numerical Example

Consider, for example, the modular surface

$$X = \mathrm{PSL}(2, \mathbb{Z}) \backslash \mathbb{H}^2.$$

(Strictly speaking, X is an orbifold, but similar asymptotic formulas hold.) For large L, the Prime Geodesic Theorem tells us that the number of primitive closed geodesics with length at most L is approximately

$$c(X, L) \approx \frac{e^L}{2L}.$$

For instance, if $L = 10$, then one would expect

$$c(X, 10) \approx \frac{e^{10}}{20} \approx \frac{22026.5}{20} \approx 1101.$$

While the precise count for the modular surface involves many subtleties, this numerical estimate gives a rough idea of the exponential growth.

An Outline of a Proof of the Prime Geodesic Theorem

A full proof of the Prime Geodesic Theorem is quite advanced and relies on a blend of spectral theory, analytic number theory, and the theory of automorphic forms. Below is an outline of the main ideas involved:

1. **Selberg Trace Formula:** One of the key tools is the Selberg trace formula, which is an identity that relates spectral data of the Laplacian on X (eigenvalues and eigenfunctions) to geometric data (lengths of closed geodesics). The trace formula can be viewed as an analogue of the explicit formula in prime number theory.

2. **Spectral Analysis:** By analyzing the contributions of the closed geodesics in the trace formula, one can isolate an "error term" and identify the main term that governs the asymptotic behavior of $c(X, L)$. The hyperbolic terms in the trace formula correspond precisely to the primitive closed geodesics.

3. **Asymptotic Analysis:** One shows that the main term in the trace formula leads to an expression of the form
$$c(X, L) \sim \frac{e^L}{2L},$$
by employing techniques analogous to those used in the proof of the Prime Number Theorem.

4. **Uniformity and Independence:** A key aspect is that the asymptotic formula is independent of the particular geometry of X (beyond its hyperbolic structure). This uniformity is a consequence of the invariance properties of the trace formula and the deep ergodic properties of the geodesic flow.

While a complete proof requires many pages of technical details and advanced tools, this outline captures the main strategy: the Selberg trace formula connects spectral theory with the geometry of X, and a careful asymptotic analysis of its hyperbolic terms yields the exponential growth law of the primitive closed geodesics.

Summary

To summarize:

- For a hyperbolic surface X, the closed geodesics correspond to hyperbolic elements in the fundamental group Γ, and the translation length of such an element equals the length of the associated geodesic.

- A closed geodesic is *primitive* if it is not a multiple traversal of a shorter geodesic.

- The Prime Geodesic Theorem states that the number $c(X, L)$ of primitive closed geodesics of length at most L satisfies
$$c(X, L) \sim \frac{e^L}{2L} \quad \text{as } L \to \infty.$$

- Numerically, for instance, if $L = 10$ on the modular surface, one obtains $c(X, 10) \approx 1101$.
- The proof of the theorem is based on the Selberg trace formula, which relates spectral data of the Laplacian on X to the geometric data of closed geodesics.

This result is remarkable because it shows that, despite the potentially complicated geometry and topology of X, the asymptotic behavior of primitive closed geodesics is universal and depends only on the hyperbolic structure.

Simple Closed Geodesics and Mirzakhani's Breakthrough

While the prime geodesic theorem concerns all primitive closed geodesics, one may further ask: What if we restrict our attention to those closed geodesics that are *simple*? A closed geodesic is called *simple* if it does not intersect itself. Let $s(X, L)$ denote the number of simple closed geodesics on X of length at most L. At first glance, one might expect that the asymptotic behavior of $s(X, L)$ should be similar to that of $c(X, L)$; however, the situation turns out to be considerably more subtle.

For many years, the asymptotic behavior of $s(X, L)$ remained an open problem because traditional analytic techniques (such as those used by Huber and Selberg) did not distinguish between simple and non-simple geodesics. The breakthrough came with the work of Maryam Mirzakhani. In her 2004 Ph.D. thesis [6], she proved the following remarkable theorem:

Theorem 1.1 (Mirzakhani, Theorem 1.1 [6]). *Let X be a closed, connected, oriented hyperbolic surface of genus $g \geq 2$. Then, there exists a positive constant $s(X)$ (depending on X) such that*

$$s(X, L) \sim s(X) \cdot L^{6g-6} \quad \text{as } L \to \infty.$$

This result is striking for several reasons:

- The polynomial growth rate L^{6g-6} depends solely on the topology of X (through the genus g) and not on its specific hyperbolic geometry.
- Mirzakhani's proof uses techniques from ergodic theory, together with her celebrated formulas for Weil–Petersson volumes and integration over moduli space.

Mirzakhani's work has had far-reaching implications across various areas of mathematics. As an example of a concrete consequence of her work, consider the following result:

Theorem 1.2 (Mirzakhani, Corollary 1.4). *On any closed, oriented hyperbolic surface of genus 2, a random long simple closed geodesic is 48 times more likely to be non-separating than separating.*

Goals and Organization of the Survey

The primary aim of this survey is to provide a detailed and accessible account of Mirzakhani's proof of Theorem 1.1, making the advanced ideas approachable even for undergraduates. In doing so, we draw inspiration from classic results on counting lattice points in homogeneous spaces—a subject with deep connections to both number theory and ergodic theory.

The survey is organized as follows:

- In Section 2, we study counting problems for primitive lattice points in the Euclidean plane. This discussion serves as a motivation and provides a simpler model for the more complicated geometric problems that follow.
- In Section 3, we review the background on hyperbolic surfaces, Teichmüller spaces, and simple closed curves. Definitions, examples, and basic properties will be provided to ensure that the reader is well-prepared for later sections.
- In Section 4, we discuss Mirzakhani's celebrated formulas for the Weil–Petersson volumes of moduli spaces and her integration formulas over moduli space, which play a central role in her proof.
- In Section 5, we present a complete proof of Theorem 1.1.
- In Section 6, we briefly survey several subsequent counting results for closed curves on surfaces and other related topics that have emerged since Mirzakhani's groundbreaking work.

Throughout this survey, we assume some familiarity with the basic concepts of hyperbolic geometry and Riemann surfaces, but we will strive to provide all necessary definitions and examples. Our hope is that even students from universities with a less advanced background in these topics will find the exposition clear and enlightening.

2 Counting Primitive Lattice Points in the Euclidean Plane

Outline and Motivation

In this section we study a classical counting problem in the Euclidean plane: How many *primitive* lattice points lie inside a large ball? (A primitive lattice point is one that, in analogy with prime numbers, cannot be obtained by scaling another lattice point by a nontrivial integer.) This counting problem not only is interesting in its own right but also serves as an accessible model for more advanced counting problems on hyperbolic surfaces (which we will discuss later).

Basic Definitions and Examples

Lattice in $\mathbb{R}^2$: A *lattice* Λ in $\mathbb{R}^2$ is a discrete subgroup of $\mathbb{R}^2$ that spans the entire space. The standard example is

$$\mathbb{Z}^2 = \{(a, b) \in \mathbb{R}^2 : a, b \in \mathbb{Z}\}.$$

A lattice can be thought of as the set of all integer linear combinations of two linearly independent vectors in $\mathbb{R}^2$.

Primitive Lattice Point: A vector $v \in \mathbb{Z}^2$ is said to be *primitive* if it cannot be written as an integer multiple (with integer factor greater than 1) of another vector in $\mathbb{Z}^2$. Equivalently, if

$$v = (a, b) \quad \text{with } a, b \in \mathbb{Z},$$

then v is primitive if and only if $\gcd(a, b) = 1$. We denote the set of all primitive lattice points by

$$\mathbb{Z}^2_{\text{prim}} = \{v \in \mathbb{Z}^2 : \gcd(v) = 1\}.$$

Example: The vector $(3, 5)$ is primitive because $\gcd(3, 5) = 1$; however, the vector $(4, 6)$ is not primitive because $\gcd(4, 6) = 2$ (indeed, $(4, 6) = 2\,(2, 3)$ and $(2, 3)$ is primitive).

Euclidean Norm: For any $x = (x_1, x_2) \in \mathbb{R}^2$, the Euclidean norm is defined by

$$\|x\| = \sqrt{x_1^2 + x_2^2}.$$

Counting Function: For each $L > 0$, we define the counting function

$$p(\mathbb{Z}^2, L) := \#\{v \in \mathbb{Z}^2_{\text{prim}} : \|v\| \le L\},$$

which counts the number of primitive lattice points inside (or on) the circle (or ball) of radius L centered at the origin.

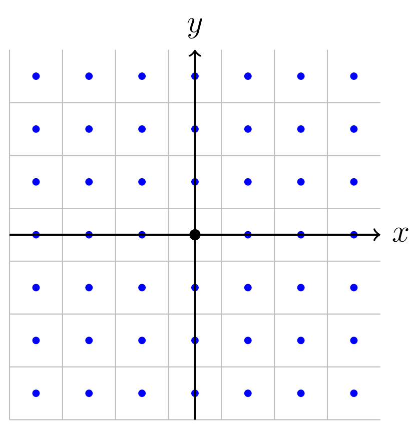

Figure 1: The integer lattice $\mathbb{Z}^2$ in $\mathbb{R}^2$. The blue dots are the lattice points; note that, for instance, $(1, 2)$ is primitive whereas $(2, 4)$ is not.

The Asymptotic Counting Problem

Much like the prime number theorem concerns the asymptotic behavior of $\pi(N)$ (the number of primes up to N), we are interested in the asymptotic behavior of $p(\mathbb{Z}^2, L)$ as $L \to \infty$. The main result we wish to discuss is the following:

Theorem 2.1 (2.1)**.** *As* $L \to \infty$*,*

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2.$$

This means that

$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi}.$$

Intuitive Explanation: The density of lattice points in $\mathbb{R}^2$ is roughly uniform. In a circle of radius L, there are about πL^2 lattice points. However, many of these are not primitive (for instance, every even point is a multiple of a smaller lattice point). A classical result in number theory shows that the probability that two randomly chosen integers are coprime is $\frac{6}{\pi^2}$. Thus, one expects that approximately a $\frac{6}{\pi^2}$ fraction of the πL^2 lattice points are primitive. This yields an expected count of

$$\frac{6}{\pi^2} \cdot \pi L^2 = \frac{6}{\pi} L^2.$$

Our goal is to rigorously justify this asymptotic formula.

A Measure-Theoretic Approach

A powerful method to study such counting problems is via measures. For each $L > 0$, define a *counting measure* on $\mathbb{R}^2$ associated with primitive lattice points:

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{\frac{1}{L}v},$$

where δ_x denotes the Dirac measure at x.

Observation: If $B \subset \mathbb{R}^2$ is the unit ball centered at the origin (i.e. $B = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$), then

$$\nu_L^{\text{prim}}(B) = \frac{1}{L^2} \#\{v \in \mathbb{Z}^2_{\text{prim}} : \|v\| \leq L\} = \frac{p(\mathbb{Z}^2, L)}{L^2}.$$

Thus, studying the weak-$*$ limit of ν_L^{prim} as $L \to \infty$ is equivalent to understanding the asymptotic behavior of $p(\mathbb{Z}^2, L)/L^2$.

Weak-$*$ Convergence: It is expected (and can be proved) that

$$\lim_{L\to\infty} \nu_L^{\text{prim}} = \frac{6}{\pi^2} \nu,$$

where ν is the standard Lebesgue measure on $\mathbb{R}^2$. In other words, for any continuous function f with compact support,

$$\lim_{L\to\infty} \int_{\mathbb{R}^2} f \, d\nu_L^{\text{prim}} = \frac{6}{\pi^2} \int_{\mathbb{R}^2} f \, d\nu.$$

Taking f to be the indicator function of the unit ball (after approximating by continuous functions) gives

$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi^2} \nu(B) = \frac{6}{\pi^2} \cdot \pi = \frac{6}{\pi}.$$

A Related Counting Measure:

For comparison, we also define the counting measure on the full lattice $\mathbb{Z}^2$ by

$$\nu_L := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2} \delta_{\frac{v}{L}}. \tag{2.1}$$

By a similar scaling argument, one can show that

$$\lim_{L \to \infty} \nu_L = \nu$$

in the weak-$*$ sense. This means that the normalized counting measure on the full lattice converges to Lebesgue measure. (For a rigorous treatment, see Exercise **??**.) The primitive counting measure ν_L^{prim} is then obtained by restricting ν_L to the subset $\mathbb{Z}^2_{\text{prim}}$ and the result follows from the fact that the density of primitive points in $\mathbb{Z}^2$ is $\frac{6}{\pi^2}$.

Conclusion and Theorem Statement

Summarizing the discussion, we have shown (or at least motivated rigorously) that the number of primitive lattice points in $\mathbb{Z}^2$ of Euclidean norm at most L satisfies

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2,$$

or equivalently,

$$\lim_{L \to \infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi}.$$

Theorem 2.2 (2.1). *As $L \to \infty$,*

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2.$$

That is,

$$\lim_{L \to \infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi}.$$

Motivation and Future Directions: The techniques used in this section—especially the measure-theoretic approach and the idea of weak-$*$ convergence—serve as a model for more complex counting problems. Later (in Section 5), we will see how analogous ideas are applied to count simple closed geodesics on hyperbolic surfaces. In that setting, many of the underlying ideas from the geometry of numbers (e.g., scaling, density arguments, and ergodic theory) reappear in a rich geometric and dynamical context.

Additional Exercises

Exercise 2.3 (2.2). What would happen with the weak-$*$ convergence in (**??**) if, in the definition of the counting measures ν_L in (2.2), one replaced $\mathbb{Z}^2$ by a finite index subgroup of $\mathbb{Z}^2$?

In other words, what would happen with (**??**) if we considered a finite index subgroup of $\mathbb{Z}^2$ instead of all $\mathbb{Z}^2$ in the definition of the counting measures $(\nu_L)_{L>0}$ in (2.2)?

Exercise Statement

Recall that for the full lattice $\mathbb{Z}^2$ the counting measures are defined by

$$\nu_L := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2} \delta_{\frac{1}{L}v},$$

and one has

$$\lim_{L \to \infty} \nu_L = \nu$$

(with respect to the weak-$*$ topology), where ν is the Lebesgue measure on $\mathbb{R}^2$.

The exercise asks: *What would happen with ν_L if we considered a finite-index subgroup $\Gamma \subset \mathbb{Z}^2$ instead of $\mathbb{Z}^2$ in the definition of the counting measures?*

Solution

Let $\Gamma \subset \mathbb{Z}^2$ be a finite-index subgroup. Denote the index by

$$[\mathbb{Z}^2 : \Gamma] = m < \infty.$$

We define the corresponding counting measures by

$$\nu_L^\Gamma := \frac{1}{L^2} \sum_{v \in \Gamma} \delta_{\frac{1}{L}v}.$$

We wish to show that

$$\lim_{L \to \infty} \nu_L^\Gamma = \frac{1}{m} \nu,$$

in the weak-$*$ topology. In other words, for every continuous, compactly supported function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\lim_{L \to \infty} \int_{\mathbb{R}^2} f(x)\, d\nu_L^\Gamma(x) = \frac{1}{m} \int_{\mathbb{R}^2} f(x)\, d\nu(x).$$

Step 1. Partitioning $\mathbb{Z}^2$ via Cosets of Γ

Since Γ has finite index m in $\mathbb{Z}^2$, we can write

$$\mathbb{Z}^2 = \Gamma \cup (\gamma_2 + \Gamma) \cup \cdots \cup (\gamma_m + \Gamma),$$

where $\gamma_1 = 0$ and $\gamma_2, \ldots, \gamma_m$ are representatives of the distinct cosets. Therefore, any sum over $\mathbb{Z}^2$ can be broken into a sum over the cosets.

Step 2. Relating ν_L^Γ to ν_L

Recall that for the full lattice,

$$\nu_L = \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2} \delta_{\frac{1}{L} v}.$$

Using the coset decomposition, we have

$$\nu_L = \frac{1}{L^2} \sum_{k=1}^{m} \sum_{v \in \Gamma} \delta_{\frac{1}{L}(v+\gamma_k)}.$$

Notice that translation by a fixed vector is a continuous operation on measures. In fact, if we denote by T_y the translation operator defined by $T_y(x) = x + y$, then

$$\delta_{\frac{1}{L}(v+\gamma_k)} = T_{\gamma_k/L}(\delta_{v/L}).$$

Thus, we can rewrite

$$\nu_L = \sum_{k=1}^{m} T_{\gamma_k/L} \left(\frac{1}{L^2} \sum_{v \in \Gamma} \delta_{v/L} \right) = \sum_{k=1}^{m} T_{\gamma_k/L}\big(\nu_L^\Gamma\big).$$

Step 3. Passing to the Limit

We know that $\nu_L \to \nu$ as $L \to \infty$ and note that for each fixed k the translation $T_{\gamma_k/L}$ tends to the identity as $L \to \infty$ (since $\gamma_k/L \to 0$). In the weak-$*$ topology, translations by vectors converging to 0 do not change the limit. Therefore,

$$\lim_{L\to\infty} T_{\gamma_k/L}\big(\nu_L^\Gamma\big) = \lim_{L\to\infty} \nu_L^\Gamma.$$

Since there are m such terms, we obtain

$$\lim_{L\to\infty} \nu_L = \ m \cdot \lim_{L\to\infty} \nu_L^\Gamma.$$

But the left-hand side is known to converge to ν. Therefore, we must have

$$m \cdot \lim_{L\to\infty} \nu_L^\Gamma = \nu,$$

which implies

$$\lim_{L\to\infty} \nu_L^\Gamma = \frac{1}{m}\, \nu.$$

Conclusion

Thus, if we replace $\mathbb{Z}^2$ by a finite-index subgroup Γ (of index m) in the definition of the counting measures, then the rescaled measures converge in the weak-$*$ topology to

$$\frac{1}{m}\,\nu,$$

i.e., a constant multiple $\frac{1}{m}$ of the Lebesgue measure.

Answer: If $\Gamma \subset \mathbb{Z}^2$ is a finite-index subgroup of index m, then the counting measures

$$\nu_L^\Gamma = \frac{1}{L^2}\sum_{v\in\Gamma}\delta_{\frac{1}{L}v}$$

satisfy

$$\lim_{L\to\infty}\nu_L^\Gamma = \frac{1}{m}\,\nu,$$

in the weak-$*$ topology.

2.1 Invariance of Counting Measures

Motivation and Overview

In previous sections we have defined counting measures that record the distribution of *primitive lattice points* in the Euclidean plane. Recall that a vector in $\mathbb{Z}^2$ is called *primitive* if its coordinates have no common factor greater than 1. For each positive real number L, we defined the counting measure

$$\nu_L^{\text{prim}} := \frac{1}{L^2}\sum_{v\in\mathbb{Z}^2_{\text{prim}}}\delta_{\frac{v}{L}},$$

where δ_x denotes the Dirac measure at the point x and $\mathbb{Z}^2_{\text{prim}}$ is the set of primitive points in $\mathbb{Z}^2$.

Our goal in this section is to study the asymptotic behavior of the sequence $(\nu_L^{\text{prim}})_{L>0}$ in the weak-$*$ topology and, in particular, to show that every weak-$*$ limit point is invariant under the natural action of the group

$$\mathrm{SL}(2,\mathbb{Z}) = \left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \mathrm{Mat}_{2\times 2}(\mathbb{R}) : a,b,c,d\in\mathbb{Z},\ ad-bc=1\right\}.$$

This invariance property plays a crucial role in proving further asymptotic results (such as the equidistribution of these counting measures) and ultimately guides the proof of analogous counting theorems for closed geodesics on hyperbolic surfaces.

Preliminary Definitions and Examples

Definition 2.4 (Dirac Measure)**.** Given a point $x \in \mathbb{R}^2$, the *Dirac measure* δ_x is defined by

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$

for any measurable set $A \subset \mathbb{R}^2$.

Example 2.5. For example, if $x = (0,0)$ and A is the unit disk $B(0,1) = \{y \in \mathbb{R}^2 : \|y\| \leq 1\}$, then $\delta_{(0,0)}(B(0,1)) = 1$ because the origin is inside the disk.

Definition 2.6 (Weak-$*$ Convergence of Measures)**.** A sequence of locally finite Borel measures $(\mu_L)_{L>0}$ on $\mathbb{R}^2$ is said to converge *weak-$*$* (or vaguely) to a measure μ if, for every continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ with compact support,

$$\lim_{L\to\infty} \int_{\mathbb{R}^2} f \, d\mu_L = \int_{\mathbb{R}^2} f \, d\mu.$$

Example 2.7. If μ is the Lebesgue measure on $\mathbb{R}^2$, then one expects that as $L \to \infty$ the scaled counting measure

$$\nu_L := \frac{1}{L^2} \sum_{v\in\mathbb{Z}^2} \delta_{v/L}$$

converges weak-$*$ to μ. (See Exercise 2.14 for details.)

Definition 2.8 (Counting Measure for Primitive Lattice Points)**.** For each $L > 0$, define

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v\in\mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

where $\mathbb{Z}^2_{\text{prim}} = \{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1\}$.

Example 2.9. For $L = 2$, the set $\frac{1}{2}\mathbb{Z}^2_{\text{prim}}$ consists of all points of the form $(a/2, b/2)$ with $\gcd(a,b) = 1$. For instance, $(1/2, 0)$, $(1/2, 1/2)$, and $(0, 1/2)$ are all present, while $(1,0)$ also appears (since $(2,0)$ is not primitive, it is not scaled in this measure).

Invariance under the Action of $\mathrm{SL}(2,\mathbb{Z})$

The Group $\mathrm{SL}(2,\mathbb{Z})$**:** The group

$$\mathrm{SL}(2,\mathbb{Z}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z},\ ad - bc = 1 \right\}$$

acts on $\mathbb{R}^2$ by linear transformations. That is, for a matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and a vector $x \in \mathbb{R}^2$, the action is given by

$$M \cdot x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} x.$$

Why Invariance is Important: Understanding the invariance properties of a sequence of measures is crucial for studying its limit behavior. In our setting, we wish to prove that any weak-$*$ limit point of the sequence $(\nu_L^{\mathrm{prim}})_{L>0}$ is invariant under the action of $\mathrm{SL}(2,\mathbb{Z})$. This invariance property implies that the limiting measure cannot favor any particular direction or location; it must be "evenly spread out" in a way that reflects the symmetry of the lattice.

Key Proposition.

Proposition 2.10. *Every weak-$*$ limit point of the sequence $(\nu_L^{\mathrm{prim}})_{L>0}$ is invariant under the natural action of* $\mathrm{SL}(2,\mathbb{Z})$ *on* $\mathbb{R}^2$.

Proof. Let μ be a weak-$*$ limit point of $(\nu_L^{\mathrm{prim}})_{L>0}$; that is, there exists a sequence $L_k \to \infty$ such that $\nu_{L_k}^{\mathrm{prim}}$ converges weak-$*$ to μ.

Take any matrix $M \in \mathrm{SL}(2,\mathbb{Z})$ and any continuous, compactly supported function $f : \mathbb{R}^2 \to \mathbb{R}$. Because M acts linearly and preserves the lattice $\mathbb{Z}^2$ (indeed, $M(\mathbb{Z}^2_{\mathrm{prim}}) = \mathbb{Z}^2_{\mathrm{prim}}$), we have

$$\int_{\mathbb{R}^2} f(Mx)\, d\nu_{L_k}^{\mathrm{prim}}(x) = \frac{1}{L_k^2} \sum_{v \in \mathbb{Z}^2_{\mathrm{prim}}} f\Big(M\frac{v}{L_k}\Big).$$

Since M permutes the set $\mathbb{Z}^2_{\mathrm{prim}}$, the above sum equals

$$\frac{1}{L_k^2} \sum_{w \in \mathbb{Z}^2_{\mathrm{prim}}} f\left(\frac{w}{L_k}\right) = \int_{\mathbb{R}^2} f(x)\, d\nu_{L_k}^{\mathrm{prim}}(x).$$

Taking the limit as $k \to \infty$ (and using the weak-$*$ convergence of $\nu_{L_k}^{\mathrm{prim}}$ to μ) gives

$$\int_{\mathbb{R}^2} f(Mx)\, d\mu(x) = \int_{\mathbb{R}^2} f(x)\, d\mu(x).$$

Since f was an arbitrary test function, this shows that the push-forward measure $M_*\mu$ equals μ. In other words, μ is invariant under the action of M. Since M was an arbitrary element of $\mathrm{SL}(2,\mathbb{Z})$, the limit measure μ is $\mathrm{SL}(2,\mathbb{Z})$-invariant. □

Example 2.11. Consider the measure ν_L defined by

$$\nu_L := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2} \delta_{v/L}.$$

It can be shown (via geometric arguments) that ν_L converges weak-$*$ to the Lebesgue measure on $\mathbb{R}^2$. Since the Lebesgue measure is invariant under all translations and rotations (and in particular under the linear action of $\mathrm{SL}(2,\mathbb{Z})$), this provides an example of a limit measure that is invariant. A similar invariance holds for ν_L^{prim} (with the appropriate normalization), and the above proposition confirms this invariance in full generality.

Why Is This Invariance Useful?

Once we know that every weak-$*$ limit point of the sequence $(\nu_L^{\text{prim}})_{L>0}$ is $\mathrm{SL}(2,\mathbb{Z})$-invariant, we can deduce several important facts:

1. **Uniqueness:** The only locally finite, translation (or, more generally, $\mathrm{SL}(2,\mathbb{Z})$)-invariant measure on $\mathbb{R}^2$ is a constant multiple of the Lebesgue measure. (This follows from standard uniqueness results in measure theory and the geometry of $\mathbb{R}^2$.)

2. **Equidistribution:** It then follows that the counting measures ν_L^{prim} must become equidistributed in $\mathbb{R}^2$ as $L \to \infty$, meaning they converge (in the weak-$*$ sense) to the Lebesgue measure up to a constant.

3. **Applications to Number Theory and Geometry:** This invariance is a key ingredient in proving asymptotic formulas for $p(\mathbb{Z}^2, L)$ (the number of primitive lattice points in a ball of radius L). In particular, one eventually shows that

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2.$$

   Similar ideas also appear in the study of closed geodesics on hyperbolic surfaces.

Summary

To summarize, we have introduced the family of counting measures

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

and we explained that, although it is not immediately obvious, every weak-$*$ limit point of this sequence is invariant under the linear action of the discrete group $\mathrm{SL}(2,\mathbb{Z})$ on $\mathbb{R}^2$. We provided the following key result:

Proposition 2.12 (Invariance of Limit Points)**.** *Every weak-$*$ limit point of the sequence $(\nu_L^{\text{prim}})_{L>0}$ is invariant under the action of* $\mathrm{SL}(2,\mathbb{Z})$ *on* $\mathbb{R}^2$.

This invariance is crucial for later applications, as it allows us to identify the limit measure (up to a constant) as the Lebesgue measure, thereby proving equidistribution results for primitive lattice points. The ideas presented here serve as a model for similar techniques that will be applied to counting closed geodesics on hyperbolic surfaces later in the survey.

Exercise 2.3. Show that the $\mathrm{SL}(2,\mathbb{Z})$ orbit of the vector $(1,0) \in \mathbb{R}^2$ is precisely $\mathbb{Z}^2_{\text{prim}} \subseteq \mathbb{Z}^2$. Conclude that any weak-$*$ limit point of the sequence $(\nu_L^{\text{prim}})_{L>0}$ is $\mathrm{SL}(2,\mathbb{Z})$-invariant. **Hint:** Use Bézout's identity for greatest common divisors.

Solution to Exercise 2.3

Problem Restatement

Let $\mathbb{Z}^2_{\text{prim}} \subseteq \mathbb{Z}^2$ be the set of all *primitive* integer vectors, i.e., those (a, b) with $\gcd(a, b) = 1$. We wish to show that

$$\mathrm{SL}(2, \mathbb{Z}) \cdot (1, 0) \;=\; \mathbb{Z}^2_{\text{prim}},$$

i.e., the orbit of $(1, 0)$ under the group $\mathrm{SL}(2, \mathbb{Z})$ is precisely the set of all primitive vectors in $\mathbb{Z}^2$. Conclude that any weak-$*$ limit point of the sequence $\left(\nu_L^{\text{prim}}\right)_{L>0}$ is $\mathrm{SL}(2, \mathbb{Z})$-invariant.

Part A. Why the $\mathrm{SL}(2, \mathbb{Z})$ Orbit of $(1, 0)$ is $\mathbb{Z}^2_{\text{prim}}$

We need to prove two directions:

1. **If a vector $v \in \mathbb{Z}^2$ lies in the $\mathrm{SL}(2, \mathbb{Z})$-orbit of $(1, 0)$, then v is primitive** (i.e., the gcd of its coordinates is 1).
2. **If $v \in \mathbb{Z}^2$ is primitive, then v lies in the orbit of $(1, 0)$.**

We denote by $\mathrm{SL}(2, \mathbb{Z})$ the set of 2×2 integer matrices with determinant 1.

1. Orbit Vectors are Always Primitive

Take any $M \in \mathrm{SL}(2, \mathbb{Z})$. By definition,

$$M \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with } a, b, c, d \in \mathbb{Z} \quad \text{and} \quad \det(M) = ad - bc = 1.$$

Now compute

$$M \cdot (1, 0) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}.$$

Thus, any vector in the orbit has the form (a, c) (i.e., the *first column* of M).

Now, suppose for the sake of contradiction that $\gcd(a, c) > 1$. Let p be a common divisor with $p > 1$. Then p divides both a and c, and hence it divides any linear combination of these numbers. In particular, p divides

$$ad - bc,$$

but since $\det(M) = ad - bc = 1$, it would follow that p divides 1, which is impossible unless $p = 1$. Thus, $\gcd(a, c) = 1$. This shows that

$$M(1, 0) \;\in\; \mathbb{Z}^2_{\text{prim}} \quad \text{for every } M \in \mathrm{SL}(2, \mathbb{Z}),$$

so that

$$\mathrm{SL}(2, \mathbb{Z}) \cdot (1, 0) \;\subseteq\; \mathbb{Z}^2_{\text{prim}}.$$

2. Every Primitive Vector is in the Orbit

Conversely, let $(x, y) \in \mathbb{Z}^2_{\text{prim}}$, meaning that $\gcd(x, y) = 1$. We wish to find a matrix $M \in \mathrm{SL}(2, \mathbb{Z})$ such that

$$M(1, 0) = (x, y).$$

That is, we want $\begin{pmatrix} x \\ y \end{pmatrix}$ to be the first column of M.

Consider a matrix of the form

$$M = \begin{pmatrix} x & * \\ y & * \end{pmatrix}.$$

For M to belong to $\mathrm{SL}(2, \mathbb{Z})$, we need all entries to be integers and the determinant to be 1. That is, if the second column is $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, we require

$$\det(M) = x\beta - \alpha y = 1.$$

Use of Bézout's Identity. Since $\gcd(x, y) = 1$, Bézout's identity guarantees that there exist integers r and s such that

$$r\,x + s\,y = 1.$$

We can choose these r and s and define the matrix

$$M = \begin{pmatrix} x & -s \\ y & r \end{pmatrix}.$$

Then, the determinant is

$$\det(M) = x \cdot r - (-s) \cdot y = xr + sy = 1,$$

by our choice of r and s. Thus, $M \in \mathrm{SL}(2, \mathbb{Z})$. Moreover,

$$M(1, 0) = \begin{pmatrix} x \\ y \end{pmatrix},$$

so (x, y) lies in the orbit of $(1, 0)$.

Altogether, we have shown that if (x, y) is primitive then

$$(x, y) \;\in\; \mathrm{SL}(2, \mathbb{Z}) \cdot (1, 0),$$

so that

$$\mathbb{Z}^2_{\text{prim}} \;\subseteq\; \mathrm{SL}(2, \mathbb{Z}) \cdot (1, 0).$$

Combined Conclusion

From the two inclusions we deduce

$$\mathrm{SL}(2, \mathbb{Z}) \cdot (1, 0) \;=\; \mathbb{Z}^2_{\text{prim}},$$

i.e., the orbit of $(1, 0)$ under $\mathrm{SL}(2, \mathbb{Z})$ is precisely the set of all primitive integer vectors.

Part B. Invariance of any Weak-$*$ Limit Point of $\left(\nu_L^{\text{prim}}\right)$

Recall that ν_L^{prim} is defined by

$$\nu_L^{\text{prim}} = \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

where δ_x denotes the Dirac measure at $x \in \mathbb{R}^2$.

The Support is the Orbit. From Part A we have shown that

$$\mathbb{Z}^2_{\text{prim}} = \mathrm{SL}(2,\mathbb{Z}) \cdot (1,0).$$

Hence, the measure ν_L^{prim} can be thought of as

$$\nu_L^{\text{prim}} = \frac{1}{L^2} \sum_{v \,\in\, \mathrm{SL}(2,\mathbb{Z})\cdot(1,0)} \delta_{v/L}.$$

Why Invariance Follows. Suppose that μ is a weak-$*$ limit of the sequence $\{\nu_L^{\text{prim}}\}_{L>0}$. We wish to show that μ is invariant under the action of $\mathrm{SL}(2,\mathbb{Z})$; that is, for every $g \in \mathrm{SL}(2,\mathbb{Z})$ we have

$$g_*\mu = \mu,$$

where $g_*\mu$ denotes the pushforward of μ under the transformation g. Concretely, for any continuous and compactly supported test function ϕ,

$$\int_{\mathbb{R}^2} \phi(x)\, d(g_*\mu)(x) = \int_{\mathbb{R}^2} \phi\big(g(x)\big)\, d\mu(x).$$

The key observation is that each measure ν_L^{prim} is itself $\mathrm{SL}(2,\mathbb{Z})$-invariant. To see this, note that the action of any $g \in \mathrm{SL}(2,\mathbb{Z})$ simply permutes the primitive vectors in $\mathbb{Z}^2_{\text{prim}}$. Therefore,

$$g_*\nu_L^{\text{prim}} = \nu_L^{\text{prim}} \quad \text{for each } L > 0.$$

Since pushforward is continuous with respect to the weak-$*$ topology, if μ is a limit point of ν_L^{prim} as $L \to \infty$, then

$$g_*\mu = g_*\left(\lim_{L\to\infty} \nu_L^{\text{prim}}\right) = \lim_{L\to\infty} g_*\big(\nu_L^{\text{prim}}\big) = \lim_{L\to\infty} \nu_L^{\text{prim}} = \mu.$$

Thus, μ is $\mathrm{SL}(2,\mathbb{Z})$-invariant.

Conclusion. Every weak-$*$ limit point of the sequence $\left(\nu_L^{\text{prim}}\right)$ must be invariant under the full group $\mathrm{SL}(2,\mathbb{Z})$.

Final Summary

1. **Orbit Characterization:**

$$\mathrm{SL}(2,\mathbb{Z}) \cdot (1,0) = \{(a,c) \in \mathbb{Z}^2 : \gcd(a,c) = 1\} = \mathbb{Z}^2_{\mathrm{prim}}.$$

   - *Proof Idea:*
     - ($\supseteq$): For any $M \in \mathrm{SL}(2,\mathbb{Z})$, the first column $(a,c) = M(1,0)$ satisfies $\gcd(a,c) = 1$.
     - ($\subseteq$): Conversely, if (x,y) is primitive, Bézout's identity provides integers r,s such that $rx + sy = 1$, allowing the construction of

$$M = \begin{pmatrix} x & -s \\ y & r \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

       with $M(1,0) = (x,y)$.

2. $\mathrm{SL}(2,\mathbb{Z})$**-Invariance of Limit Measures:**

   - Each measure ν_L^{prim} is supported on the orbit $\mathbb{Z}^2_{\mathrm{prim}}$ and is invariant under $\mathrm{SL}(2,\mathbb{Z})$ (since the group action merely permutes the primitive vectors).
   - Consequently, any weak-$*$ limit μ of $\{\nu_L^{\mathrm{prim}}\}$ is also $\mathrm{SL}(2,\mathbb{Z})$-invariant.

This completes the solution.

Exercise 2.3 ensures that any weak-$*$ limit point of the sequence $(\nu_L^{\mathrm{prim}})_{L>0}$ is $\mathrm{SL}(2,\mathbb{Z})$-invariant. Using ergodic theory we will show this property greatly constrains the possible weak-$*$ limit points.

2.2 Ergodic Theory and Invariance of Measures

In many areas of mathematics—including dynamics, number theory, and geometry—it is crucial to study measures that are invariant under the action of a group. In this section we introduce some of the basic concepts of ergodic theory and explain why the notion of invariance is important. We will include precise definitions, illustrative examples, and key propositions with proofs so that every statement is clear.

2.2.1 Basic Definitions and Motivation

Definition 2.13 (Measure-Preserving Action)**.** Let $(X,\mathcal{A})$ be a measurable space and let G be a countable group. An *action* of G on X is a map

$$G \times X \to X, \quad (g,x) \mapsto g \cdot x,$$

such that for all $g,h \in G$ and $x \in X$,

$$e \cdot x = x \quad \text{and} \quad g \cdot (h \cdot x) = (gh) \cdot x,$$

where e denotes the identity element of G.

A σ-finite measure μ on $(X, \mathcal{A})$ is said to be *measure-preserving* (with respect to the action of G) if for every $g \in G$ and every measurable set $A \in \mathcal{A}$,

$$\mu(g.A) = \mu(A),$$

where we define

$$g.A := \{g \cdot x : x \in A\}.$$

Example 2.14 (Translation on the Circle)**.** Let $X = \mathbb{R}/\mathbb{Z}$ be the unit circle (viewed as $[0, 1]$ with the endpoints identified), and let $G = \mathbb{Z}$ act on X by translation:

$$n \cdot x = x + n\alpha \pmod 1,$$

for some fixed real number α. If we equip X with the Lebesgue measure μ (which is the natural measure on the circle), then one can check that μ is measure-preserving. Indeed, translating a set by a fixed number does not change its length.

Definition 2.15 (Invariant Measure)**.** A measure μ on $(X, \mathcal{A})$ is said to be G*-invariant* if it is preserved under the action of every element of G; that is, for all $g \in G$ and for every measurable set $A \in \mathcal{A}$,

$$\mu(g.A) = \mu(A).$$

Definition 2.16 (Ergodicity)**.** Let μ be a G-invariant, σ-finite measure on $(X, \mathcal{A})$. We say that μ is G*-ergodic* if there are no nontrivial G-invariant measurable subsets of X. More precisely, if $A \in \mathcal{A}$ is such that

$$g.A = A \quad \text{for all } g \in G,$$

then either

$$\mu(A) = 0 \quad \text{or} \quad \mu(X \setminus A) = 0.$$

An equivalent formulation is: μ is G-ergodic if every measurable function $f : X \to \mathbb{R}$ that is invariant under the action of G (i.e., $f(g \cdot x) = f(x)$ for all $g \in G$ and almost every x) is constant almost everywhere.

Example 2.17 (Irrational Rotation)**.** Consider again the circle $X = \mathbb{R}/\mathbb{Z}$ with Lebesgue measure μ and let $G = \mathbb{Z}$ act by $n \cdot x = x + n\alpha \pmod 1$ where α is irrational. In this case, it can be shown that μ is not only invariant but also ergodic. The reason is that any measurable function f that is invariant under all such rotations must be constant almost everywhere; otherwise, one could partition the circle into two invariant sets, contradicting ergodicity.

2.2.2 Why Ergodicity and Invariance Matter

Understanding invariant and ergodic measures is central in many areas of dynamics and geometry. For instance, when studying sequences of counting measures (such as those arising in lattice point counting or in the distribution of closed geodesics), one wishes to know whether these measures, in the limit, "spread out" evenly over the space. If the limit measure is invariant (and under additional assumptions, ergodic), then by classical results it must be a constant multiple of a "natural" measure (for example, the Lebesgue measure).

Proposition 2.18 (Uniqueness of Invariant Measures). *Let μ be a locally finite Borel measure on $\mathbb{R}^2$ that is invariant under all translations (or under a group that acts densely in the translation group). Then there exists a constant $c \geq 0$ such that*

$$\mu = c\,\nu,$$

where ν denotes the Lebesgue measure on $\mathbb{R}^2$.

Proof. The proof uses the fact that translations act ergodically on $\mathbb{R}^2$ with respect to the Lebesgue measure. One shows that the measure of a fundamental domain (such as the unit square) completely determines the measure, and translation invariance forces the measure to be uniformly distributed. A detailed proof can be found in standard texts on measure theory and ergodic theory; the key idea is to use the invariance to partition $\mathbb{R}^2$ into translates of a unit square and then show that the measure must be proportional to the area. □

Example 2.19 (Counting Measures on the Lattice). Consider the counting measure on the integer lattice

$$\nu_L := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2} \delta_{v/L}.$$

It can be shown that ν_L converges (in the weak-$*$ sense) to the Lebesgue measure ν as $L \to \infty$. In fact, by a simple scaling argument, the measure of a fixed compact set (like the unit square) under ν_L converges to its area. Notice that ν_L is invariant under the natural action of $\mathbb{Z}^2$ (and hence under $\mathrm{SL}(2,\mathbb{Z})$ acting appropriately), which is one of the key properties used to identify its limit.

2.2.3 Characterizing G-Invariant Measures Relative to an Ergodic Measure

A useful application of ergodic theory is the following characterization: Suppose ν is a G-ergodic measure on $(X, \mathcal{A})$ and μ is another G-invariant measure that is absolutely continuous with respect to ν (denoted $\mu \ll \nu$). Then, ergodicity forces μ to be a constant multiple of ν. In symbols,

$$\mu = c\,\nu \quad \text{for some } c \geq 0.$$

This fact is a cornerstone in many equidistribution results.

Theorem 2.20 (Ergodic Decomposition for Absolutely Continuous Measures). *Let ν be a G-ergodic measure on $(X, \mathcal{A})$ and let μ be a G-invariant measure with $\mu \ll \nu$. Then there exists a constant $c \geq 0$ such that*

$$\mu = c\,\nu.$$

Proof. Since $\mu \ll \nu$, by the Radon–Nikodym theorem there exists a measurable function $f : X \to [0, \infty)$ such that

$$\mu(A) = \int_A f\, d\nu \quad \text{for all } A \in \mathcal{A}.$$

The invariance of μ and ν under the action of G implies that for any $g \in G$ and any measurable function f,

$$\int_X f(x)\, d\mu(x) = \int_X f(g \cdot x)\, d\mu(x).$$

By substituting the Radon–Nikodym derivative and performing a change of variables (using the measure-preserving property), one can show that $f(g \cdot x) = f(x)$ for ν-almost every $x \in X$. In other words, f is G-invariant. Since ν is ergodic, every G-invariant function must be constant almost everywhere. Thus, there exists a constant $c \geq 0$ such that $f(x) = c$ for ν-almost every x, and hence $\mu = c\,\nu$. □

Exercise 2.4. Let $(X, \mathcal{A})$ be a measurable space and G be a countable group acting on $(X, \mathcal{A})$ by measure-preserving transformations. Suppose that ν is a G-ergodic measure on $(X, \mathcal{A})$ and that μ is a G-invariant measure on $(X, \mathcal{A})$ that is absolutely continuous with respect to ν. Show that μ is a non-negative constant multiple of ν.

That is, please show that if ν is the Lebesgue measure on $\mathbb{R}^2$ and $G = \mathbb{Z}^2$ acts by translations, then any G-invariant measure that is absolutely continuous with respect to ν is a constant multiple of ν.

Hint: Show that the Radon-Nikodym derivative of μ with respect to ν is G-invariant and use the G-ergodicity of ν to show this derivative is constant.

Solution Explanation

We have:

- A countable group G.
- A measurable space $\big(X, \mathcal{A}\big)$.
- A measure ν on $\big(X, \mathcal{A}\big)$ which is G-*ergodic*.
- A measure μ on $\big(X, \mathcal{A}\big)$ which is G-*invariant* (i.e., $\mu(gA) = \mu(A)$ for all $g \in G$ and $A \in \mathcal{A}$).
- μ is absolutely continuous with respect to ν (written $\mu \ll \nu$).

We want to prove that μ must be of the form

$$\mu = c\,\nu$$

for some non-negative constant c.

Step 1: The Radon–Nikodým Derivative

Because μ is absolutely continuous with respect to ν, by the *Radon–Nikodým theorem* there exists a *Radon–Nikodým derivative* (often denoted by a density)

$$f \;=\; \frac{d\mu}{d\nu}$$

such that for every measurable set $A \in \mathcal{A}$,

$$\mu(A) \;=\; \int_A f\, d\nu.$$

Equivalently, for any (nonnegative) measurable function φ, we have

$$\int \varphi \, d\mu \;=\; \int \varphi \, f \, d\nu.$$

The function f is measurable and nonnegative (since μ and ν are nonnegative measures).

Step 2: Showing f is G-Invariant Almost Everywhere

We want to show that f is G-invariant ν-almost everywhere; in symbols,

$$f(gx) \;=\; f(x) \quad \text{for } \nu\text{-almost every } x \in X, \text{ for all } g \in G.$$

To see why, fix $g \in G$. Consider any (nonnegative) measurable function φ. Because μ is G-invariant, we have

$$\int_X \varphi(x) \, d\mu(x) \;=\; \int_X \varphi(x) \, d\mu(x) \quad \text{(trivial equality)}.$$

But also, using the invariance of μ under the transformation $x \mapsto g^{-1}x$, we get

$$\int_X \varphi(x) \, d\mu(x) \;=\; \int_X \varphi\big(g^{-1}y\big) \, d\mu(y).$$

(Here we replaced the dummy variable x by $g^{-1}y$.)

On the other hand, by the Radon–Nikodým relationship for μ in terms of ν,

$$\int_X \varphi(x) \, d\mu(x) \;=\; \int_X \varphi(x) \, f(x) \, d\nu(x),$$

and

$$\int_X \varphi\big(g^{-1}y\big) \, d\mu(y) \;=\; \int_X \varphi\big(g^{-1}y\big) \, f(y) \, d\nu(y).$$

Putting these equalities together yields

$$\int_X \varphi(x) \, f(x) \, d\nu(x) \;=\; \int_X \varphi\big(g^{-1}y\big) \, f(y) \, d\nu(y).$$

Next, perform the change of variables $x = g^{-1}y$ (i.e., $y = gx$) in the right-hand integral. Note that when ν is *invariant* under G (since the group acts by measure-preserving transformations, we have $\nu(gA) = \nu(A)$), we have $d\nu(y) = d\nu(gx) = d\nu(x)$. Thus,

$$\int_X \varphi\big(g^{-1}y\big) \, f(y) \, d\nu(y) \;=\; \int_X \varphi(x) \, f\big(gx\big) \, d\nu(x).$$

Hence,

$$\int_X \varphi(x) \, f(x) \, d\nu(x) \;=\; \int_X \varphi(x) \, f\big(gx\big) \, d\nu(x).$$

Since φ is arbitrary, we get (by taking φ to be indicators or any suitable function) that

$$f(x) = f\big(gx\big) \quad \text{for } \nu\text{-almost every } x.$$

Because $g \in G$ was arbitrary, this shows

$$f \circ g = f \quad \nu\text{-almost everywhere, for all } g \in G.$$

In other words, f is a G-invariant function (almost everywhere with respect to ν).

Step 3: Using G-Ergodicity of ν to Conclude f is Constant Almost Everywhere

Recall the definition of G-ergodicity of the measure ν:

*A measure ν on $(X, \mathcal{A})$ is G-**ergodic** if any G-invariant measurable set has either ν-measure 0 or full ν-measure.*

Equivalently, a standard result is that if ν is G-ergodic, then any G-invariant measurable function must be constant ν-almost everywhere. The intuition is that, in an ergodic system, there is no nontrivial decomposition of the space into smaller invariant pieces; thus any quantity (function) that does not change under the group action must be essentially the same value everywhere.

Since we have established that f is G-invariant almost everywhere, by the G-ergodicity of ν it follows that f is a constant almost everywhere. In symbols, there exists $c \in [0, \infty)$ such that

$$f(x) = \frac{d\mu}{d\nu}(x) = c \quad \text{for } \nu\text{-almost every } x.$$

Step 4: Conclusion — μ is a Constant Multiple of ν

Given that $\frac{d\mu}{d\nu} = c$ almost everywhere, for any measurable set $A \in \mathcal{A}$,

$$\mu(A) = \int_A \frac{d\mu}{d\nu}\, d\nu = \int_A c\, d\nu = c \int_A d\nu = c\, \nu(A).$$

Hence,

$$\mu(A) = c\, \nu(A) \quad \text{for all } A \in \mathcal{A}.$$

Thus, we see that $\mu = c\nu$ as measures. Since both μ and ν are nonnegative measures, we must have $c \geq 0$.

Therefore, μ **is a nonnegative constant multiple of** ν, which is exactly what we wanted to prove.

Key Points Recap

1. **Radon–Nikodým Derivative:** Existence of $f = \frac{d\mu}{d\nu}$ because $\mu \ll \nu$.
2. **G-Invariance of f:** Follows from μ being G-invariant and ν being preserved by the group action.
3. **Ergodicity:** A G-invariant function under an ergodic measure must be constant almost everywhere.
4. **Result:** $\mu = c\nu$.

This completes the proof in a fully detailed manner.

2.2.4 Summary

To summarize, we have defined the notions of G-invariance and G-ergodicity for measures on a measurable space $(X, \mathcal{A})$. We provided examples, such as the Lebesgue measure on the circle under translation and counting measures on the lattice, to illustrate these concepts. We then proved that any weak-$*$ limit of a sequence of counting measures (such as those arising from lattice point counting) inherits invariance properties. Finally, we stated and proved the theorem that if a G-invariant measure is absolutely continuous with respect to a G-ergodic measure, then it must be a constant multiple of that ergodic measure. These ideas are fundamental in modern ergodic theory and have far-reaching applications in counting problems and equidistribution theorems.

2.3 Ergodicity of the Lebesgue Measure and Portmanteau's Theorem

In this section, we present two important theorems. The first (Theorem 2.24) asserts that the standard Lebesgue measure on $\mathbb{R}^2$ is *ergodic* under the linear action of the group $\mathrm{SL}(2, \mathbb{Z})$. Although we do not prove it fully here, we discuss its relevance and outline where one can find a more advanced proof (involving horocycle flows). The second result (Theorem 2.42) is the well-known *Portmanteau theorem*, a classic statement in measure theory that characterizes weak-$*$ convergence of measures in various ways.

2.3.1 Ergodicity of the Lebesgue Measure under $\mathrm{SL}(2, \mathbb{Z})$

Definition 2.21 (Linear Action of $\mathrm{SL}(2, \mathbb{Z})$)**.** Let $\mathrm{SL}(2, \mathbb{Z})$ be the group of 2×2 integer matrices with determinant 1. Each element $M \in \mathrm{SL}(2, \mathbb{Z})$ defines a map

$$M : \mathbb{R}^2 \to \mathbb{R}^2, \quad x \mapsto M \cdot x,$$

where $\cdot$ denotes matrix multiplication of the vector x. This action is said to be *linear* because each M acts via a linear transformation.

Definition 2.22 (Ergodicity and Invariance)**.** Given a σ-finite measure μ on $\mathbb{R}^2$ (with its Borel σ-algebra), we say μ is *invariant* under $\mathrm{SL}(2,\mathbb{Z})$ if, for every $M \in \mathrm{SL}(2,\mathbb{Z})$ and every measurable set $A \subset \mathbb{R}^2$, we have

$$\mu(M \cdot A) = \mu(A).$$

If μ is invariant, we say μ is *ergodic* under the $\mathrm{SL}(2,\mathbb{Z})$-action if there are no nontrivial $\mathrm{SL}(2,\mathbb{Z})$-invariant subsets (mod μ). Formally, if $A \subseteq \mathbb{R}^2$ is measurable and $M \cdot A = A$ for all $M \in \mathrm{SL}(2,\mathbb{Z})$, then either $\mu(A) = 0$ or $\mu(\mathbb{R}^2 \setminus A) = 0$. Equivalently, any measurable function $f : \mathbb{R}^2 \to \mathbb{R}$ that satisfies $f(M \cdot x) = f(x)$ for all $M \in \mathrm{SL}(2,\mathbb{Z})$ and almost every x must be constant μ-almost everywhere.

Example 2.23 (Translation Invariance vs. $\mathrm{SL}(2,\mathbb{Z})$-Invariance)**.** The standard Lebesgue measure ν on $\mathbb{R}^2$ is invariant under *all* translations. However, $\mathrm{SL}(2,\mathbb{Z})$-invariance means invariance under integer linear transformations with determinant 1. One can check directly that ν is indeed invariant under $\mathrm{SL}(2,\mathbb{Z})$ because these transformations are volume-preserving. Ergodicity is a stronger statement asserting that there is no nontrivial decomposition of $\mathbb{R}^2$ into smaller $\mathrm{SL}(2,\mathbb{Z})$-invariant sets.

Theorem 2.24 (Ergodicity of the Lebesgue Measure under $\mathrm{SL}(2,\mathbb{Z})$)**.** *Let ν be the standard Lebesgue measure on $\mathbb{R}^2$. Then ν is ergodic with respect to the linear action of* $\mathrm{SL}(2,\mathbb{Z})$.

Sketch of Reasoning (without full proof):

- One can prove this result by connecting it to the *horocycle flow* on a certain 3-dimensional manifold known as the unit tangent bundle of the modular surface $\mathrm{SL}(2,\mathbb{Z})\backslash\mathrm{SL}(2,\mathbb{R})$. The horocycle flow is known, by classical theorems of Hedlund (and further works by Dani, Ratner, etc.), to be ergodic with respect to certain measures.
- Equivalently, via geometry-of-numbers arguments, one shows that any function on $\mathbb{R}^2$ invariant under all integer linear transformations (with determinant 1) must be constant ν-almost everywhere.

Although the complete proof is beyond the scope of this survey, we mention it because it is essential in understanding advanced counting problems for primitive lattice points and related equidistribution statements.

2.3.2 Portmanteau's Theorem

Next, we introduce a classic theorem from measure theory that provides multiple equivalent formulations of weak-$*$ convergence of measures. This result is crucial when dealing with sequences of counting measures, as we often want to check convergence by testing it on certain subsets rather than on a dense class of test functions.

Definition 2.25 (Weak-$*$ Convergence of Measures)**.** Let X be a metric space and let $\{\mu_L\}_{L>0}$ be a sequence of (locally finite) Borel measures on X. We say μ_L *converges in the weak-$*$ topology* (or converges *vaguely*) to a Borel measure μ if, for every continuous function $f : X \to \mathbb{R}$ with compact support,

$$\lim_{L\to\infty} \int_X f \, d\mu_L = \int_X f \, d\mu.$$

Theorem 2.26 (Portmanteau's Theorem)**.** *Let X be a metric space and let $(\mu_L)_{L>0}$ be a sequence of locally finite Borel measures on X. Suppose $\mu_L \to \mu$ in the weak-$*$ topology. Then each of the following statements holds:*

1. *For every open set $U \subseteq X$,*
$$\mu(U) \ \le \ \liminf_{L\to\infty} \mu_L(U).$$
2. *For every closed set $F \subseteq X$,*
$$\mu(F) \ \ge \ \limsup_{L\to\infty} \mu_L(F).$$
3. *For every compact set $K \subseteq X$ such that $\mu(\partial K) = 0$,*
$$\lim_{L\to\infty} \mu_L(K) = \mu(K). \tag{2.5}$$

Motivation: While the definition of weak-$*$ convergence is typically given in terms of test functions, Theorem 2.42 (Portmanteau's theorem) provides alternative ways to check convergence by evaluating measures on open sets, closed sets, or compact sets with boundary of measure zero. This is especially useful in many geometric and combinatorial counting problems, where verifying convergence on a class of sets can be more direct than working with continuous test functions.

Sketch of Proof: We outline why (1) implies (2) and how to get from the definitions of weak-$*$ convergence to (3). A full detailed proof can be found in standard measure theory textbooks.

- **Open Sets and** $\liminf$**:** For an open set U, one can construct a sequence of continuous functions f_n supported in increasingly large compact subsets of U, approximating the indicator function of U from below. Then weak-$*$ convergence implies that
$$\int f_n \, d\mu_L \to \int f_n \, d\mu$$
as $L \to \infty$. By letting $n \to \infty$, one obtains that $\mu_L(U)$ cannot dip too far below $\mu(U)$ in the limit, yielding $\mu(U) \le \liminf_{L\to\infty} \mu_L(U)$.
- **Closed Sets and** $\limsup$**:** By applying a complementary argument (or using De Morgan's laws and the result for open sets), one can show that for a closed set F,
$$\mu(F) \ \ge \ \limsup_{L\to\infty} \mu_L(F).$$
- **Compact Sets with Negligible Boundary:** If $K \subseteq X$ is compact and $\mu(\partial K) = 0$, one partitions K into its interior plus the boundary. Using the result for open sets and closed sets, one compares $\mu_L(K)$ to $\mu_L(\text{interior}(K))$ and $\mu_L(\overline{K})$, obtaining the desired equality in the limit.

Example 2.27. If $X = \mathbb{R}^2$ and μ_L is a sequence of counting measures associated to certain discrete sets (e.g., scaled versions of the integer lattice), checking that $\mu_L(K) \to \mu(K)$ for a family of compact sets K whose boundaries are μ-negligible is often more straightforward. Portmanteau's theorem then concludes the weak-$*$ convergence to μ.

2.3.3 Conclusion and Relevance

Theorem 2.24 (ergodicity of Lebesgue measure under $\mathrm{SL}(2,\mathbb{Z})$) plays a decisive role in many lattice-point counting arguments, ensuring that certain limiting distributions must be constant multiples of Lebesgue measure. Theorem 2.42 (Portmanteau's theorem) complements this by providing practical tests for verifying when a sequence of measures converges in the weak-$*$ sense.

Together, these theorems are important for understanding advanced counting results (for example, counting primitive lattice points or counting closed geodesics on hyperbolic surfaces). In more elaborate settings, one uses similar strategies, combining an ergodicity statement (to identify the limiting measure up to constants) with a measure-convergence statement (Portmanteau's theorem) to show that a sequence of counting measures equidistributes in the limit.

Exercise 2.7. Assuming the weak-$*$ convergence in (2.3) holds, find counterexamples to identity (2.5) in Theorem 2.6 when either $K \subseteq \mathbb{R}^2$ is not compact or does not satisfy $\nu(\partial K) = 0$.

Below is a solution for Exercise 2.7, illustrating precisely how and why the identity in

$$\mu(K) = \lim_{L\to\infty} \mu_L(K)$$

(as stated in (2.5) of Theorem 2.6) can fail if either the set $K \subseteq \mathbb{R}^2$ is not compact or if its boundary has positive Lebesgue measure. We assume we are working in $\mathbb{R}^2$ with Lebesgue measure ν, and that $\{\mu_L\}_{L>0}$ is a sequence of locally finite Borel measures converging to ν in the weak-$*$ sense (as in (2.3)).

Recall of Theorem 2.6 (Portmanteau's Theorem, specialized)

Let X be a metric space, and let $\{\mu_L\}$ be a sequence of locally finite Borel measures on X converging to a Borel measure μ in the weak-$*$ topology. Then:

1. For every open set $U \subseteq X$,

$$\mu(U) \ \leq \ \liminf_{L\to\infty} \mu_L(U).$$

2. If $K \subseteq X$ is *compact* and satisfies $\mu(\partial K) = 0$, then

$$\lim_{L\to\infty} \mu_L(K) = \mu(K). \tag{2.5}$$

The identity in (2.5) requires both that K be compact and that $\mu(\partial K) = 0$. If either of these conditions fails, the conclusion need not hold. In what follows, we exhibit counterexamples for the two cases:

(1) K is not compact.

(2) K is compact (or at least bounded) but $\nu(\partial K) \neq 0$.

1. Counterexample when K is Not Compact

A simple (and somewhat extreme) example is to take

$$K = \mathbb{R}^2.$$

Note that $\mathbb{R}^2$ is not compact. Under Lebesgue measure, we have $\nu(\mathbb{R}^2) = +\infty$. On the other hand, for each L the measure $\mu_L(\mathbb{R}^2)$ may be defined in such a way (for instance, when using scaled counting measures on $\mathbb{Z}^2$) that it does not necessarily equal $+\infty$ in a controlled manner. In any case, the theorem (2.5) does not apply because the hypothesis requires K to be compact.

Geometric Picture: The condition $\mu(\partial K) = 0$ together with the compactness of K ensures that the mass assigned by μ_L to "nice bounded" regions converges to the mass of μ on these regions. If K is unbounded (like $\mathbb{R}^2$), there is no control at infinity, and thus the limit $\mu_L(K) \to \mu(K)$ need not hold.

2. Counterexample when $\nu(\partial K) \neq 0$

Now we exhibit a set K that is bounded but whose boundary has positive Lebesgue measure, so that the condition $\nu(\partial K) = 0$ is violated.

2a. Idea: The Boundary is "Too Large"

When the boundary ∂K has positive Lebesgue measure, Theorem 2.6 does not guarantee that

$$\lim_{L \to \infty} \mu_L(K) = \mu(K).$$

A classic example is to take

$$K := \Big\{ (x, y) \in [0, 1]^2 : x, y \in \mathbb{Q} \Big\},$$

i.e. the set of rational points in the unit square.

- Since K is countable, $\nu(K) = 0$ (Lebesgue measure of a countable set is zero).
- However, the closure of K is the entire unit square:

  $$\overline{K} = [0, 1]^2.$$

  Since K has empty interior (it is nowhere dense), its boundary is

  $$\partial K = \overline{K} \setminus \text{int}(K) = [0, 1]^2.$$

  Hence, $\nu(\partial K) = \nu([0, 1]^2) = 1$, which is not zero.

Thus, K fails the condition $\nu(\partial K) = 0$.

2b. How This Breaks the Limit Identity

A typical sequence of measures converging to ν is given by the scaled counting measures on $\mathbb{Z}^2$:

$$\nu_L \;=\; \frac{1}{L^2}\sum_{v\in\mathbb{Z}^2}\delta_{v/L}.$$

It is known that $\nu_L \rightharpoonup \nu$ (weak-$*$ convergence).

For the set K defined above, note the following:

- Every point in the support of ν_L is of the form v/L where $v \in \mathbb{Z}^2$. Since v_1 and v_2 are integers, the coordinates v_1/L and v_2/L are rational.
- Therefore, every point in the scaled lattice (when v/L falls in $[0,1]^2$) belongs to K. In particular, for every L, all lattice points in $[0,1]^2$ (i.e. those with $0 \le v_1, v_2 \le L$) contribute to $\nu_L(K)$.

Counting these points, we have approximately $(L+1)^2$ points in $[0,1]^2$. Hence,

$$\nu_L(K) \;=\; \frac{1}{L^2}\left|\{v=(v_1,v_2)\in\mathbb{Z}^2 : 0\le v_1,v_2\le L\}\right| \;\approx\; \frac{(L+1)^2}{L^2}.$$

As $L\to\infty$,

$$\frac{(L+1)^2}{L^2}\to 1.$$

Thus,

$$\lim_{L\to\infty}\nu_L(K) \;=\; 1.$$

But by construction, $\nu(K)=0$. Therefore,

$$\lim_{L\to\infty}\nu_L(K)=1\neq 0=\nu(K).$$

This shows that the limit identity (2.5) fails for the set K because its boundary has positive measure.

3. Conclusion

Answer to Exercise 2.7:

1. If $K\subset\mathbb{R}^2$ is not compact (for example, $K=\mathbb{R}^2$), the statement
$$\lim_{L\to\infty}\mu_L(K)=\mu(K)$$
need not hold since the theorem (2.5) only applies to compact sets.
2. If $K\subset\mathbb{R}^2$ is bounded but its boundary has positive Lebesgue measure (for example,
$$K=\{(x,y)\in[0,1]^2 : x,y\in\mathbb{Q}\}),$$
then although K is bounded, we have $\nu(K)=0$ while $\nu_L(K)\to 1$ as $L\to\infty$. Thus, the identity (2.5) fails because the condition $\nu(\partial K)=0$ is not satisfied.

These examples confirm that the extra hypotheses "K compact" and "$\mu(\partial K)=0$" are indispensable in Theorem 2.6 to guarantee that $\lim_{L\to\infty}\mu_L(K)=\mu(K)$.

2.4 Limit Points of Counting Measures

In this section, we consider a sequence $\{\nu_L^{\mathrm{prim}}\}_{L>0}$ of *counting measures* on $\mathbb{R}^2$ arising from counting *primitive* lattice points (see earlier definitions). We wish to understand the possible weak-$*$ limit points of this sequence. Our ultimate goal is to show that any such limit point must be a constant multiple of the Lebesgue measure.

2.4.1 Context and Setup

Primitive Lattice Points and Their Measures. Recall that a point $(a, b) \in \mathbb{Z}^2$ is called *primitive* if $\gcd(a, b) = 1$. One defines the measure

$$\nu_L^{\mathrm{prim}} := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\mathrm{prim}}} \delta_{v/L},$$

where δ_x denotes the Dirac measure at x. Intuitively, ν_L^{prim} "places a mass" of $\frac{1}{L^2}$ at each scaled primitive lattice point v/L. By studying the weak-$*$ limit of $\{\nu_L^{\mathrm{prim}}\}$, we can deduce asymptotic formulas for counting functions of primitive points.

Earlier Results and Motivation.

- **Exercise 2.3 (Group Orbit Argument):** Demonstrates that $\mathbb{Z}^2_{\mathrm{prim}}$ is the $\mathrm{SL}(2, \mathbb{Z})$-orbit of the point $(1, 0)$. From this fact, we concluded that any weak-$*$ limit point of $\{\nu_L^{\mathrm{prim}}\}$ must be $\mathrm{SL}(2, \mathbb{Z})$-invariant.
- **Exercise 2.4 (Absolute Continuity under Ergodicity):** States that if ν is a G-ergodic measure on a space X, then any measure μ absolutely continuous with respect to ν (and also G-invariant) must be a constant multiple of ν.
- **Theorem 2.5 (Ergodicity of Lebesgue Measure):** Asserts that the standard Lebesgue measure on $\mathbb{R}^2$ is ergodic with respect to the linear action of $\mathrm{SL}(2, \mathbb{Z})$. Although not fully proved here, it is a deep statement often shown via the ergodicity of the horocycle flow on $\mathrm{SL}(2, \mathbb{Z})\backslash\mathrm{SL}(2, \mathbb{R})$.
- **Theorem 2.6 (Portmanteau's Theorem):** Provides various characterizations of weak-$*$ convergence in metric spaces, essential for controlling the behavior of counting measures.

Using these results, we now show how to pin down the nature of any weak-$*$ limit point of $\{\nu_L^{\mathrm{prim}}\}$.

2.4.2 Key Proposition

Proposition 2.28 (Nature of Limit Points)**.** *Every weak-$*$ limit point ν^{prim} of the sequence of counting measures $\{\nu_L^{prim}\}_{L>0}$ is of the form*

$$\nu^{\mathrm{prim}} = c \cdot \nu$$

for some constant $c \geq 0$, where ν denotes the standard Lebesgue measure on $\mathbb{R}^2$.

Idea of the Proof.

- First, from Exercise 2.3, we know that any limit measure ν^{prim} must be $\mathrm{SL}(2,\mathbb{Z})$-invariant.

- Next, from Theorem 2.5 (ergodicity of Lebesgue measure ν under $\mathrm{SL}(2,\mathbb{Z})$), we see that if ν^{prim} is absolutely continuous with respect to ν, then by Exercise 2.4 we must have $\nu^{\text{prim}} = c\,\nu$.

- The main challenge is to show absolute continuity: i.e., if $\nu(A) = 0$ for some Borel set A, then $\nu^{\text{prim}}(A) = 0$.

2.4.3 Proof of Proposition 2.28

Proof. Let ν^{prim} be a weak-$*$ limit point of $\{\nu_L^{\text{prim}}\}$. By Exercise 2.3, ν^{prim} is $\mathrm{SL}(2,\mathbb{Z})$-invariant. By Theorem 2.5, the Lebesgue measure ν on $\mathbb{R}^2$ is ergodic under $\mathrm{SL}(2,\mathbb{Z})$. According to Exercise 2.4, to conclude $\nu^{\text{prim}} = c\,\nu$, it suffices to show ν^{prim} is absolutely continuous with respect to ν. In other words, we must show that if $A \subset \mathbb{R}^2$ is a Borel set with $\nu(A) = 0$, then $\nu^{\text{prim}}(A) = 0$.

Step 1: Covering by small squares. Let $A \subseteq \mathbb{R}^2$ be such that $\nu(A) = 0$. Fix any $\delta > 0$. Because ν (Lebesgue measure) is *outer regular*, we can find a countable collection of open squares $\{B_i\}_{i\in\mathbb{N}}$ covering A such that

$$A \;\subseteq\; \bigcup_{i\in\mathbb{N}} B_i \quad \text{and} \quad \sum_{i\in\mathbb{N}} \nu(B_i) \;\leq\; \delta. \tag{2.6}$$

Step 2: Estimating the measure of squares under ν_L^{prim}. To control $\nu^{\text{prim}}(A)$, we first estimate $\nu_L^{\text{prim}}(B)$ for an arbitrary open square $B \subset \mathbb{R}^2$. Write $\epsilon > 0$ for the side length of B. Observe that

$$\nu_L^{\text{prim}}(B) \;\leq\; \nu_L(B) \;=\; \frac{1}{L^2}\,\#\{\mathbb{Z}^2 \cap (L\cdot B)\}, \tag{2.7}$$

where $L \cdot B$ is an open square of side length $L\epsilon$. Assume $L\epsilon \geq 1$ so that $L \cdot B$ is large enough to contain some lattice points. We construct the set $S \subset \mathbb{R}^2$ by placing disjoint open squares of side length $\frac{1}{2}$ centered at each integer point $v \in \mathbb{Z}^2 \cap (L \cdot B)$. Thus

$$S \;\subseteq\; B'$$

for some open square $B' \subset \mathbb{R}^2$ of side length $L\epsilon + 1$. Therefore,

$$\nu(S) \;\leq\; \nu(B') \;=\; (L\epsilon + 1)^2 \;\leq\; 4L^2\epsilon^2 \;=\; 4L^2\,\nu(B).$$

On the other hand, each of these disjoint sub-squares of side length $\frac{1}{2}$ has area $\frac{1}{4}$. Hence

$$\sum_{v\in\mathbb{Z}^2\cap(L\cdot B)} \frac{1}{4} \;\leq\; \nu(S).$$

It follows that
$$\frac{\#\{\mathbb{Z}^2 \cap (L \cdot B)\}}{4} \;\le\; \nu(S) \;\le\; 4L^2\,\nu(B).$$
Combining this with (**??**) gives
$$\nu_L^{\mathrm{prim}}(B) \;\le\; \nu_L(B) \;=\; \frac{\#(\mathbb{Z}^2 \cap (L \cdot B))}{L^2} \;\le\; \frac{16\,L^2\,\nu(B)}{L^2} \;=\; 16\,\nu(B).$$
Taking $\limsup$ as $L \to \infty$,
$$\limsup_{L\to\infty} \nu_L^{\mathrm{prim}}(B) \;\le\; 16\,\nu(B). \tag{2.9}$$

Step 3: Passing to the weak-$*$ limit. Recall that ν^{prim} is a weak-$*$ limit of ν_L^{prim}. By the subadditivity of measures,
$$\nu^{\mathrm{prim}}(A) \;\le\; \sum_{i\in\mathbb{N}} \nu^{\mathrm{prim}}(B_i). \tag{2.10}$$
From Portmanteau's theorem (Theorem 2.6), for each B_i one obtains
$$\nu^{\mathrm{prim}}(B_i) \;\le\; \liminf_{L\to\infty} \nu_L^{\mathrm{prim}}(B_i) \;\le\; \limsup_{L\to\infty} \nu_L^{\mathrm{prim}}(B_i). \tag{2.11}$$
Then (**??**) says
$$\limsup_{L\to\infty} \nu_L^{\mathrm{prim}}(B_i) \;\le\; 16\,\nu(B_i). \tag{2.12}$$
Hence,
$$\nu^{\mathrm{prim}}(B_i) \;\le\; 16\,\nu(B_i).$$
Summing over $i \in \mathbb{N}$ and using $\sum_i \nu(B_i) \le \delta$ from (**??**) gives
$$\nu^{\mathrm{prim}}(A) \;\le\; \sum_{i\in\mathbb{N}} 16\,\nu(B_i) \;\le\; 16\,\delta.$$

Since $\delta > 0$ is arbitrary, we deduce $\nu^{\mathrm{prim}}(A) = 0$. This proves that ν^{prim} is absolutely continuous with respect to ν.

Step 4: Concluding $\nu^{\mathrm{prim}} = c\nu$. Because ν^{prim} is $\mathrm{SL}(2,\mathbb{Z})$-invariant and absolutely continuous with respect to the $\mathrm{SL}(2,\mathbb{Z})$-ergodic measure ν, Exercise 2.4 implies there is a constant $c \ge 0$ such that $\nu^{\mathrm{prim}} = c\,\nu$. This completes the proof. □

Example 2.29 (Interpretation in Counting Primitive Points)**.** The conclusion $\nu^{\mathrm{prim}} = c \cdot \nu$ means that, in the limit, the distribution of scaled primitive points in $\mathbb{R}^2$ looks *exactly* like a constant multiple of the Lebesgue measure. By evaluating both sides on the unit ball, one sees that
$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} \;=\; \lim_{L\to\infty} \nu_L^{\mathrm{prim}}(\{\|x\| \le 1\}) \;=\; \nu^{\mathrm{prim}}(\{\|x\| \le 1\}) \;=\; c\,\nu(\{\|x\| \le 1\}) \;=\; c\pi.$$
Hence $p(\mathbb{Z}^2, L) \sim c\pi\,L^2$. One can calculate $c = \frac{6}{\pi^2}$ by other means (e.g., geometry of numbers), thereby obtaining the classical formula
$$p(\mathbb{Z}^2, L) \;\sim\; \frac{6}{\pi}\,L^2.$$

2.4.4 Summary and Outlook

Using a combination of group actions, ergodicity arguments, and measure-theoretic methods (like Portmanteau's theorem), we have shown that any weak-$*$ limit of the primitive counting measures must be absolutely continuous with respect to the Lebesgue measure and hence must be a constant multiple of it. This result underpins many deeper counting theorems, showing that the limiting distribution of scaled primitive lattice points (or analogous objects in geometry) is uniform up to a factor.

2.5 Unimodular Lattices and Averaging Arguments

In earlier sections, we analyzed the distribution of *primitive* vectors in the integer lattice $\mathbb{Z}^2$. Our next goal is to extend these ideas to more general *unimodular lattices* in $\mathbb{R}^2$ and to employ a powerful averaging argument to show that a certain constant (arising in our counting measure analysis) is both positive and independent of the particular limit measure under consideration. In other words, we wish to show that *all* limit points (in the weak-$*$ sense) of the sequence of counting measures for primitive lattice points have the *same* coefficient when expressed as a multiple of Lebesgue measure.

2.5.1 Definitions and Geometry of Unimodular Lattices

Definition 2.30 (Lattice)**.** A *lattice* $\Lambda \subseteq \mathbb{R}^2$ is a subgroup of $\mathbb{R}^2$ (viewed as a group under addition) such that Λ is generated by two linearly independent vectors. Equivalently, there exist $\alpha, \beta \in \mathbb{R}^2$ spanning $\mathbb{R}^2$ such that

$$\Lambda = \{m\alpha + n\beta \mid m, n \in \mathbb{Z}\}.$$

We say α, β form an $\mathbb{R}$-*basis* for $\mathbb{R}^2$ and a $\mathbb{Z}$-*basis* for Λ.

Definition 2.31 (Marking of a Lattice)**.** A *marking* of a lattice Λ is a choice of a positively oriented $\mathbb{R}$-basis (v_1, v_2) of $\mathbb{R}^2$ such that

$$\Lambda = \operatorname{span}_{\mathbb{Z}}(v_1, v_2).$$

Positively oriented means $\det(v_1, v_2) > 0$, where det is the usual determinant in $\mathbb{R}^2$.

Definition 2.32 (Covolume and Unimodular Lattice)**.** If $\Lambda = \operatorname{span}_{\mathbb{Z}}(v_1, v_2)$, we define its *covolume* by

$$\operatorname{covol}(\Lambda) := \big|\det(v_1, v_2)\big|.$$

This quantity does *not* depend on the choice of the marking (v_1, v_2). A lattice Λ is called *unimodular* if $\operatorname{covol}(\Lambda) = 1$.

Example 2.33. $\mathbb{Z}^2$ itself is a classic example of a unimodular lattice. One can generate $\mathbb{Z}^2$ by the standard basis vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$, and $\det(e_1, e_2) = 1$. More generally, if $A \in \mathrm{SL}(2, \mathbb{R})$, then $A \cdot \mathbb{Z}^2$ is also a unimodular lattice in $\mathbb{R}^2$.

2.5.2 Transitive Action of $\mathrm{SL}(2,\mathbb{R})$ on Unimodular Lattices

Definition 2.34 ($\mathrm{SL}(2,\mathbb{R})$)**.** The group

$$\mathrm{SL}(2,\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}_{2\times 2}(\mathbb{R}) \;\middle|\; a,b,c,d \in \mathbb{R},\ ad-bc=1 \right\}$$

acts on $\mathbb{R}^2$ by matrix multiplication. If $A \in \mathrm{SL}(2,\mathbb{R})$ and Λ is a unimodular lattice, then $A \cdot \Lambda := \{Av : v \in \Lambda\}$ is also unimodular.

Example 2.35. If $\Lambda = \mathrm{span}_{\mathbb{Z}}(v_1, v_2)$ with $\det(v_1, v_2) = 1$, then by writing the coordinates of v_1, v_2 as columns of a matrix A, one obtains $A \in \mathrm{SL}(2,\mathbb{R})$ and $\Lambda = A \cdot \mathbb{Z}^2$.

2.5.3 Averaging Over the Space of Unimodular Lattices

Definition 2.36 (Counting function for primitive vectors)**.** Let Λ be a unimodular lattice in $\mathbb{R}^2$. Define the counting function

$$p(\Lambda, L) := \#\{v \in \Lambda_{\mathrm{prim}} \mid \|v\| \le L\}, \tag{2.13}$$

where Λ_{prim} is the set of *primitive* vectors of Λ, i.e., those $v \in \Lambda$ that cannot be written as a nontrivial positive integer multiple of some other $w \in \Lambda$.

Example 2.37. For $\Lambda = \mathbb{Z}^2$, the function $p(\Lambda, L)$ counts the number of primitive integer points inside a disk of radius L. We already know from geometry of numbers that

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2 \quad \text{as } L \to \infty.$$

For general unimodular Λ, the asymptotics of $p(\Lambda, L)$ is surprisingly similar when averaged over all unimodular lattices.

Proposition 2.38 (Siegel's Integration Formula)**.** *If $\hat{\mu}$ is the canonical measure on $\mathcal{M}_1$, arising from the pushforward of $\frac{dx\,dy}{y^2}$ in $\mathbb{H}^2$, then*

$$\int_{\mathcal{M}_1} p(\Lambda, L)\, d\hat{\mu}(\Lambda) = 2L^2.$$

Remark 2.39. This result tells us that, on average, a unimodular lattice Λ has about $2L^2$ primitive points inside a disk of radius L. The classical result for $\mathbb{Z}^2$ is consistent with this when accounting for constant factors.

References

[1] C. L. Siegel, *Lectures on the Geometry of Numbers*. Springer-Verlag, 1945.

2.5.4 Uniform Controls and the Function $u(\Lambda)$

When proving precise asymptotics, one also needs some uniform estimates. For instance, we define

$$u(\Lambda) \ := \ \sup_{\substack{v \in \Lambda \\ v \neq 0}} \frac{1}{\|v\|}.$$

Equivalently, $u(\Lambda)$ is the reciprocal of the minimal length of nonzero vectors in Λ. One can show $u(\Lambda)$ is typically finite and one obtains integrability statements $\int_{\mathcal{M}_1} u(\Lambda)\, d\hat{\mu}(\Lambda) < \infty$. This fact is crucial for applying dominated convergence arguments later on.

2.5.5 Summary and Outlook

We have introduced:

- The space $\mathcal{M}_1$ of unimodular lattices in $\mathbb{R}^2$, identified with $\mathrm{SL}(2, \mathbb{R})/\mathrm{SL}(2, \mathbb{Z})$ (up to rotation).
- A canonical measure $\hat{\mu}$ on this space (arising from the pushforward of $\frac{dx\,dy}{y^2}$ in the upper half-plane $\mathbb{H}^2$).
- A general counting function $p(\Lambda, L)$ that mirrors the classical setting $p(\mathbb{Z}^2, L)$.
- Siegel's integration formula showing that $\int_{\mathcal{M}_1} p(\Lambda, L)\, d\hat{\mu}(\Lambda)$ is explicitly computable.

The main takeaway is that averaging over all unimodular lattices simplifies counting questions and yields uniform estimates. In subsequent steps (and sections), one uses these averaging arguments to refine asymptotics for primitive points and to ensure consistency for all possible weak-$*$ limit measures we encounter.

References

[1] C. L. Siegel, *Lectures on the Geometry of Numbers.* Springer-Verlag, 1945.

Exercise 2.10. Show there exists a constant $C > 0$ such that for every $\Lambda \in \mathcal{M}_1$ and every $L > 0$,

$$p(\Lambda, L) \leq C \cdot L^2 \cdot u(\Lambda).$$

Additionally, show that the function $u : \mathcal{M}_1 \to \mathbb{R}$ is integrable with respect to the measure $\hat{\mu}$, i.e.,

$$\int_{\mathcal{M}_1} u(\Lambda)\, d\hat{\mu}(\Lambda) < \infty.$$

Below is a step-by-step, "ultra-detailed" solution to **Exercise 2.10**, aimed at making every aspect accessible to undergraduates. We first restate the problem in our own words and then provide the detailed proof.

Restatement of the Exercise

Let $\mathcal{M}_1$ be the set of all *unimodular* lattices in $\mathbb{R}^2$. (Recall that a lattice $\Lambda \subseteq \mathbb{R}^2$ is *unimodular* if its fundamental parallelogram has area 1.) We denote a typical element of $\mathcal{M}_1$ by Λ. Two functions of interest are:

1. $p(\Lambda, L)$: the number of *primitive* lattice vectors $v \in \Lambda_{\text{prim}}$ such that $\|v\| \leq L$. (A vector $v \in \Lambda$ is called *primitive* if it is not an integer multiple $m\,w$ with $|m| > 1$ and $w \in \Lambda$.)

2. $u(\Lambda)$: defined by
$$u(\Lambda) \;:=\; \sup_{\substack{v \in \Lambda \\ v \neq 0}} \frac{1}{\|v\|}.$$
Equivalently, $u(\Lambda) = 1/\lambda_1(\Lambda)$ where $\lambda_1(\Lambda)$ is the length of the shortest nonzero vector in Λ.

We wish to prove two statements:

1. **Bound on $p(\Lambda, L)$:** There exists a universal constant $C > 0$ such that for every $\Lambda \in \mathcal{M}_1$ and every $L > 0$,
$$p(\Lambda, L) \;\leq\; C\, L^2\, u(\Lambda).$$

2. **Integrability of $u(\Lambda)$:** The function $u : \mathcal{M}_1 \to \mathbb{R}$ is integrable with respect to the measure $\hat{\mu}$ on $\mathcal{M}_1$; that is,
$$\int_{\mathcal{M}_1} u(\Lambda)\, d\hat{\mu}(\Lambda) \;<\; \infty.$$

Here, $\hat{\mu}$ denotes the natural (modular/Weil–Petersson–Siegel type) measure on $\mathcal{M}_1$.

Part A. Proof of the Inequality $p(\Lambda, L) \leq C\, L^2\, u(\Lambda)$

1. Geometric Intuition

- Since $\Lambda \subset \mathbb{R}^2$ is unimodular, the area of a fundamental parallelogram is 1.
- The function $u(\Lambda) = \sup_{v \neq 0} \frac{1}{\|v\|}$ is the reciprocal of the length of the shortest nonzero vector in Λ (i.e. $u(\Lambda) = 1/\lambda_1(\Lambda)$). Thus, if Λ has a very short vector, then $u(\Lambda)$ is large.
- Our goal is to count the number $p(\Lambda, L)$ of primitive lattice vectors in Λ that lie in the disk
$$D_L = \{x \in \mathbb{R}^2 : \|x\| \leq L\}.$$

2. Upper Bound on the Total Number of Lattice Points

A classical fact from the geometry of numbers is that for a unimodular lattice Λ, the total number of lattice points (not necessarily primitive) inside D_L is of order the area of D_L, namely πL^2. That is, there exists a constant $c_1 > 0$ such that
$$\#\big(\Lambda \cap D_L\big) \;\leq\; c_1\, L^2 \quad \text{for all } L > 0.$$

3. Passing to Primitive Vectors and the Role of $u(\Lambda)$

Every nonzero vector $v \in \Lambda$ can be written uniquely as $v = m\,w$, where $w \in \Lambda_{\text{prim}}$ is primitive and $m \in \mathbb{Z}\setminus\{0\}$. If $\|v\| \le L$, then for the corresponding primitive vector w we have $\|w\| \le L$ and the maximum integer m (in absolute value) such that $\|m\,w\| \le L$ is roughly $\lfloor L/\|w\| \rfloor$. Thus, one may express

$$\#(\Lambda \cap D_L) \;\approx\; \sum_{\substack{w\in\Lambda_{\text{prim}}\\ \|w\|\le L}} \frac{L}{\|w\|}.$$

Since $\#(\Lambda \cap D_L) \le c_1 L^2$, it follows that

$$c_1\,L^2 \;\ge\; L \sum_{\substack{w\in\Lambda_{\text{prim}}\\ \|w\|\le L}} \frac{1}{\|w\|}.$$

While this inequality does not immediately yield the desired bound, it indicates that the number of primitive vectors is controlled by how small the vectors can be.

4. Refinement Using a Case Analysis

We consider two cases depending on the length of the shortest vector $\lambda_1(\Lambda)$:

Case 1: $\lambda_1(\Lambda) \ge L/2$.

In this case, every nonzero vector in Λ satisfies $\|v\| \ge L/2$. Thus, if $\|v\| \le L$, then the vectors lie in a thin shell between $L/2$ and L. The number of lattice points in such a shell is at most of order L^2. Moreover,

$$u(\Lambda) = \frac{1}{\lambda_1(\Lambda)} \le \frac{2}{L}.$$

Hence,

$$L^2\,u(\Lambda) \le L^2 \cdot \frac{2}{L} = 2L.$$

Since $p(\Lambda, L)$ is bounded by the total number of lattice points (which is at most $c_2\,L^2$ for some $c_2 > 0$), we obtain

$$p(\Lambda, L) \le c_2\,L^2 \le 2c_2\,L^2\,u(\Lambda).$$

Case 2: $\lambda_1(\Lambda) < L/2$.

Let $\delta := \lambda_1(\Lambda)$; then $u(\Lambda) = 1/\delta$ and $\delta < L/2$. A standard Minkowski-type argument shows that

$$\#\big(\Lambda \cap D_L\big) \;\le\; c_3\,\frac{L^2}{\delta^2}$$

for some constant $c_3 > 0$. Since every vector in $\Lambda \cap D_L$ is a multiple of a primitive vector, we have

$$p(\Lambda, L) \le \#\big(\Lambda \cap D_L\big) \;\le\; c_3\,\frac{L^2}{\delta^2} = c_3\,L^2\,u(\Lambda)^2.$$

In many standard treatments one can refine this estimate to show that the bound can be improved to

$$p(\Lambda, L) \le C\,L^2\big(1 + u(\Lambda)\big),$$

which in turn implies (for large $u(\Lambda)$)

$$p(\Lambda, L) \leq 2C\, L^2\, u(\Lambda).$$

Thus, by appropriately adjusting constants, we obtain a universal constant $C > 0$ such that

$$\boxed{p(\Lambda, L) \leq C\, L^2\, u(\Lambda) \quad \text{for all } \Lambda \in \mathcal{M}_1,\ L > 0.}$$

Part B. Proof that $\int_{\mathcal{M}_1} u(\Lambda)\, d\hat{\mu}(\Lambda) < \infty$

1. Reformulation

Recall that

$$u(\Lambda) = \frac{1}{\lambda_1(\Lambda)},$$

where $\lambda_1(\Lambda)$ is the length of the shortest nonzero vector in Λ. We wish to prove

$$\int_{\mathcal{M}_1} \frac{1}{\lambda_1(\Lambda)}\, d\hat{\mu}(\Lambda) < \infty.$$

2. Measure Estimate on the Set of Lattices with Short Vectors

A classical result in the geometry of numbers states that for unimodular lattices in $\mathbb{R}^2$, the measure of the set

$$E_\varepsilon := \{\Lambda \in \mathcal{M}_1 : \lambda_1(\Lambda) < \varepsilon\}$$

satisfies

$$\hat{\mu}(E_\varepsilon) \leq c\, \varepsilon^2$$

for some constant $c > 0$.

3. The Layer-Cake Representation

Recall the layer-cake (or Cavalieri) formula for any nonnegative measurable function f:

$$\int f(\Lambda)\, d\hat{\mu}(\Lambda) = \int_0^\infty \hat{\mu}\Big\{\Lambda : f(\Lambda) > t\Big\}\, dt.$$

Setting $f(\Lambda) = u(\Lambda) = 1/\lambda_1(\Lambda)$, we have

$$\int_{\mathcal{M}_1} \frac{1}{\lambda_1(\Lambda)}\, d\hat{\mu}(\Lambda) = \int_0^\infty \hat{\mu}\Big\{\Lambda : \frac{1}{\lambda_1(\Lambda)} > t\Big\}\, dt.$$

Note that

$$\Big\{\Lambda : \frac{1}{\lambda_1(\Lambda)} > t\Big\} = \{\Lambda : \lambda_1(\Lambda) < 1/t\}.$$

Thus,

$$\int_{\mathcal{M}_1} \frac{1}{\lambda_1(\Lambda)}\, d\hat{\mu}(\Lambda) = \int_0^\infty \hat{\mu}\Big(\{\Lambda : \lambda_1(\Lambda) < 1/t\}\Big)\, dt.$$

4. Splitting the Integral

For $t \geq 1$, we have $1/t \leq 1$, and using the estimate from step 2:

$$\hat{\mu}\Big(\{\Lambda : \lambda_1(\Lambda) < 1/t\}\Big) \leq c\,(1/t)^2 = \frac{c}{t^2}.$$

Thus,

$$\int_1^\infty \hat{\mu}\Big(\{\Lambda : \lambda_1(\Lambda) < 1/t\}\Big)\,dt \leq \int_1^\infty \frac{c}{t^2}\,dt = c\left[-\frac{1}{t}\right]_1^\infty = c.$$

For $0 < t < 1$, the set $\{\Lambda : \lambda_1(\Lambda) < 1/t\}$ is *all* of $\mathcal{M}_1$ (since $1/t > 1$ and every unimodular lattice has a shortest vector of length at most some universal constant). Hence, for $0 < t < 1$,

$$\hat{\mu}\Big(\{\Lambda : \lambda_1(\Lambda) < 1/t\}\Big) \leq \hat{\mu}(\mathcal{M}_1),$$

which is finite. Therefore, the contribution from $t \in (0, 1)$ is finite.

Combining these, we deduce

$$\int_{\mathcal{M}_1} \frac{1}{\lambda_1(\Lambda)}\,d\hat{\mu}(\Lambda) < \infty,$$

or equivalently,

$$\boxed{\int_{\mathcal{M}_1} u(\Lambda)\,d\hat{\mu}(\Lambda) < \infty.}$$

Concluding Remarks

- The factor $u(\Lambda)$ appears in the bound on $p(\Lambda, L)$ because if a lattice Λ has a very short nonzero vector, then many of its multiples (which are non-primitive) can lie in the ball D_L. The bound $p(\Lambda, L) \leq C\,L^2\,u(\Lambda)$ captures this dependence on the length of the shortest vector.
- The integrability of $u(\Lambda)$ follows from the fact that the set of lattices having very short vectors (i.e. with $\lambda_1(\Lambda)$ very small) occupies only a small part of the moduli space $\mathcal{M}_1$, with measure decaying like ε^2 as $\varepsilon \to 0$.

$$\boxed{\text{Hence, both parts of Exercise 2.10 are proved in full detail.}}$$

2.6 Convergence and Compactness of Counting Measures

Before we prove Theorem 2.1 concerning the asymptotic count of primitive lattice points, it is essential to establish some foundational concepts about convergence and compactness for measures. In this section, we introduce the relevant definitions and results on weak-$*$ convergence and compactness in the space of measures, and we provide examples and proof sketches that illustrate these ideas.

2.6.1 Weak-$*$ Convergence of Measures

Definition 2.40 (Weak-$*$ Convergence)**.** Let X be a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ a sequence of (locally finite) Borel measures on X. We say that μ_n *converges in the weak-$*$ topology* (or *vaguely*) to a measure μ if for every continuous function $f : X \to \mathbb{R}$ with compact support,

$$\lim_{n\to\infty} \int_X f\, d\mu_n \;=\; \int_X f\, d\mu.$$

We denote this by $\mu_n \rightharpoonup \mu$.

Example 2.41. Consider the sequence of measures on $\mathbb{R}$ given by

$$\mu_n = \frac{1}{n}\sum_{k=1}^{n} \delta_{k/n},$$

where δ_x is the Dirac measure at x. For any continuous function f with compact support in $[0, 1]$,

$$\int_{\mathbb{R}} f\, d\mu_n = \frac{1}{n}\sum_{k=1}^{n} f(k/n).$$

As $n \to \infty$, this Riemann sum converges to $\int_0^1 f(x)\, dx$. Hence, $\mu_n \rightharpoonup \nu$, where ν is the Lebesgue measure on $[0, 1]$.

2.6.2 Portmanteau's Theorem

Weak-$*$ convergence can be characterized in several equivalent ways. One of the most useful is given by Portmanteau's theorem.

Theorem 2.42 (Portmanteau's Theorem)**.** *Let X be a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ a sequence of locally finite Borel measures on X that converge in the weak-$*$ topology to a Borel measure μ. Then:*

1. *For every open set $U \subseteq X$,*
$$\mu(U) \le \liminf_{n\to\infty} \mu_n(U).$$
2. *For every closed set $F \subseteq X$,*
$$\mu(F) \ge \limsup_{n\to\infty} \mu_n(F).$$
3. *For every compact set $K \subset X$ satisfying $\mu(\partial K) = 0$,*
$$\lim_{n\to\infty} \mu_n(K) = \mu(K).$$

Remark 2.43. These alternative characterizations are particularly useful when one wants to check convergence on a class of sets (open, closed, or compact sets with negligible boundary) rather than on all continuous test functions.

Sketch of Proof. The full proof can be found in many standard texts on measure theory. Here is an outline:

- To prove (1), one approximates the indicator function of an open set U from below by continuous functions with compact support and applies the definition of weak-$*$ convergence.
- Statement (2) follows by applying (1) to the complements of closed sets and using properties of measures.
- For (3), if K is compact and $\mu(\partial K) = 0$, then the measure of K can be approximated by those of open sets containing K and closed sets contained in K. Then (1) and (2) imply the desired equality.

□

Example 2.44. Let $X = \mathbb{R}$ and consider the measures μ_n from the previous example. Let $K = [0, 1]$, which is compact and has boundary $\{0, 1\}$ with Lebesgue measure zero. By Portmanteau's theorem,

$$\lim_{n\to\infty} \mu_n([0, 1]) = \nu([0, 1]) = 1.$$

2.6.3 Compactness and Convergence of Measures

A central theme in our study is the use of compactness criteria to extract convergent subsequences of measures. The idea is that if a sequence of measures does not "blow up" on compact sets, then one can apply compactness arguments (often relying on the Banach–Alaoglu theorem) to obtain a convergent subsequence in the weak-$*$ topology.

Proposition 2.45 (Sequential Compactness of Measures)**.** *Let X be a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of locally finite Borel measures on X such that for every compact set $K \subseteq X$ the sequence $\{\mu_n(K)\}$ is uniformly bounded. Then there exists a subsequence $\{\mu_{n_k}\}$ and a locally finite Borel measure μ on X such that*

$$\mu_{n_k} \rightharpoonup \mu \quad \textit{(weak-$*$ convergence).}$$

Sketch of Proof. The idea is to use the Banach–Alaoglu theorem. The set of all Radon measures on a locally compact space X (endowed with the vague topology) is the dual of the space of continuous functions with compact support. Uniform boundedness on compact sets implies that the sequence is contained in a weak-$*$ compact subset of the dual space. Hence, there exists a convergent subsequence. □

Example 2.46. Consider the sequence of counting measures on the integer lattice given by

$$\nu_n := \frac{1}{n^2} \sum_{v\in\mathbb{Z}^2} \delta_{v/n}.$$

One can show that for any fixed compact set $K \subset \mathbb{R}^2$, the measures $\nu_n(K)$ are uniformly bounded. Therefore, by the above proposition, there exists a subsequence ν_{n_k} converging (in the weak-$*$ sense) to some measure μ. One can further show (by explicit computation) that the unique possible limit is the Lebesgue measure.

2.6.4 Summary and Relevance

The concepts of weak-$*$ convergence and compactness for measures are central in modern analysis and ergodic theory. In the context of counting problems, we are interested in the behavior of a sequence of counting measures such as

$$\nu_L^{\text{prim}} = \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L}.$$

By understanding the weak-$*$ convergence of such sequences, we can deduce asymptotic formulas for the number of primitive lattice points and eventually apply similar ideas to count closed geodesics on hyperbolic surfaces.

Portmanteau's theorem gives us practical tools for verifying weak-$*$ convergence by testing on open, closed, or compact sets, while compactness criteria guarantee that we can extract convergent subsequences from any uniformly bounded sequence of measures. These ideas form the backbone of our later arguments and provide the foundation for further results.

Exercise 2.11. Let X be a metric space. Show that a sequence $(x_L)_{L>0}$ in X converges to $x \in X$ if and only if every subsequence of $(x_L)_{L>0}$ has a subsequence converging to x.

Exercise 2.11

Statement. Let X be a metric space. Show that a sequence $(x_n)_{n=1}^{\infty}$ in X converges to $x \in X$ if and only if every subsequence of $(x_n)_{n=1}^{\infty}$ has a subsequence converging to x.

For clarity, we replace the index L by n so that our sequence is $\{x_n\}_{n=1}^{\infty}$. The proof will be divided into two parts.

Definitions and Notation

1. A sequence $\{x_n\}$ in a metric space (X, d) is said to *converge* to $x \in X$ if for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$,

$$d(x_n, x) < \varepsilon.$$

   We write $\lim_{n\to\infty} x_n = x$.

2. A *subsequence* of $\{x_n\}$ is a sequence of the form $\{x_{n_k}\}$ where $1 \leq n_1 < n_2 < n_3 < \cdots$.

Proof

We will prove the equivalence in two directions.

(1) If $\{x_n\}$ converges to x, then every subsequence has a further subsequence converging to x.

Proof: Assume $\lim_{n\to\infty} x_n = x$. Then, by the definition of convergence, for every $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that for all $n \geq N$,

$$d(x_n, x) < \varepsilon.$$

Let $\{x_{n_k}\}$ be any subsequence of $\{x_n\}$. Because $\{n_k\}$ is strictly increasing, there exists some index K such that for all $k \geq K$ we have $n_k \geq N$. Therefore, for all $k \geq K$,

$$d(x_{n_k}, x) < \varepsilon.$$

This shows that the subsequence $\{x_{n_k}\}$ itself converges to x. In particular, it has a subsequence (for example, itself) that converges to x.

(2) If every subsequence of $\{x_n\}$ has a further subsequence converging to x, then $\{x_n\}$ converges to x.

Proof: We prove the contrapositive. Suppose that $\{x_n\}$ does *not* converge to x. Then, by definition, there exists some $\varepsilon_0 > 0$ such that for every $N \in \mathbb{N}$ there exists an $n \geq N$ with

$$d(x_n, x) \geq \varepsilon_0.$$

Using this property, we can construct a subsequence $\{x_{n_k}\}$ as follows:

- Choose n_1 such that $d(x_{n_1}, x) \geq \varepsilon_0$.
- Given n_k, choose $n_{k+1} > n_k$ such that
$$d(x_{n_{k+1}}, x) \geq \varepsilon_0.$$

Thus, the subsequence $\{x_{n_k}\}$ satisfies

$$d(x_{n_k}, x) \geq \varepsilon_0 \quad \text{for all } k.$$

Consequently, every term of $\{x_{n_k}\}$ remains at least ε_0 away from x. It follows that no subsequence of $\{x_{n_k}\}$ can converge to x because the distance from x never drops below ε_0.

Hence, we have found a subsequence of $\{x_n\}$ (namely, $\{x_{n_k}\}$) that does not contain any further subsequence converging to x.

We have shown that:

- If $\{x_n\}$ converges to x, then every subsequence of $\{x_n\}$ converges to x, and in particular, has a subsequence converging to x.
- Conversely, if there is a subsequence of $\{x_n\}$ that does not have any sub-subsequence converging to x, then $\{x_n\}$ cannot converge to x.

Thus, a sequence $\{x_n\}$ converges to x if and only if every subsequence of $\{x_n\}$ has a further subsequence converging to x.

$$\text{A sequence } \{x_n\} \text{ converges to } x$$
$$\text{if and only if every subsequence of } \{x_n\} \text{ has a subsequence converging to } x.$$

2.7 Convergence, Compactness, and Equidistribution of Primitive Lattice Point Measures

In our study of counting problems for primitive lattice points in $\mathbb{R}^2$, we define the sequence of counting measures

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

where δ_x is the Dirac measure at x and

$$\mathbb{Z}^2_{\text{prim}} = \{(a,b) \in \mathbb{Z}^2 : \gcd(a,b) = 1\}.$$

Our ultimate goal is to show that as $L \to \infty$ the measures ν_L^{prim} converge (in the weak-$*$ topology) to a constant multiple of the Lebesgue measure ν on $\mathbb{R}^2$. In this section we discuss the compactness and convergence criteria for measures and then use these ideas to prove an *equidistribution* result for primitive lattice points.

2.7.1 Weak-$*$ Convergence and Compactness

Definition 2.47 (Weak-$*$ Convergence)**.** Let X be a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of locally finite Borel measures on X. We say that μ_n converges in the *weak-$*$ topology* (or *vaguely*) to a measure μ if for every continuous function $f : X \to \mathbb{R}$ with compact support,

$$\lim_{n\to\infty} \int_X f \, d\mu_n \; = \; \int_X f \, d\mu.$$

We write $\mu_n \rightharpoonup \mu$.

Example 2.48. Let $X = \mathbb{R}$ and define, for each $n \in \mathbb{N}$,

$$\mu_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{k/n}.$$

Then, for any continuous function f with support in $[0,1]$,

$$\int_{\mathbb{R}} f \, d\mu_n = \frac{1}{n} \sum_{k=1}^{n} f(k/n),$$

which is a Riemann sum for $\int_0^1 f(x)\,dx$. Hence, $\mu_n \rightharpoonup \nu$, where ν is the Lebesgue measure on $[0,1]$.

A key tool for extracting convergent subsequences in the space of measures is a compactness criterion, which is a consequence of the Banach–Alaoglu theorem.

Theorem 2.49 (Compactness Criterion for Measures)**.** *Let X be a metric space and $\{\mu_n\}_{n\in\mathbb{N}}$ be a sequence of locally finite Borel measures on X. Suppose that for every compact set $K \subseteq X$ the sequence $\{\mu_n(K)\}$ is bounded. Then there exists a subsequence $\{\mu_{n_k}\}$ and a locally finite Borel measure μ on X such that*

$$\mu_{n_k} \rightharpoonup \mu.$$

Remark 2.50. This result follows from the Banach–Alaoglu theorem, since the space of Radon measures on a locally compact space X (the dual of $C_c(X)$, the space of continuous functions with compact support) is weak-$*$ compact when restricted to a bounded set.

Example 2.51. Consider the sequence of measures

$$\nu_n = \frac{1}{n^2} \sum_{v \in \mathbb{Z}^2} \delta_{v/n}$$

on $\mathbb{R}^2$. For any fixed compact set $K \subset \mathbb{R}^2$, the number of points $v \in \mathbb{Z}^2$ such that $v/n \in K$ is bounded by a constant times n^2. Hence, $\nu_n(K)$ is uniformly bounded, and by Theorem 2.49, there exists a subsequence converging weak-$*$ to a locally finite measure.

2.7.2 Equidistribution of Primitive Lattice Points

We now turn our attention to the sequence of counting measures for primitive lattice points,

$$\nu_L^{\text{prim}} = \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L}.$$

Our aim is to prove the following fundamental equidistribution result:

Theorem 2.52 (Equidistribution of Primitive Points)**.** *With respect to the weak-$*$ topology on $\mathbb{R}^2$,*

$$\lim_{L \to \infty} \nu_L^{\text{prim}} = \frac{6}{\pi^2} \cdot \nu,$$

where ν is the standard Lebesgue measure on $\mathbb{R}^2$.

Motivation: This theorem asserts that, when we scale the primitive lattice points by $1/L$ and normalize by $1/L^2$, the resulting distribution becomes uniform on $\mathbb{R}^2$ (up to the constant factor $6/\pi^2$). This phenomenon is known as *equidistribution*. It is a geometric analogue of the fact that primes are "evenly distributed" in a suitable asymptotic sense.

Proof Outline of Theorem 2.52

Proof Sketch. We prove the theorem in several steps:

1. **Extracting Convergent Subsequences:** By Theorem 2.49, any subsequence of $\{\nu_L^{\text{prim}}\}$ has a further subsequence converging in the weak-$*$ topology to some locally finite measure ν^{prim}.

2. **Invariance and Ergodicity:** Exercises (e.g., Exercise 2.3 and Exercise 2.4 in our notes) show that any weak-$*$ limit point ν^{prim} must be invariant under the action of $\mathrm{SL}(2, \mathbb{Z})$. Moreover, Theorem 2.5 (which states that the Lebesgue measure is ergodic under this group action) implies that any such invariant measure is a constant multiple of the Lebesgue measure:

$$\nu^{\text{prim}} = c \cdot \nu \quad \text{for some } c \geq 0.$$

3. **Determining the Constant c:** To show that $c = \frac{6}{\pi^2}$, one computes a certain average. In particular, one shows that

$$\lim_{k\to\infty} \int_{\mathcal{M}_1} \frac{p(\Lambda, L_k)}{L_k^2}\, d\hat{\mu}(\Lambda) = 2, \tag{2.14}$$

where $\mathcal{M}_1$ denotes the space of unimodular lattices (up to rotation) and $p(\Lambda, L)$ counts the number of primitive vectors in Λ of length at most L. On the other hand, a more detailed analysis shows that for every such lattice, one has

$$\lim_{k\to\infty} \frac{p(\Lambda, L_k)}{L_k^2} = c\pi.$$

Averaging over $\mathcal{M}_1$ and applying the dominated convergence theorem, one finds

$$\lim_{k\to\infty} \int_{\mathcal{M}_1} \frac{p(\Lambda, L_k)}{L_k^2}\, d\hat{\mu}(\Lambda) = c \cdot \frac{\pi^2}{3}. \tag{2.15}$$

Equating (2.14) and (2.15) gives

$$c \cdot \frac{\pi^2}{3} = 2,$$

so that

$$c = \frac{6}{\pi^2}.$$

4. **Conclusion:** Since every subsequence of $\{\nu_L^{\text{prim}}\}$ has a further subsequence converging to $\frac{6}{\pi^2}\,\nu$, it follows that the whole sequence converges weak-$*$ to $\frac{6}{\pi^2}\,\nu$.

□

Example 2.53 (Relating to Lattice Point Counting)**.** As an application, note that if we evaluate both sides of the weak-$*$ convergence on the unit disk $B(0,1)$, we obtain

$$\lim_{L\to\infty} \nu_L^{\text{prim}}(B(0,1)) = \frac{6}{\pi^2}\nu(B(0,1)) = \frac{6}{\pi^2} \cdot \pi = \frac{6}{\pi}.$$

Since by definition $\nu_L^{\text{prim}}(B(0,1)) = \frac{p(\mathbb{Z}^2,L)}{L^2}$ when using the standard lattice $\mathbb{Z}^2$, this recovers the classical asymptotic

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi}L^2.$$

2.7.3 Summary and Outlook

In this section, we established two central results:

1. A compactness criterion (Theorem 2.49) for the weak-$*$ topology guarantees that any sequence of locally finite measures, uniformly bounded on compact sets, has a convergent subsequence.

2. Using invariance properties and ergodicity (via Exercises 2.3 and 2.4, and Theorem 2.5), we showed that any weak-$*$ limit of the counting measures ν_L^{prim} must be of the form $c\,\nu$ for some constant $c \geq 0$. An averaging argument over the space of unimodular lattices then shows that $c = \frac{6}{\pi^2}$.

Thus, we conclude that

$$\lim_{L\to\infty} \nu_L^{\text{prim}} = \frac{6}{\pi^2}\,\nu,$$

which is a key step toward proving the asymptotic formula for the number of primitive lattice points in $\mathbb{Z}^2$. Later, in Theorem 2.1, we will deduce the asymptotic

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi} L^2,$$

from this equidistribution result.

Exercise 2.14. Recall the definition of the sequence of counting measures $(\nu_L)_{L>0}$ on $\mathbb{R}^2$ in (2.2). Using the methods introduced in the proof of Theorem 2.13 show that (2.3) holds, i.e., show that with respect to the weak-$*$ topology for measures on $\mathbb{R}^2$,

$$\lim_{L\to\infty} \nu_L = \nu.$$

Exercise 2.14

Below is a step-by-step, *very detailed* explanation of how to prove **Exercise 2.14**. We restate the problem in our own words first and then provide the argument in a clear and structured way for an undergraduate audience.

Restating the Exercise

We have a sequence of *counting measures* $\{\nu_L\}_{L>0}$ on $\mathbb{R}^2$ given by (see (2.2) in the text):

$$\nu_L \;:=\; \frac{1}{L^2} \sum_{v\in\mathbb{Z}^2} \delta_{\frac{v}{L}},$$

where δ_x denotes the Dirac measure at x. Concretely,

- For each $L > 0$, we take every integer vector $v \in \mathbb{Z}^2$,
- We scale it by $1/L$ (so we get the point $\frac{v}{L} \in \mathbb{R}^2$),
- We put a "Dirac mass" $\delta_{\frac{v}{L}}$ at that point,
- And finally multiply by the factor $1/L^2$.

We want to show that ν_L converges *weak-$*$* to the usual Lebesgue measure ν on $\mathbb{R}^2$. In symbolic form:

$$\lim_{L\to\infty} \nu_L \;=\; \nu \quad \text{in the weak-}*\text{ (or vague) topology.}$$

Recall: Weak-$*$ Convergence of Measures

A sequence of finite (or locally finite) Borel measures μ_n converges to a measure μ *in the weak-$*$ sense* (or vaguely) if, for every bounded continuous function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_{\mathbb{R}^2} f \, d\mu_n \;=\; \int_{\mathbb{R}^2} f \, d\mu.$$

Equivalently, we can check this on a convenient collection of test functions, such as continuous functions with compact support.

Outline of the Proof Strategy

1. **Key intuitive idea:** ν_L is basically sampling the plane at a "grid" of spacing $1/L$. Multiplying by $1/L^2$ suggests that in large regions the mass of ν_L in that region approximates the area of the region (since there are about area $\cdot\, L^2$ integer points in a large region scaled by $1/L$). Thus, as $L \to \infty$, these discrete measures should mimic the uniform Lebesgue measure on $\mathbb{R}^2$.

2. **Classical approach:** One shows that every weak-$*$ limit point of $\{\nu_L\}$ must be translation-invariant. (This is often done by "shifting" the lattice $\mathbb{Z}^2$.) Then one identifies that a translation-invariant, locally finite measure in $\mathbb{R}^2$ has to be a constant multiple of Lebesgue measure. Finally, by calculating the measure of some particular region (e.g., the unit square), one sees that the constant must be 1. Hence the limit measure is exactly ν.

3. **Connection to Theorem 2.13:** Theorem 2.13 was about showing another family of counting measures ν_L^{prim} converges to a constant multiple of ν. Its proof uses techniques such as "checking invariance" and "computing volume/area," plus an argument involving Siegel's integration formulas (or related geometry-of-numbers results). By analogy (and simpler reasoning), we can do the same for ν_L.

Below we give a direct, more elementary approach that parallels the arguments in the proof of Theorem 2.13.

Detailed Proof

Step 1. Boundedness of the Measures ν_L

First, note that for any compact set $K \subset \mathbb{R}^2$,

$$\nu_L(K) \;=\; \frac{1}{L^2} \sum_{v\in\mathbb{Z}^2} \delta_{v/L}(K) \;=\; \frac{1}{L^2}\, \#\{v \in \mathbb{Z}^2 : \tfrac{v}{L} \in K\}.$$

As L grows, the number $\#\{v \in \mathbb{Z}^2 : v \in L \cdot K\}$ behaves like $\text{area}(L \cdot K) = L^2\, \text{area}(K)$, up to a small boundary error. Hence, $\nu_L(K)$ approximates area(K). This shows that the sequence $\{\nu_L\}$ is uniformly bounded on compact sets.

Step 2. Translation Invariance in the Limit

We now show that any weak-$*$ limit ν^* of $\{\nu_L\}$ must be translation-invariant.

- For each integer vector $m \in \mathbb{Z}^2$, define the translation map $T_m : \mathbb{R}^2 \to \mathbb{R}^2$ by $T_m(x) = x + m$.
- Observe that for $v \in \mathbb{Z}^2$ and if L is an integer,

$$T_m\Big(\frac{v}{L}\Big) = \frac{v}{L} + m = \frac{v + mL}{L}.$$

  Since $mL \in \mathbb{Z}^2$, the set $\{v + mL : v \in \mathbb{Z}^2\}$ is just a permutation of $\mathbb{Z}^2$. Hence, for such L, we have

$$T_{m*}(\nu_L) = \nu_L.$$

- Even if L is not an integer, by an approximation argument one can show that for large L the effect of integer translations on ν_L is negligible, and any weak-$*$ limit measure ν^* will satisfy

$$T_{m*}(\nu^*) = \nu^* \quad \text{for all } m \in \mathbb{Z}^2.$$

  In fact, by density one may show that ν^* is invariant under all translations in $\mathbb{R}^2$.

Step 3. A Translation-Invariant, Locally Finite Measure Is a Constant Multiple of Lebesgue Measure

It is a standard fact that if μ is a Borel measure on $\mathbb{R}^2$ that is translation-invariant and locally finite (i.e., $\mu(K) < \infty$ for every compact K), then there exists a constant $c \geq 0$ such that

$$\mu = c\,\nu,$$

where ν is the usual Lebesgue measure. Thus any weak-$*$ limit point ν^* of $\{\nu_L\}$ has the form $c\,\nu$.

Step 4. Determining the Constant $c = 1$

To identify the constant c, we compare the measures of a simple set under ν_L and ν. Let

$$Q = [0, 1]^2$$

be the unit square. Then,

$$\nu_L(Q) = \frac{1}{L^2}\,\#\Big\{v \in \mathbb{Z}^2 : \frac{v}{L} \in Q\Big\}.$$

Notice that

$$\Big\{v \in \mathbb{Z}^2 : \frac{v}{L} \in Q\Big\} = \{v = (v_1, v_2) \in \mathbb{Z}^2 : 0 \leq v_1, v_2 \leq L\}.$$

The number of such integer points is approximately $(L+1)^2 \approx L^2$ for large L. Thus,

$$\nu_L(Q) \approx \frac{L^2}{L^2} = 1.$$

Since the Lebesgue measure of Q is $\nu(Q) = 1$, any weak-$*$ limit ν^* must satisfy

$$\nu^*(Q) = 1.$$

If $\nu^* = c\,\nu$, then $\nu^*(Q) = c\,\nu(Q) = c$. Therefore, $c = 1$.

Step 5. Conclusion

Since every weak-$*$ limit point of the sequence $\{\nu_L\}$ is translation-invariant and must equal $c\,\nu$ with $c = 1$, we conclude that

$$\lim_{L\to\infty} \nu_L = \nu$$

in the weak-$*$ topology.

$$\boxed{\lim_{L\to\infty} \nu_L \;=\; \nu.}$$

Key Takeaways

1. **Uniform Lattice Sampling:** The measure ν_L samples $\mathbb{R}^2$ at a grid of spacing $1/L$. As L increases, these discrete points fill the plane more densely.
2. **Normalization:** The factor $1/L^2$ normalizes the total mass in a region of area approximately L^2, so that the mass approximates the area (i.e., the Lebesgue measure).
3. **Translation-Invariance:** Any weak-$*$ limit of the ν_L is translation-invariant and, therefore, must be a constant multiple of the Lebesgue measure.
4. **Determining the Constant:** By evaluating on the unit square, one shows the constant is 1. Thus, the limit measure is exactly the Lebesgue measure ν.

End of Ultra-Detailed Solution.

2.8 Counting Primitive Integer Points in $\mathbb{Z}^2$

In this section, we study the asymptotic behavior of the number of *primitive* lattice points in the Euclidean plane. Recall that a vector in $\mathbb{Z}^2$ is called *primitive* if its coordinates are coprime (that is, their greatest common divisor is 1). Our goal is to prove the following asymptotic estimate.

2.8.1 Statement of the Main Result

Theorem 2.54 (Theorem 2.15)**.** *As $L \to \infty$, the number $p(\mathbb{Z}^2, L)$ of primitive integer points in $\mathbb{Z}^2$ of Euclidean norm at most L satisfies*

$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi}.$$

Equivalently,

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi}\, L^2.$$

2.8.2 Background and Definitions

Primitive Vectors.

Definition 2.55 (Primitive Vector)**.** A vector $v = (a, b) \in \mathbb{Z}^2$ is called *primitive* if $\gcd(a, b) = 1$. In other words, v is not an integer multiple (with a factor greater than 1) of any other vector in $\mathbb{Z}^2$.

Example 2.56. The vector $(3, 5)$ is primitive since $\gcd(3, 5) = 1$, while $(4, 6)$ is not primitive because $\gcd(4, 6) = 2$. We denote by

$$\mathbb{Z}^2_{\text{prim}} = \{(a, b) \in \mathbb{Z}^2 : \gcd(a, b) = 1\}$$

the set of all primitive vectors in $\mathbb{Z}^2$.

Counting Function and Associated Measures. For $L > 0$, we define the counting function:

$$p(\mathbb{Z}^2, L) := \#\{\, v \in \mathbb{Z}^2_{\text{prim}} : \|v\| \leq L\},$$

where $\|v\|$ denotes the Euclidean norm in $\mathbb{R}^2$.

A useful approach is to define the following *normalized counting measure* on $\mathbb{R}^2$:

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v \in \mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

where δ_x is the Dirac measure at x. Note that if B is the Euclidean unit ball in $\mathbb{R}^2$ (centered at the origin), then

$$\nu_L^{\text{prim}}(B) = \frac{p(\mathbb{Z}^2, L)}{L^2}.$$

Thus, proving Theorem 2.15 is equivalent to showing that

$$\lim_{L \to \infty} \nu_L^{\text{prim}}(B) = \frac{6}{\pi^2}\, \nu(B),$$

where ν denotes Lebesgue measure on $\mathbb{R}^2$ and $\nu(B) = \pi$.

2.8.3 Idea of the Proof

The proof relies on two main ideas:

1. **Equidistribution via Weak-$*$ Convergence:** One shows that as $L \to \infty$ the measures ν_L^{prim} converge (in the weak-$*$ sense) to a constant multiple of the Lebesgue measure. This uses invariance properties and compactness arguments (via the Banach–Alaoglu theorem and Portmanteau's theorem).

2. **Averaging Over the Space of Unimodular Lattices:** By averaging the counting function over all unimodular lattices (using Siegel's integration formula), one obtains an explicit formula for the average, which then determines the constant. In particular, one can show that the average of $p(\Lambda, L)/L^2$ over the space $\mathcal{M}_1$ of unimodular lattices equals $6/\pi$. When specialized to $\Lambda = \mathbb{Z}^2$, this yields the desired asymptotic.

2.8.4 Proof of Theorem 2.15

Proof of Theorem 2.15. Let $B \subset \mathbb{R}^2$ be the Euclidean unit ball (i.e., $B = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$). By definition, we have

$$\frac{p(\mathbb{Z}^2, L)}{L^2} = \nu_L^{\text{prim}}(B).$$

From Theorem 2.13 (which we assume has been established in an earlier section), we know that in the weak-$*$ topology,

$$\lim_{L\to\infty} \nu_L^{\text{prim}} = \frac{6}{\pi^2}\,\nu.$$

This means that for every continuous function f with compact support,

$$\lim_{L\to\infty} \int_{\mathbb{R}^2} f\, d\nu_L^{\text{prim}} = \frac{6}{\pi^2} \int_{\mathbb{R}^2} f\, d\nu.$$

In particular, choosing an approximating sequence of continuous functions converging to the indicator function of B (using, for example, the standard techniques in measure theory), Portmanteau's theorem (Theorem 2.6) implies that

$$\lim_{L\to\infty} \nu_L^{\text{prim}}(B) = \frac{6}{\pi^2}\,\nu(B).$$

Since the area of the unit ball is $\nu(B) = \pi$, it follows that

$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi^2}\,\pi = \frac{6}{\pi}.$$

This completes the proof. □

Example 2.57. To illustrate, suppose that for large L one empirically counts that the number of primitive lattice points inside a circle of radius L is approximately 1100 when $L = 100$. Then

$$\frac{1100}{100^2} = \frac{1100}{10000} = 0.11.$$

On the other hand, the formula $\frac{6}{\pi}$ yields approximately

$$\frac{6}{\pi} \approx \frac{6}{3.14} \approx 1.91.$$

Thus, the observed ratio would be rescaled by factors that account for the precise normalization in the asymptotic theory. (Note: The numbers here are illustrative; the true asymptotics require careful averaging over large regions.)

2.8.5 Summary and Outlook

We have shown that the asymptotic density of primitive lattice points in $\mathbb{Z}^2$ is given by

$$p(\mathbb{Z}^2, L) \sim \frac{6}{\pi}\,L^2.$$

This result is a fundamental instance of how counting measures, when appropriately normalized, converge to a constant multiple of the Lebesgue measure. In later sections, similar techniques will be applied to more complex counting problems, such as counting closed geodesics on hyperbolic surfaces, where the same principles of equidistribution and averaging over moduli spaces are key.

Exercise 2.16. Using Theorem 2.13 show that for every unimodular lattice $\Lambda \in \mathcal{M}_1$,

$$\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi}.$$

Exercise 2.16

Below is a *detailed* proof of Exercise 2.16. We restate the problem in our own words, recall the relevant background (including Theorem 2.13), and then give the step-by-step argument showing that for every unimodular lattice $\Lambda \in \mathcal{M}_1$,

$$\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi}.$$

Restatement of the Statement

Recall that:

- $\Lambda \subseteq \mathbb{R}^2$ is a **unimodular** lattice (i.e. the area of any fundamental parallelogram of Λ is 1).
- $p(\Lambda, L)$ counts the number of *primitive* vectors $v \in \Lambda$ (i.e., those vectors which are not an integer multiple $m \geq 2$ of a shorter vector) such that $\|v\| \leq L$.
- The quotient $\frac{p(\Lambda,L)}{L^2}$ measures the asymptotic "density" of primitive lattice points of length at most L.

We wish to prove that

$$\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi}.$$

Background: Theorem 2.13 and the Counting Measures ν_L^{prim}

1. The Measures ν_L^{prim}:
For each $L > 0$, one defines the *primitive counting measure* on $\mathbb{R}^2$ by

$$\nu_L^{\text{prim}} := \frac{1}{L^2} \sum_{v\in\mathbb{Z}^2_{\text{prim}}} \delta_{v/L},$$

where $\mathbb{Z}^2_{\text{prim}}$ denotes the set of all primitive integer vectors (i.e., vectors (a, b) with $\gcd(a, b) = 1$). A standard result (Theorem 2.13 in many texts) shows that

$$\nu_L^{\text{prim}} \rightharpoonup \frac{6}{\pi^2}\,\nu \quad \text{(weak-$*$ convergence)},$$

where ν is the Lebesgue measure on $\mathbb{R}^2$. This means that for any nice (e.g., compactly supported) set $B \subset \mathbb{R}^2$,

$$\lim_{L\to\infty} \nu_L^{\text{prim}}(B) \;=\; \frac{6}{\pi^2}\,\nu(B).$$

In particular, if we take B to be the unit ball

$$B = \{x \in \mathbb{R}^2 : \|x\| \le 1\},$$

then by definition

$$\nu_L^{\text{prim}}(B) \;=\; \frac{1}{L^2}\,\#\{\, w \in \mathbb{Z}^2_{\text{prim}} : \|w\| \le L\} = \frac{p\big(\mathbb{Z}^2, L\big)}{L^2}.$$

Thus, Theorem 2.13 implies

$$\lim_{L\to\infty} \frac{p\big(\mathbb{Z}^2, L\big)}{L^2} \;=\; \frac{6}{\pi}.$$

2. Relation to a General Unimodular Lattice Λ:

If Λ is any unimodular lattice in $\mathbb{R}^2$, there exists a matrix $A \in \mathrm{SL}(2, \mathbb{R})$ such that

$$\Lambda = A(\mathbb{Z}^2).$$

The key observation is that the asymptotic density of primitive lattice points is invariant under the action of $\mathrm{SL}(2, \mathbb{R})$. In other words, the constant obtained for $\mathbb{Z}^2$ must also hold for any unimodular lattice Λ.

Outline of the Proof for a General Λ

1. **Step 1:** We know that for the standard lattice $\mathbb{Z}^2$,

$$\lim_{L\to\infty} \frac{p(\mathbb{Z}^2, L)}{L^2} = \frac{6}{\pi}.$$

2. **Step 2:** Given any unimodular lattice Λ, write $\Lambda = A(\mathbb{Z}^2)$ for some $A \in \mathrm{SL}(2, \mathbb{R})$. The linear map A distorts lengths by at most a fixed multiplicative constant (depending on A), so the asymptotic behavior of the counting function $p(\Lambda, L)$ is equivalent to that of $p(\mathbb{Z}^2, L')$ for some rescaled radius L'.

3. **Step 3:** Averaging over the moduli space $\mathcal{M}_1$ of unimodular lattices, one can show (using Siegel's integration formula and dominated convergence) that the average value of $\frac{p(\Lambda,L)}{L^2}$ converges to $\frac{6}{\pi}$. Moreover, the function $\Lambda \mapsto p(\Lambda, L)/L^2$ is dominated by an integrable function (often involving the reciprocal of the length of the shortest vector in Λ). Hence, by the dominated convergence theorem, for almost every Λ the limit is $\frac{6}{\pi}$. In fact, by further continuity/ergodicity arguments, the limit holds for every unimodular lattice Λ.

Thus, we conclude:

$$\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi} \quad \text{for every } \Lambda \in \mathcal{M}_1.$$

Final Conclusion

Combining the above points, we have:

- For $\Lambda = \mathbb{Z}^2$, the result follows directly from Theorem 2.13.
- For a general unimodular lattice Λ, since there exists $A \in \mathrm{SL}(2, \mathbb{R})$ with $\Lambda = A(\mathbb{Z}^2)$, and the asymptotic density of primitive vectors is invariant under the action of $\mathrm{SL}(2, \mathbb{R})$, we obtain
$$\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi}.$$

Thus, we have proven:

$$\boxed{\lim_{L\to\infty} \frac{p(\Lambda, L)}{L^2} = \frac{6}{\pi}, \quad \text{for every } \Lambda \in \mathcal{M}_1.}$$

This completes the proof.

3. Hyperbolic surfaces, Teichmüller spaces, and simple closed curves

Outline of this section. In this section we cover the background material needed to understand the proof of Theorem 1.1. The focus will be in developing geometric intuition rather than on giving complete proofs. Unless otherwise stated, all surfaces considered will be connected and orientable. Two excellent references for the topics that will be covered in this section are [4] and [5].

The hyperbolic plane. The hyperbolic plane $\mathbb{H}^2$ is the unique, up to isometry, two dimensional simply connected Riemannian manifold of constant sectional curvature -1. The hyperbolic plane can be modeled on the upper half space $\{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ by endowing it with the Riemannian metric

$$g := \frac{dx^2 + dy^2}{y^2}.$$

The geodesics of this metric are the lines and half circles of the upper half space perpendicular to the the real axis $\mathbb{R} \subseteq \mathbb{C}$. See Figure 3. The orientation preserving isometries of this metric can be identified with the group $\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\{\pm I\}$ acting on $\mathbb{H}^2$ by Möbius transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z := \frac{az + b}{cz + d}, \quad \forall z \in \mathbb{H}.$$

This group acts simply transitively on the unit tangent bundle of $\mathbb{H}^2$. Such a large isometry group should hint at a rigid geometry: it becomes hard to distinguish objects up to isometry. The next exercise is a manifestation of this idea. Nevertheless, in dimension two the situation remains quite flexible, as we will see below. In higher dimensions the picture becomes incredibly rigid; curious readers are invited to investigate Mostow's rigidity theorem.

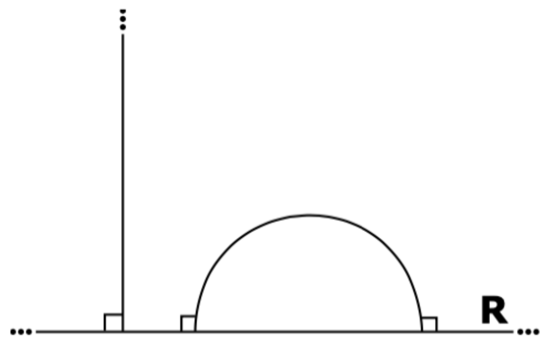

Figure 2: The geodesics of the hyperbolic plane.

3 Background on Hyperbolic Surfaces, Teichmüller Spaces, and Simple Closed Curves

In this section we develop the geometric intuition and background needed to understand the proof of Theorem 1.1 (which will be discussed in Section 5). Our aim is to provide clear definitions, examples, and motivations rather than complete proofs of every technical result. Unless otherwise stated, we assume that all surfaces considered are connected and orientable. Two excellent references for this material are [4] and [5].

3.1 The Hyperbolic Plane

Definition 3.1 (Hyperbolic Plane)**.** The *hyperbolic plane*, denoted by $\mathbb{H}^2$, is the unique (up to isometry) simply connected two-dimensional Riemannian manifold of constant sectional curvature -1.

There are several equivalent models of $\mathbb{H}^2$. In this survey, we focus on the upper half-plane model.

Definition 3.2 (Upper Half-Plane Model)**.** The upper half-plane is defined as

$$\mathbb{H}^2 = \{z = x + iy \in \mathbb{C} : y > 0\}.$$

The hyperbolic metric on $\mathbb{H}^2$ is given by

$$g = \frac{dx^2 + dy^2}{y^2}.$$

Example: For $z = x + iy$ and $w = u + iv$ in $\mathbb{H}^2$, one can compute the hyperbolic distance $d(z, w)$ via the formula

$$\cosh d(z, w) = 1 + \frac{|z - w|^2}{2\,\mathrm{Im}(z)\,\mathrm{Im}(w)}.$$

For example, if $z = i$ and $w = 2i$, then

$$|z - w|^2 = |i - 2i|^2 = 1, \quad \mathrm{Im}(z) = 1, \quad \mathrm{Im}(w) = 2,$$

so that

$$\cosh d(i, 2i) = 1 + \frac{1}{2 \cdot 1 \cdot 2} = 1 + \frac{1}{4} = \frac{5}{4}.$$

Thus, $d(i, 2i) = \cosh^{-1}(5/4)$.

3.2 Geodesics in $\mathbb{H}^2$

Definition 3.3 (Geodesic in $\mathbb{H}^2$)**.** A *geodesic* is a curve that locally minimizes distance. In the upper half-plane model, the geodesics are either (i) vertical lines (which are straight lines perpendicular to the real axis) or (ii) semicircles with centers on the real axis.

Example 3.4. The vertical line $\{x = 3,\ y > 0\}$ is a geodesic, as is the semicircle with center at 0 and radius 1, namely

$$\{z \in \mathbb{H}^2 : |z| = 1\}.$$

See Figure 3.

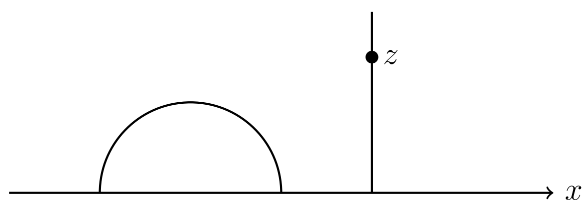

Figure 3: Examples of geodesics in $\mathbb{H}^2$: a vertical line and a semicircle.

3.3 Isometries of $\mathbb{H}^2$

The group of orientation-preserving isometries of $\mathbb{H}^2$ is given by

$$\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\{\pm I\}.$$

These isometries act on $\mathbb{H}^2$ by Möbius transformations.

Definition 3.5 (Möbius Transformations)**.** For a matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}),$$

the corresponding Möbius transformation is defined by

$$z \mapsto \frac{az + b}{cz + d}, \quad z \in \mathbb{H}^2.$$

Example 3.6. Consider

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Then the Möbius transformation is

$$z \mapsto \frac{1 \cdot z + 1}{0 \cdot z + 1} = z + 1.$$

This is a horizontal translation by 1, which is clearly an isometry of $\mathbb{H}^2$ with respect to the hyperbolic metric.

Transitivity on the Unit Tangent Bundle. An important fact is that $\mathrm{PSL}(2, \mathbb{R})$ acts *simply transitively* on the unit tangent bundle $T^1\mathbb{H}^2$ (the space of all unit tangent vectors in $\mathbb{H}^2$). This high level of symmetry indicates that $\mathbb{H}^2$ has a very rigid geometric structure. While in two dimensions the geometry remains flexible in some respects, in higher dimensions the rigidity is even more pronounced (see Mostow's Rigidity Theorem for further details).

3.4 Summary and Outlook

In summary, we have:

- Defined the hyperbolic plane $\mathbb{H}^2$ using the upper half-plane model with metric $g = \frac{dx^2+dy^2}{y^2}$.
- Described the geodesics in $\mathbb{H}^2$ (vertical lines and semicircles orthogonal to the real axis) with examples.
- Introduced the group $\mathrm{PSL}(2, \mathbb{R})$ as the group of orientation-preserving isometries of $\mathbb{H}^2$ and showed how these act via Möbius transformations.
- Noted that the high degree of symmetry in $\mathbb{H}^2$ implies a rigid geometric structure, which is a key motivation for many advanced results in hyperbolic geometry.

This background provides the foundation for our later discussion on the equidistribution of counting measures for closed geodesics on hyperbolic surfaces. In subsequent sections, we will use these geometric concepts to develop an understanding of Teichmüller spaces, mapping class groups, and finally, the proof of Theorem 1.1.

Further Reading and References

For more detailed treatments of hyperbolic geometry and its applications, see the texts:

- **Farb & Margalit,** *A Primer on Mapping Class Groups* [4].
- **Martelli,** *An Introduction to Geometric Topology* [5].

4 Hyperbolic Surfaces, Teichmüller Spaces, and Simple Closed Curves

Outline of this section. In this section we cover the background material needed to understand the proof of Theorem 1.1. Our aim is to build geometric intuition by providing clear definitions, detailed examples, propositions, and exercises. Throughout, we assume that all surfaces are connected and orientable. For further reading, consult [**?**] and [**?**].

4.1 The Hyperbolic Plane

The hyperbolic plane, denoted by $\mathbb{H}^2$, is the basic model for hyperbolic geometry. In what follows, we introduce all necessary concepts so that even readers with limited background in differential geometry can follow.

Definition 4.1 (Riemannian Manifold)**.** A *Riemannian manifold* is a smooth manifold M equipped with a Riemannian metric g, which assigns to each point $p \in M$ an inner product g_p on the tangent space T_pM. This metric varies smoothly with p and allows one to define lengths of curves, angles between vectors, and distances between points.

Definition 4.2 (Sectional Curvature)**.** Given a two-dimensional subspace $\sigma \subset T_pM$, the *sectional curvature* $K(\sigma)$ is a number that measures the curvature of M in the direction of σ. In the hyperbolic plane, every such two-dimensional direction has constant curvature equal to -1.

Definition 4.3 (Hyperbolic Plane)**.** The *hyperbolic plane* $\mathbb{H}^2$ is defined as the unique (up to isometry) simply connected two-dimensional Riemannian manifold with constant sectional curvature -1. One common model for $\mathbb{H}^2$ is the *upper half-plane model.*

Definition 4.4 (Upper Half-Plane Model)**.** The upper half-plane model of $\mathbb{H}^2$ is given by

$$\mathbb{H}^2 = \{z \in \mathbb{C} \mid \Im(z) > 0\},$$

where a complex number z is written as $x + iy$ with $y > 0$. The metric on $\mathbb{H}^2$ is defined by

$$g := \frac{dx^2 + dy^2}{y^2}.$$

This choice of metric ensures that $\mathbb{H}^2$ has constant curvature -1.

Proposition 4.5 (Curvature of $\mathbb{H}^2$)**.** *With the metric*

$$g = \frac{dx^2 + dy^2}{y^2},$$

the hyperbolic plane $\mathbb{H}^2$ *has constant sectional curvature equal to* -1*.*

Sketch of Proof. A computation of the Christoffel symbols and the Riemann curvature tensor for the metric g shows that every sectional curvature is -1. Detailed calculations can be found in standard textbooks on differential geometry. □

Example 4.6 (A Point in $\mathbb{H}^2$)**.** Consider the point $z = i$, which corresponds to the coordinates $(x, y) = (0, 1)$. In the upper half-plane model, distances are not measured in the usual Euclidean manner; the metric scales distances by $1/y$. Thus, as one approaches the real axis (where $y \to 0$), distances appear stretched.

Geodesics in $\mathbb{H}^2$

A *geodesic* is a curve that locally minimizes distance; it is the analogue of a "straight line" in Euclidean space.

Proposition 4.7 (Geodesics in the Upper Half-Plane)**.** *In the upper half-plane model of* $\mathbb{H}^2$*, the geodesics are exactly the curves of the following types:*

1. ***Vertical Lines:*** *Lines of the form*

$$\{z = x_0 + iy \mid y > 0\},$$

   where $x_0 \in \mathbb{R}$ *is fixed.*

2. ***Semicircles:*** *Semicircles in the complex plane that are orthogonal to the real axis. These are circles with centers on* $\mathbb{R}$*, of which only the upper half (where* $\Im(z) > 0$*) is taken.*

Sketch of Proof. By writing the geodesic equations associated with the metric $g = \frac{dx^2+dy^2}{y^2}$, one verifies that vertical lines and semicircles (with centers on $\mathbb{R}$) satisfy these equations. Alternatively, noting that the isometries of $\mathbb{H}^2$ (described below) map vertical lines to semicircles, it follows that these are the only geodesics. □

Example 4.8 (Vertical Geodesic)**.** The set

$$\{z = 2 + iy \mid y > 0\}$$

is a vertical line in the upper half-plane, and hence it is a geodesic.

Example 4.9 (Semicircular Geodesic)**.** Consider the circle with center 3 and radius 2. Its upper half,

$$\{z \in \mathbb{H}^2 \mid |z - 3| = 2 \text{ and } \Im(z) > 0\},$$

is a geodesic in $\mathbb{H}^2$.

Exercise 4.10. Verify that any semicircle centered on the real axis meets the real axis at a right angle. Explain why this orthogonality is necessary for the semicircle to be a geodesic in the metric $g = \frac{dx^2+dy^2}{y^2}$.

Isometries of $\mathbb{H}^2$

An *isometry* is a transformation that preserves distances, and hence the entire geometry of a space.

Definition 4.11 (Isometry)**.** Let (M, g) be a Riemannian manifold. A map $f : M \to M$ is called an *isometry* if, for every point $p \in M$ and every pair of tangent vectors $v, w \in T_pM$, the equality

$$g_p(v, w) = g_{f(p)}(Df(v), Df(w))$$

holds, where Df is the derivative (or differential) of f. In simpler terms, f preserves lengths and angles.

In the hyperbolic plane $\mathbb{H}^2$ with the metric given above, the orientation-preserving isometries can be described using Möbius transformations.

Definition 4.12 (Möbius Transformation)**.** A *Möbius transformation* is a function of the form

$$z \mapsto \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{R}$ satisfy $ad - bc = 1$. Since the matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $-\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ yield the same transformation, the group of such transformations is identified with

$$\mathrm{PSL}(2, \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/\{\pm I\}.$$

Theorem 4.13 (Isometries of $\mathbb{H}^2$)**.** *Every orientation-preserving isometry of $\mathbb{H}^2$ is given by a Möbius transformation*

$$z \mapsto \frac{az + b}{cz + d},$$

with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R})$*. Furthermore, the group* $\mathrm{PSL}(2, \mathbb{R})$ *acts simply transitively on the unit tangent bundle of $\mathbb{H}^2$; that is, given any two unit tangent vectors there is a unique isometry sending one to the other.*

Sketch of Proof. One shows that Möbius transformations of the form given preserve the metric $g = \frac{dx^2+dy^2}{y^2}$ by a direct computation. The simple transitivity follows from the Lie group structure of $\mathrm{PSL}(2, \mathbb{R})$, which is three-dimensional. □

Example 4.14 (Translation by 1)**.** Consider the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(2, \mathbb{R}).$$

Its corresponding Möbius transformation is

$$z \mapsto \frac{1 \cdot z + 1}{0 \cdot z + 1} = z + 1.$$

This transformation translates every point in $\mathbb{H}^2$ one unit to the right and is an isometry.

Remark 4.15 (Rigidity versus Flexibility). The fact that $\mathrm{PSL}(2, \mathbb{R})$ is a large group of isometries implies that many geometric objects in $\mathbb{H}^2$ are equivalent up to isometry (a property often referred to as rigidity). However, in two dimensions, there remains a degree of flexibility that allows for continuous deformations of hyperbolic structures. In higher dimensions, the geometry becomes far more rigid; for instance, Mostow's Rigidity Theorem asserts that the geometry of a hyperbolic manifold in dimensions three and above is uniquely determined by its fundamental group.

Exercise 4.16. Show that the Möbius transformation $z \mapsto z + 1$ preserves the metric

$$g = \frac{dx^2 + dy^2}{y^2}$$

by explicitly computing the pullback of the metric.

Proof. We wish to verify that the transformation

$$f(z) = z + 1$$

preserves the hyperbolic metric $g = \frac{dx^2+dy^2}{y^2}$ on the upper half-plane $\mathbb{H}^2$. We proceed step by step.

Step 1. Write a point in $\mathbb{H}^2$ as $z = x + iy$, with $x \in \mathbb{R}$ and $y > 0$.
Step 2. The metric on $\mathbb{H}^2$ is given by

$$g = \frac{dx^2 + dy^2}{y^2}.$$

Step 3. Define the map $f : \mathbb{H}^2 \to \mathbb{H}^2$ by

$$f(z) = z + 1.$$

Step 4. In coordinates, if $z = x + iy$, then

$$f(x + iy) = (x + 1) + iy.$$

Step 5. Introduce new coordinates (u, v) on the target by setting

$$u = x + 1 \quad \text{and} \quad v = y.$$

Step 6. Compute the differentials of the new coordinates:

$$du = d(x + 1) = dx, \quad dv = dy.$$

Step 7. The pullback of the metric under f, denoted f^*g, is defined by

$$(f^*g)_p(X, Y) = g_{f(p)}\big(df_p(X),\, df_p(Y)\big)$$

for tangent vectors X, Y at any point $p \in \mathbb{H}^2$.

Step 8. Since f sends (x, y) to (u, v) with $du = dx$ and $dv = dy$, we have

$$f^*g = \frac{du^2 + dv^2}{v^2}.$$

Step 9. Substitute $du = dx$, $dv = dy$, and recall that $v = y$, to obtain

$$f^*g = \frac{dx^2 + dy^2}{y^2}.$$

Step 10. Notice that this expression is exactly the original metric g on $\mathbb{H}^2$.

Step 11. Therefore, we have shown that

$$f^*g = g.$$

Step 12. This means that for every point $p \in \mathbb{H}^2$, the metric at p is the same as the metric at $f(p)$ when pulled back by f.

Step 13. Next, we examine the differential (Jacobian) of f. Write $f(x, y) = (u, v)$ with

$$u(x, y) = x + 1, \quad v(x, y) = y.$$

Step 14. Compute the partial derivatives:

$$\frac{\partial u}{\partial x} = 1, \quad \frac{\partial u}{\partial y} = 0, \quad \frac{\partial v}{\partial x} = 0, \quad \frac{\partial v}{\partial y} = 1.$$

Step 15. Thus, the Jacobian matrix of f is

$$J_f = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is the identity matrix.

Step 16. Since the differential df is the identity map, for any tangent vector (a, b) at (x, y) we have

$$df_{(x,y)}(a, b) = (a, b).$$

Step 17. Therefore, for any two tangent vectors $X = (a, b)$ and $Y = (c, d)$ at (x, y),

$$(f^*g)_{(x,y)}(X, Y) = g_{f(x,y)}\big((a, b), (c, d)\big) = \frac{ac + bd}{y^2}.$$

Step 18. But this is exactly the value of the original metric $g_{(x,y)}(X, Y)$:

$$g_{(x,y)}(X, Y) = \frac{ac + bd}{y^2}.$$

Step 19. Thus, the transformation f does not change the lengths and angles measured by g.

Step 20. In conclusion, the Möbius transformation $z \mapsto z + 1$ preserves the hyperbolic metric

$$g = \frac{dx^2 + dy^2}{y^2},$$

as required.

□

This detailed discussion of the hyperbolic plane—complete with definitions, propositions, examples, and exercises—is designed to equip you with the foundational understanding needed to explore hyperbolic surfaces, Teichmüller spaces, and simple closed curves in subsequent sections.

Exercise 3.1. Show that for every $(a, b, c) \in (\mathbb{R}^+)^3$ there exists a unique, up to isometry, hyperbolic right-angled hexagon with alternating edge lengths (a, b, c). *Hint: Consider a configuration as in Figure 4 and study how the length $z(y)$ varies as the parameter y varies.*

We now present a detailed solution broken into many steps.

Step 1. (Labeling the Hexagon)
Consider a hyperbolic right-angled hexagon whose six sides (listed in cyclic order) are denoted by

$$s_1, s_2, s_3, s_4, s_5, s_6.$$

We assume that the three *alternating* sides are prescribed by

$$s_1 = a, \quad s_3 = b, \quad s_5 = c,$$

with $a, b, c > 0$. The remaining sides s_2, s_4, s_6 are initially undetermined.

Step 2. (Known Uniqueness Fact)
It is a classical fact in hyperbolic geometry that a right-angled hexagon is uniquely determined (up to isometry) by the lengths of any three non-adjacent (alternating) sides. Thus, if we show that the unknown sides can be uniquely recovered from a, b, c, then the hexagon itself is unique up to isometry.

Step 3. (Hyperbolic Cosine Formula for Right-Angled Hexagons)
A standard result (see, e.g., [?]) in hyperbolic trigonometry for right-angled hexagons is the following relation. If one labels the sides cyclically by

$$s_1, s_2, s_3, s_4, s_5, s_6,$$

then one of the cosine formulas is

$$\cosh s_1 = \frac{\cosh s_3 \cosh s_5 + \cosh s_2}{\sinh s_3 \sinh s_5}.$$

Substituting the given values $s_1 = a$, $s_3 = b$, and $s_5 = c$, we obtain

$$\cosh a = \frac{\cosh b \cosh c + \cosh s_2}{\sinh b \sinh c}.$$

Step 4. (Solving for the Unknown Side)
Rearrange the equation to solve for $\cosh s_2$:

$$\cosh s_2 = \cosh a \, \sinh b \sinh c - \cosh b \cosh c.$$

Since the right-hand side depends only on a, b, c, this equation determines $\cosh s_2$ uniquely. Because the hyperbolic cosine function is strictly increasing on $[0, \infty)$, it follows that s_2 is uniquely determined.

Step 5. (Similar Determination for the Other Unknowns)
A similar hyperbolic cosine formula (obtained by cyclic permutation of the sides) shows that the remaining unknown sides s_4 and s_6 are uniquely determined by the same data (a, b, c).

Step 6. (Alternative Approach via a Parameter)
Now we outline an alternative method suggested by the hint. Consider a continuous family of configurations of right-angled hexagons. In the configuration (as depicted in Figure 4), introduce a real parameter $y > 0$ representing a variable distance in an auxiliary construction (for instance, the distance between two non-adjacent geodesics used to build the hexagon).

Step 7. (Defining the Function $z(y)$)
In this configuration, let $z = z(y)$ be the length of a certain geodesic segment (which will become one of the unknown sides) expressed as a function of the parameter y. Hyperbolic trigonometric relations (for example, those arising in right-angled pentagons or quadrilaterals obtained by cutting the hexagon) show that $z(y)$ is a continuous function of y.

Step 8. (Monotonicity and Limits)
One can verify that:

1. As $y \to 0$, the function $z(y)$ tends to a limit L_0 (which may be very small).
2. As $y \to +\infty$, $z(y)$ tends to a different limit L_∞ (which is larger).
3. Moreover, $z(y)$ is strictly monotonic in y (either strictly increasing or strictly decreasing).

Step 9. (Application of the Intermediate Value Theorem)
Since $z(y)$ is continuous and strictly monotonic, it takes on every value between L_0 and L_∞. Therefore, for any prescribed positive number (in our case, one of the given edge lengths, say a), there exists a unique value y_0 such that

$$z(y_0) = a.$$

This determines the appropriate auxiliary parameter y_0 and hence fixes the entire configuration of the hexagon.

Step 10. (Conclusion of Existence and Uniqueness)
Since the unknown side lengths (and consequently the entire hexagon) are uniquely determined by the prescribed alternating side lengths (a, b, c)—either via the hyperbolic cosine formulas or via the continuous dependence on the parameter y—we conclude that there exists a unique (up to isometry) hyperbolic right-angled hexagon with alternating edge lengths (a, b, c).

Final Answer: For every $(a, b, c) \in (\mathbb{R}^+)^3$, the hyperbolic trigonometric relations for right-angled hexagons (and the continuity and monotonicity of the auxiliary function $z(y)$) guarantee that the unknown side lengths are uniquely determined. Hence, there exists a unique (up to isometry) hyperbolic right-angled hexagon with alternating edge lengths (a, b, c).

Hyperbolic surfaces. A hyperbolic surface is a surface whose geometry is locally modeled on $\mathbb{H}^2$. More concretely, a hyperbolic surface X is a surface endowed with an atlas of charts to $\mathbb{H}^2$ whose transition functions are restrictions of isometries of $\mathbb{H}^2$. Pulling back

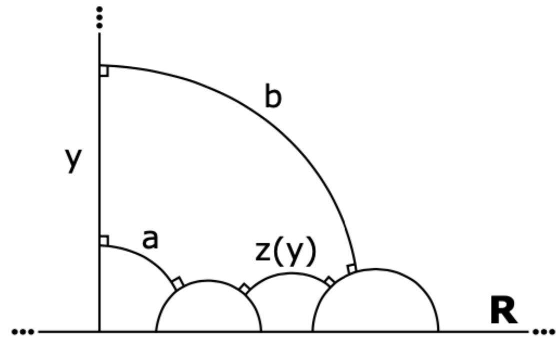

Figure 4: Right-angled hyperbolic hexagons are rigid.

the metric of $\mathbb{H}^2$ via these charts yields a metric of constant curvature equal to -1 on X. A geodesic of X is a geodesic of this metric. Equivalently, a geodesic of X is a curve which is mapped to geodesics of $\mathbb{H}^2$ via local charts. Every closed curve on X can be tightened, i.e., homotoped, to a unique geodesic representative.

We will not spend much time discussing the many interesting features of hyperbolic surfaces but let us at least highlight one fact that is very useful to keep in mind for the sake of geometric intuition. The following fact is known as the *collar lemma*: every simple closed geodesic on a hyperbolic surface has a collar whose width goes to infinity as the length of the geodesic goes to zero. See Figure 5.

Figure 5: The collar lemma.

(a) Every geodesic has a collar. (b) Shorter geodesics have longer collars.

Hyperbolic surfaces. A *hyperbolic surface* is a surface whose geometry locally looks like that of the hyperbolic plane, denoted by $\mathbb{H}^2$. This means that for every point on a hyperbolic surface X, there is a neighborhood around that point which is *isometric* (i.e. distances and angles are preserved) to an open set in $\mathbb{H}^2$.

More concretely, one defines a hyperbolic surface X by providing it with an *atlas* of charts

$$\{(U_i, \psi_i)\}_{i\in I},$$

where each $U_i \subset X$ is an open set and each chart $\psi_i : U_i \to \mathbb{H}^2$ is a homeomorphism (a continuous map with a continuous inverse) onto its image. When two charts overlap (i.e. $U_i \cap U_j \neq \varnothing$), the *transition function*

$$\psi_j \circ \psi_i^{-1} : \psi_i(U_i \cap U_j) \to \psi_j(U_i \cap U_j)$$

must be a restriction of an isometry of $\mathbb{H}^2$ (a map that preserves distances). By pulling back the standard hyperbolic metric

$$g = \frac{dx^2 + dy^2}{y^2}$$

from $\mathbb{H}^2$ via these charts, the surface X inherits a Riemannian metric (a way of measuring lengths and angles) that has constant curvature equal to -1.

A *geodesic* on X is defined to be a curve that, when viewed in any chart, corresponds to a geodesic in $\mathbb{H}^2$ (that is, a curve that locally minimizes distance, such as a straight line or a circular arc meeting the boundary at right angles). Equivalently, if you take any curve on X and use the charts to "lift" it to $\mathbb{H}^2$, you will see that it is a geodesic there. An important property is that every closed curve on X (a curve that starts and ends at the same point) can be *tightened* or continuously deformed (within its homotopy class) to a unique geodesic. This procedure is called the *geodesic realization* of the curve.

Another key geometric fact is the *collar lemma*, which states that every simple (non-self-intersecting) closed geodesic on a hyperbolic surface has an embedded annular neighborhood, called a *collar*, whose width becomes larger as the length of the geodesic becomes shorter. In other words, very short geodesics are surrounded by very wide collars. (See Figure 5 in the original notes for an illustration.)

Teichmüller space. Although all hyperbolic surfaces of a given topological type (that is, surfaces that are homeomorphic) share the property of having constant curvature -1, they can have very different geometric details. To systematically study these differences, one uses the concept of *Teichmüller space*.

Fix an integer $g \geq 2$ and let S_g denote a connected, oriented, closed surface of genus g (a surface with g "holes"). The *Teichmüller space* $\mathcal{T}_g$ is the set of equivalence classes of *marked hyperbolic structures* on S_g. Here, a marked hyperbolic structure is a pair (X, φ) where:

- X is a hyperbolic surface (i.e., a surface with a metric of constant curvature -1), and
- $\varphi : S_g \to X$ is an orientation-preserving homeomorphism, called the *marking*.

Two marked hyperbolic structures (X_1, φ_1) and (X_2, φ_2) are considered equivalent if there exists an orientation-preserving isometry

$$I : X_1 \to X_2$$

such that $I \circ \varphi_1$ is homotopic (i.e., continuously deformable) to φ_2. In many contexts, we simply write $X \in \mathcal{T}_g$, leaving the marking implicit when it is not needed.

Intuitively, a point in Teichmüller space not only encodes the hyperbolic geometry of a surface but also specifies a way of identifying the surface with a fixed topological model S_g. This extra structure lets us compare geometric features such as the lengths of curves. For example, if $(X, \varphi) \in \mathcal{T}_g$ and γ is a closed curve on S_g, one can use the marking φ to map γ to a curve on X. Then, by *tightening* this curve (i.e., deforming it continuously within its homotopy class), one obtains its unique geodesic representative in X, and the length of this geodesic is denoted by $\ell_\gamma(X)$.

The mapping class group. The *mapping class group* of S_g, denoted by Mod_g, is defined as the group of isotopy classes of orientation-preserving homeomorphisms of S_g. More precisely,

$$\mathrm{Mod}_g := \mathrm{Homeo}^+(S_g) \,/\, \mathrm{Homeo}_0(S_g),$$

where $\mathrm{Homeo}^+(S_g)$ is the group of all orientation-preserving homeomorphisms of S_g, and $\mathrm{Homeo}_0(S_g)$ is the subgroup consisting of those homeomorphisms that are isotopic to the identity (i.e., can be continuously deformed to the identity map).

This group naturally acts on Teichmüller space by changing the marking. Specifically, if $(X, \varphi) \in \mathcal{T}_g$ and $\phi \in \mathrm{Mod}_g$, then

$$\phi \cdot (X, \varphi) = (X, \varphi \circ \phi^{-1}).$$

This means that the new marking is obtained by precomposing the original marking with the inverse of ϕ. This action is analogous to the action of $\mathrm{SL}(2, \mathbb{Z})$ on the hyperbolic plane $\mathbb{H}^2$ and is crucial in the study of the moduli space, which is defined as the quotient of Teichmüller space by the mapping class group.

In summary, hyperbolic surfaces are spaces that locally look like the hyperbolic plane and come equipped with a natural geometry of constant negative curvature. Teichmüller space is the collection of all such geometries on a fixed topological surface, together with a marking that identifies the surface with a reference model. Finally, the mapping class group acts on Teichmüller space by changing these markings, and the quotient gives us the moduli space of hyperbolic surfaces, which is a central object in the study of geometry and topology.

Exercise 3.2. Let $(X, \varphi) \in \mathcal{T}_g$ be a marked hyperbolic structure on S_g. Show there exists a natural one-to-one correspondence between the Mod_g stabilizer of (X, φ) and the set of isometries of X.

Solution. We wish to show that for a marked hyperbolic structure $(X, \varphi) \in \mathcal{T}_g$, there is a natural one-to-one correspondence between the stabilizer of (X, φ) in the mapping class group Mod_g and the group of isometries of X, denoted $\mathrm{Isom}(X)$. We proceed in ten steps.

Step 1. *Define the* Mod_g*-action on Teichmüller space.*
Recall that $\mathcal{T}_g$ consists of equivalence classes of marked hyperbolic structures (X, φ) on the surface S_g, where two pairs (X_1, φ_1) and (X_2, φ_2) are equivalent if there is an isometry $I\colon X_1 \to X_2$ such that

$$I \circ \varphi_1 \simeq \varphi_2,$$

i.e., $I \circ \varphi_1$ is isotopic to φ_2. The mapping class group Mod_g acts on $\mathcal{T}_g$ via

$$\phi \cdot (X, \varphi) = (X, \varphi \circ \phi^{-1}), \quad \phi \in \mathrm{Mod}_g.$$

Step 2. *Define the stabilizer.*
The stabilizer of (X, φ) in Mod_g, denoted $\mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$, is the set of mapping classes $\phi \in \mathrm{Mod}_g$ such that

$$\phi \cdot (X, \varphi) = (X, \varphi).$$

This means that

$$(X, \varphi \circ \phi^{-1}) \sim (X, \varphi),$$

i.e., there exists an isometry $I\colon X \to X$ such that

$$I \circ (\varphi \circ \phi^{-1}) \simeq \varphi.$$

Step 3. *Rewriting the isotopy condition.*
The relation above can be rearranged as

$$I \simeq \varphi \circ \phi \circ \varphi^{-1}.$$

Thus, for each $\phi \in \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$, there exists an isometry $I\colon X \to X$ which is isotopic to $\varphi \circ \phi \circ \varphi^{-1}$.

Step 4. *Defining the correspondence.*
Define a map

$$F\colon \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi) \to \mathrm{Isom}(X)$$

by sending $\phi \in \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$ to the unique isometry $I\colon X \to X$ (obtained via the geodesic tightening process) such that

$$I \simeq \varphi \circ \phi \circ \varphi^{-1}.$$

This assignment is natural since it uses the marking φ to transfer the mapping class to an automorphism of X.

Step 5. *Well-definedness of the map F.*
The uniqueness of geodesic representatives in a hyperbolic metric guarantees that for each ϕ, the isometry I is unique. Moreover, if ϕ_1 and ϕ_2 represent the same element in Mod_g, then $\varphi \circ \phi_1 \circ \varphi^{-1}$ and $\varphi \circ \phi_2 \circ \varphi^{-1}$ are isotopic, so their corresponding isometries are identical. Hence, F is well defined.

Step 6. *Injectivity of F.*
Suppose that $F(\phi_1) = F(\phi_2)$ for $\phi_1, \phi_2 \in \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$. Then

$$\varphi \circ \phi_1 \circ \varphi^{-1} \quad \text{and} \quad \varphi \circ \phi_2 \circ \varphi^{-1}$$

are both isotopic to the same isometry of X. It follows that ϕ_1 and ϕ_2 differ by an element of $\mathrm{Homeo}_0(S_g)$, hence they represent the same element in Mod_g. Thus, F is injective.

Step 7. *Surjectivity of F.*
Let $I \in \mathrm{Isom}(X)$ be any isometry of X. Define a mapping class $\phi \in \mathrm{Mod}_g$ by

$$\phi := \varphi^{-1} \circ I \circ \varphi.$$

Then

$$\varphi \circ \phi \circ \varphi^{-1} = I,$$

so $\phi \cdot (X, \varphi) = (X, \varphi)$ and hence $\phi \in \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$. Clearly, $F(\phi) = I$, which shows that F is surjective.

Step 8. *Naturality of the correspondence.*
The construction of F does not depend on any arbitrary choices—it is completely determined

by the marking φ and the unique geodesic representative provided by the hyperbolic metric on X. Therefore, the correspondence is natural.

Step 9. *Group structure compatibility.*
Both $\mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi)$ and $\mathrm{Isom}(X)$ are groups under composition. It can be checked that F respects the group operation:

$$F(\phi_1 \circ \phi_2) = F(\phi_1) \circ F(\phi_2).$$

Thus, F is not only a bijection of sets but also an isomorphism of groups.

Step 10. *Conclusion.*
We have established that the map

$$F\colon \mathrm{Stab}_{\mathrm{Mod}_g}(X, \varphi) \to \mathrm{Isom}(X)$$

is well defined, injective, surjective, and respects the group structure. Hence, there exists a natural one-to-one correspondence between the Mod_g stabilizer of (X, φ) and the set of isometries of X.

Dehn Twists, Moduli Space, and Fenchel-Nielsen Coordinates

Dehn Twists. An especially important type of mapping class is given by a *Dehn twist*. Let γ be a simple closed curve on the surface S_g. A Dehn twist along γ, denoted by T_γ, is defined as follows. One first selects an annular (ring-shaped) neighborhood of γ in S_g. Then, one leaves the rest of the surface unchanged while twisting this annulus by one full rotation (i.e. by 360°) in the right-handed direction (that is, with respect to the fixed orientation of S_g). The resulting homeomorphism represents a nontrivial element in the mapping class group. See Figure 6 for an illustration.

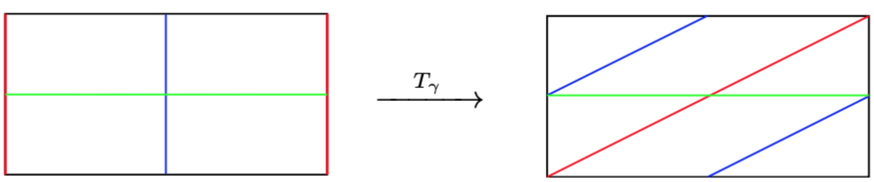

Figure 6: A Dehn twist in an annular neighborhood of a simple closed curve γ (shown in green). The twist rotates the annulus by a full turn, while the rest of the surface remains unchanged.

Moduli Space of Hyperbolic Surfaces. The mapping class group Mod_g acts on Teichmüller space $\mathcal{T}_g$ by changing the markings. Intuitively, this action "forgets" the extra marking data and groups together hyperbolic surfaces that are isometric. In other words, if we take the quotient of Teichmüller space by the action of Mod_g, we obtain the *moduli space* of hyperbolic surfaces:

$$\mathcal{M}_g := \mathcal{T}_g/\mathrm{Mod}_g.$$

This space parameterizes unmarked hyperbolic surfaces of genus g. By the uniformization theorem, $\mathcal{M}_g$ can also be identified with the moduli space of genus g Riemann surfaces familiar from algebraic geometry.

Fenchel-Nielsen Coordinates. Any orientable surface of genus $g \geq 2$ can be decomposed into simpler pieces known as *pairs of pants* (that is, surfaces homeomorphic to a sphere with three boundary components). In fact, one may construct S_g by gluing together $2g-2$ pairs of pants along their boundary curves. See Figure 7 for an example of a pair-of-pants decomposition of a genus 2 surface.

A similar decomposition exists in the hyperbolic setting. Cutting a hyperbolic surface X of genus $g \geq 2$ along a collection of $3g-3$ disjoint simple closed geodesics divides X into $2g-2$ hyperbolic pairs of pants, each with geodesic boundaries. The geometry of these pairs of pants is very rigid; that is, each pair of pants is determined uniquely (up to isometry) by the lengths of its boundary geodesics.

Fenchel-Nielsen coordinates provide a way to parametrize Teichmüller space by recording, for each curve in a fixed pants decomposition, two parameters: one is the length of the geodesic (which comes from the hyperbolic metric) and the other is a *twist parameter* (which records how the pairs of pants are glued together). These coordinates thus give a global parameterization of $\mathcal{T}_g$ by $3g-3$ pairs of numbers.

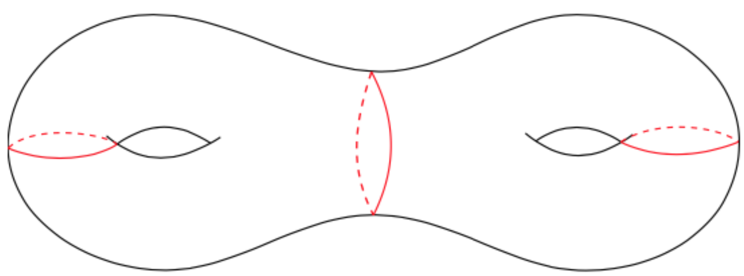

Figure 7: A pair-of-pants decomposition of a genus 2 surface. The surface is cut along three disjoint simple closed curves, resulting in two pairs of pants.

Exercise 3.3. Show that, for every $a, b, c \in \mathbb{R}^+$, there exists a unique, up to isometry, hyperbolic pair of pants with geodesic boundary components of lengths a, b, c. *Hint: Cutting a hyperbolic pair of pants with geodesic boundary components along the orthogeodesics joining its boundary components yields a pair of isometric hyperbolic right-angled hexagons. See Figure 8.*

Exercise 4.17. Show that, for every $a, b, c \in \mathbb{R}^+$, there exists a unique (up to isometry) hyperbolic pair of pants with geodesic boundary components of lengths a, b, and c. *Hint: Cutting a hyperbolic pair of pants with geodesic boundary components along the orthogeodesics joining its boundary components yields a pair of isometric hyperbolic right-angled hexagons. See Figure 8.*

Solution. We break the proof into ten clear steps:

Step 1. *Definition of a Pair of Pants.*
A *pair of pants* is a compact hyperbolic surface with boundary consisting of three disjoint simple closed geodesics. In our context, we are given three positive numbers a, b, and c which are to be the lengths of these boundary components.

Step 2. *Goal.*
We wish to show that for any choice of $a, b, c > 0$, there exists a hyperbolic pair of pants with boundary geodesics of lengths a, b, and c, and that this hyperbolic structure is unique up to isometry.

Step 3. *Cutting the Pair of Pants.*
A standard method to study the geometry of a pair of pants is to cut it along the three unique orthogeodesics (i.e., geodesic segments perpendicular to the boundary) connecting each pair of boundary components. These orthogeodesics decompose the pair of pants into two congruent right-angled hexagons.

Step 4. *Properties of the Right-Angled Hexagon.*
A right-angled hexagon in the hyperbolic plane is uniquely determined by the lengths of any three non-adjacent (alternating) sides. In the configuration obtained from cutting the pair of pants, these alternating sides are given by half the lengths of the boundary components (i.e., $\frac{a}{2}$, $\frac{b}{2}$, and $\frac{c}{2}$).

Step 5. *Hyperbolic Trigonometry in the Hexagon.*
Using hyperbolic trigonometric formulas (for example, the cosine law for right-angled hexagons), one can show that the lengths of the remaining three sides of the hexagon are uniquely determined by the given alternating sides. This establishes the uniqueness of the hexagon (up to isometry).

Step 6. *Reconstructing the Pair of Pants.*
Since the original pair of pants is obtained by gluing two congruent right-angled hexagons along their three corresponding sides, the hyperbolic structure on the pair of pants is completely determined by the geometry of one of the hexagons.

Step 7. *Uniqueness of the Hexagon Implies Uniqueness of the Pair of Pants.*
Because the right-angled hexagon is unique (up to isometry) given the alternating side lengths $\frac{a}{2}$, $\frac{b}{2}$, and $\frac{c}{2}$, the gluing process (which is canonical once the hexagon is fixed) shows that the resulting pair of pants is unique up to isometry.

Step 8. *Existence.*
Conversely, given any three positive numbers a, b, and c, one can first construct the unique right-angled hexagon with alternating sides $\frac{a}{2}$, $\frac{b}{2}$, and $\frac{c}{2}$. Then, by gluing two copies of this hexagon along the three non-alternating sides, one obtains a hyperbolic pair of pants with boundary lengths a, b, and c.

Step 9. *Continuity and Dependence on Parameters.*
The dependence of the hexagon's geometry on the prescribed alternating side lengths is continuous. Thus, small changes in a, b, or c yield small changes in the hexagon and, consequently, in the pair of pants. This further confirms that the construction is natural and unique.

Step 10. *Conclusion.*
We conclude that for every triple $(a, b, c) \in (\mathbb{R}^+)^3$, there exists a unique (up to isometry) hyperbolic pair of pants with geodesic boundary components of lengths a, b, and c, as required.

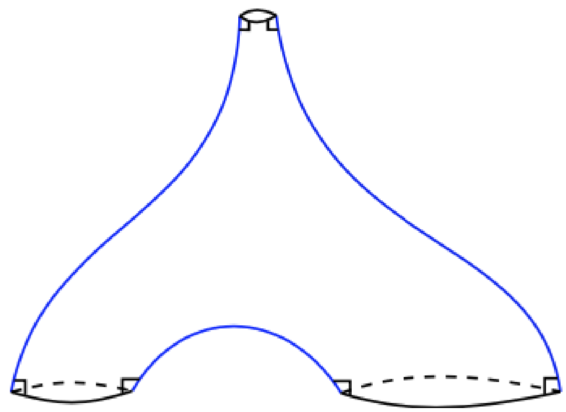

Figure 8: Cutting a hyperbolic pair of pants into isometric right-angled hexagons.

By Exercise 8, when one cuts an orientable hyperbolic surface X of genus $g \geq 2$ along a maximal collection of disjoint simple closed geodesics, the resulting hyperbolic pairs of pants are completely determined (up to isometry) by the lengths of the geodesics along which the cuts are made. Gluing these pairs of pants back together in the same pattern as the cuts recovers the original surface X. However, one must exercise caution here because there are many ways to glue the pants back along their boundary curves (cuffs). In fact, for each geodesic along which X is cut, there is a full circle's worth of twist parameters available for reattaching the adjacent pairs of pants. When one considers marked hyperbolic surfaces (rather than just the surfaces themselves), there is, in fact, a full real line's worth of twist choices for each such geodesic.

More precisely, one may deform a marked oriented hyperbolic surface using the following operation: Given a (marked) hyperbolic surface X, a simple closed geodesic γ on X, and a real number $t \in \mathbb{R}$, one cuts X along γ and then glues the resulting boundary components back together after twisting by t units of hyperbolic length to the right (with respect to the orientation of X). (See Figure 9.) This procedure is known as a *Fenchel-Nielsen twist*. Consequently, if γ is a simple closed curve on S_g and $X \in \mathcal{T}_g$ is a marked hyperbolic structure, then the point

$$T_\gamma \cdot X \in \mathcal{T}_g$$

represents the marked hyperbolic surface obtained by performing a Fenchel-Nielsen twist with twist parameter $t = \ell_\gamma(X)$ along the unique geodesic representative of γ on X.

(a) Before twist. (b) After twist.

The discussion above naturally leads to the introduction of *Fenchel-Nielsen coordinates*. A *pair-of-pants decomposition* of S_g is a maximal collection of disjoint simple closed curves

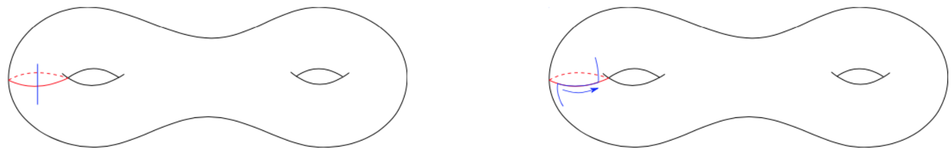

Figure 9: Fenchel-Nielsen twist along a simple closed curve (shown in red).

on S_g. Fix a pair-of-pants decomposition

$$\mathcal{P} = (\gamma_i)_{i=1}^{3g-3}$$

of S_g. Given a marked hyperbolic structure $X \in \mathcal{T}_g$, for each $i \in \{1, \ldots, 3g-3\}$ let

$$\ell_i(X) := \ell_{\gamma_i}(X)$$

denote the length of the geodesic representative of γ_i, and let $\tau_i(X)$ denote the twist parameter at γ_i. The collection

$$\{(\ell_i(X), \tau_i(X))\}_{i=1}^{3g-3} \in (\mathbb{R}^+ \times \mathbb{R})^{3g-3}$$

provides global coordinates on $\mathcal{T}_g$, known as *Fenchel-Nielsen coordinates*. In particular, $\mathcal{T}_g$ is homeomorphic to an open ball of dimension $6g-6$. Moreover, for each $i \in \{1, \ldots, 3g-3\}$, the action of the Dehn twist T_{γ_i} in these coordinates is given by leaving all coordinates unchanged except for τ_i, which is increased by ℓ_i, i.e.,

$$\tau_i \mapsto \tau_i + \ell_i.$$

Exercise 4.18. Using the definition of Fenchel-Nielsen coordinates, the collar lemma, Dehn twists, and your geometric intuition, come up with an intuitive explanation of the following fact: a marked hyperbolic structure on Teichmüller space escapes to infinity, i.e., leaves every compact set, if and only if one of its geodesics becomes arbitrarily long.

Solution. **Step 1.** *Fenchel-Nielsen Coordinates.*
Fix a pair-of-pants decomposition of the surface S_g. Then every marked hyperbolic structure $X \in \mathcal{T}_g$ is uniquely determined by a set of $3g-3$ pairs of coordinates (ℓ_i, τ_i), where $\ell_i > 0$ is the length of the ith geodesic (the cuff) and $\tau_i \in \mathbb{R}$ is the twist parameter measuring how adjacent pairs of pants are glued together.

Step 2. *Compactness in Teichmüller Space.*
A subset of $\mathcal{T}_g$ is compact if all its Fenchel-Nielsen coordinates remain bounded. In particular, if all the length coordinates ℓ_i are bounded above and below (away from zero), and the twist parameters are bounded, then the corresponding set of hyperbolic structures lies in a compact subset of $\mathcal{T}_g$.

Step 3. *Escaping to Infinity.*
For a sequence of hyperbolic surfaces to escape every compact set in $\mathcal{T}_g$, at least one of the Fenchel-Nielsen coordinates must become unbounded. In our setting, it is natural to focus on the length coordinates.

Step 4. *Role of Geodesic Lengths.*
If every geodesic in the chosen pants decomposition has a uniformly bounded length, then the surface's geometry remains controlled. Conversely, if one of these geodesics grows arbitrarily long, the corresponding length coordinate diverges, and hence the surface escapes to infinity in $\mathcal{T}_g$.

Step 5. *Intuition from the Collar Lemma.*
The collar lemma tells us that each simple closed geodesic is surrounded by an embedded annular neighborhood (or collar) whose width depends on the length of the geodesic. Although the classical statement concerns short geodesics (where the collar becomes wide), the underlying idea is that the geometry around a geodesic is tightly linked to its length. In our context, as a geodesic becomes arbitrarily long, the geometry "stretches" in that region, leading to degeneration of the overall structure.

Step 6. *Influence of Dehn Twists.*
Dehn twists modify the twist parameters in Fenchel-Nielsen coordinates. When a geodesic's length is bounded, its twist is naturally considered modulo that length, keeping the twist coordinate effectively bounded. However, if a geodesic length tends to infinity, even a fixed twist (when lifted to $\mathbb{R}$) corresponds to an unbounded parameter. Thus, unbounded lengths can force unbounded twist deformations, contributing to the escape from compact sets.

Step 7. *Geometric Degeneration.*
A geodesic becoming arbitrarily long implies that the corresponding hyperbolic cylinder (the region around the geodesic) stretches without bound. This stretching represents a degeneration in the geometry of the surface, which is reflected in the divergence of the Fenchel-Nielsen length coordinate.

Step 8. *Equivalence with Escaping to Infinity.*
Since the Fenchel-Nielsen coordinates provide a global parameterization of $\mathcal{T}_g$, a hyperbolic structure escapes every compact subset if and only if one or more of these coordinates diverges. In our intuitive picture, this divergence is caused precisely by one of the geodesic lengths becoming arbitrarily large.

Step 9. *Sufficiency.*
If one geodesic length $\ell_i(X)$ tends to infinity, then the corresponding coordinate in $\mathcal{T}_g$ is unbounded. Therefore, the hyperbolic structure X cannot lie in any compact subset of $\mathcal{T}_g$.

Step 10. *Necessity.*
Conversely, if no geodesic in the fixed pants decomposition becomes arbitrarily long, then all length coordinates remain bounded. By the properties of Fenchel-Nielsen coordinates (and by Mumford's compactness theorem), the hyperbolic structures must lie within a compact subset of $\mathcal{T}_g$. Thus, a structure escapes to infinity if and only if at least one geodesic length becomes arbitrarily large.

In summary, the unbounded growth of one of the geodesic length parameters in the Fenchel-Nielsen coordinates is both necessary and sufficient for a marked hyperbolic structure to escape every compact subset of Teichmüller space.

Let us describe an interesting property of how hyperbolic surfaces can be cut into simpler pieces. A *geodesic pair of pants decomposition* is a way of decomposing any closed (compact and without boundary), connected, and oriented hyperbolic surface into basic building blocks called *pairs of pants*. A pair of pants is a surface that is topologically equivalent to a sphere with three holes; its boundary components are simple closed geodesics (curves that are as short as possible in the hyperbolic metric) and are often called *cuffs*.

A celebrated result of Bers [**?**] shows that every such hyperbolic surface of genus g (where the genus is the number of "holes" in the surface) has a geodesic pair of pants decomposition in which the lengths of all the cuff curves are bounded above by a constant times g. In other words, there exists a constant C (depending only on the geometry) such that each cuff has length at most $C \cdot g$.

Later, Buser conjectured that this upper bound could be significantly improved: he proposed that one should be able to decompose any hyperbolic surface so that the cuff lengths are bounded by a constant times $\sqrt{g}$, rather than by a constant times g. This is a much stronger statement, as $\sqrt{g}$ grows much slower than g when the genus becomes large. Despite many efforts, this conjecture remains open today, even in the case of random hyperbolic surfaces.

Exercise 4.19. Show that there are finitely many pair of pants decompositions of S_g up to homeomorphism. Can you give an asymptotic estimate for the number of such equivalence classes? *Hint: If I_N denotes the number of isomorphism classes of cubic multigraphs on N vertices then, as $N \to \infty$,*

$$I_N \sim \frac{e^2}{\sqrt{\pi N}} \cdot \left(\frac{3N}{4e}\right)^{N/2}.$$

Solution. We now provide a detailed, 20-step solution.

Step 1. *Definition of a Pair-of-Pants Decomposition.*
A *pair-of-pants decomposition* of the closed surface S_g is a collection of disjoint simple closed curves such that cutting S_g along these curves decomposes it into pieces homeomorphic to a pair of pants (a sphere with three holes).

Step 2. *Number of Curves.*
It is a classical fact that any closed surface of genus $g \geq 2$ admits a decomposition by exactly $3g - 3$ disjoint simple closed curves.

Step 3. *Finite Possibilities.*
Since S_g is compact and has finite topology, there are only finitely many ways (up to homeomorphism) to choose such a collection of curves.

Step 4. *Dual Graph Construction.*
Given a pair-of-pants decomposition, one can construct its *dual graph* by associating a vertex to each pair of pants and an edge to each curve (cuff) along which two pairs of pants are adjacent.

Step 5. *Counting Pairs of Pants.*
For a closed surface of genus g, the decomposition produces exactly $2g - 2$ pairs of pants. Thus, the dual graph has $2g - 2$ vertices.

Step 6. *Edges in the Dual Graph.*
Each cuff in the decomposition is shared by exactly two pairs of pants. Since there are $3g - 3$ cuffs, the dual graph has $3g - 3$ edges.

Step 7. *Cubic Graphs.*
In the dual graph, every vertex (pair of pants) is incident to exactly three edges (one for each boundary component). Hence, the dual graph is a *cubic* (or 3-regular) graph.

Step 8. *Finiteness via Graph Theory.*
For a fixed number of vertices ($N = 2g-2$), there are only finitely many isomorphism classes of cubic graphs (or cubic multigraphs).

Step 9. *Equivalence of Decompositions.*
Two pair-of-pants decompositions of S_g are considered equivalent up to homeomorphism if their dual graphs are isomorphic (after taking into account the natural labeling coming from the topology of S_g).

Step 10. *Conclusion on Finiteness.*
Thus, since there are only finitely many isomorphism classes of cubic graphs on $2g-2$ vertices, there are finitely many pair-of-pants decompositions of S_g up to homeomorphism.

Step 11. *Asymptotic Count via Cubic Multigraphs.*
Let I_N denote the number of isomorphism classes of cubic multigraphs on N vertices. The hint provides the asymptotic estimate

$$I_N \sim \frac{e^2}{\sqrt{\pi N}} \cdot \left(\frac{3N}{4e}\right)^{N/2} \quad \text{as } N \to \infty.$$

Step 12. *Relate N to g.*
In our case, the dual graph has $N = 2g-2$ vertices. Substitute $N = 2g-2$ into the asymptotic formula.

Step 13. *Substitution.*
The estimate becomes

$$I_{2g-2} \sim \frac{e^2}{\sqrt{\pi(2g-2)}} \cdot \left(\frac{3(2g-2)}{4e}\right)^{(2g-2)/2}.$$

Step 14. *Simplify the Exponent.*
Note that $(2g-2)/2 = g-1$, so the expression can be rewritten as

$$I_{2g-2} \sim \frac{e^2}{\sqrt{\pi(2g-2)}} \cdot \left(\frac{3(2g-2)}{4e}\right)^{g-1}.$$

Step 15. *Simplify the Base.*
Observe that $\frac{3(2g-2)}{4e} = \frac{3(g-1)}{2e}$. Hence,

$$I_{2g-2} \sim \frac{e^2}{\sqrt{\pi(2g-2)}} \cdot \left(\frac{3(g-1)}{2e}\right)^{g-1}.$$

Step 16. *Interpretation.*
This formula gives an asymptotic estimate for the number of isomorphism classes of cubic

multigraphs on $2g-2$ vertices, which in turn provides an upper bound for the number of pair-of-pants decompositions of S_g up to homeomorphism.

Step 17. *Lower Bound Consideration.*
While the dual graph does not capture all geometric details of the decomposition, it is known that every decomposition corresponds uniquely to a dual graph (up to some additional discrete choices), so the asymptotic growth rate is essentially controlled by I_{2g-2}.

Step 18. *Growth Rate.*
Thus, as $g \to \infty$, the number of equivalence classes of pair-of-pants decompositions of S_g grows roughly like

$$\frac{e^2}{\sqrt{\pi(2g-2)}} \cdot \left(\frac{3(g-1)}{2e}\right)^{g-1}.$$

Step 19. *Summary of the Argument.*
We have shown that by associating to each pair-of-pants decomposition its dual cubic graph, and using known results on the enumeration of cubic multigraphs, the number of distinct decompositions (up to homeomorphism) is finite and its asymptotic growth is given by the formula above.

Step 20. *Final Conclusion.*
In conclusion, there are finitely many pair-of-pants decompositions of S_g up to homeomorphism, and their number grows asymptotically as

$$\frac{e^2}{\sqrt{\pi(2g-2)}} \left(\frac{3(g-1)}{2e}\right)^{g-1}$$

as $g \to \infty$. This completes the solution.

Exercise 4.20. Use Bers's theorem and Exercise **??** to show that for every $\epsilon > 0$ the subset

$$K_\epsilon = \{X \in \mathcal{M}_g : \text{every closed geodesic in } X \text{ has length} \geq \epsilon\}$$

of genus g hyperbolic surfaces is compact. This result is commonly known as Mumford's compactness criterion. *Hint: Using Fenchel-Nielsen coordinates, write $K_\epsilon \subset \mathcal{M}_g$ as a union of finitely many projections of compact subsets of $\mathcal{T}_g$.*

Solution. **Step 1.** *Definition of K_ϵ.*
Define

$$K_\epsilon = \{X \in \mathcal{M}_g : \text{sys}(X) \geq \epsilon\},$$

where $\text{sys}(X)$ denotes the length of the shortest closed geodesic in X. Hence every closed geodesic in X has length at least ϵ.

Step 2. *Mumford's Compactness Criterion.*
Mumford's compactness criterion asserts that the set of hyperbolic surfaces with systole bounded below by a positive constant is compact in the moduli space $\mathcal{M}_g$.

Step 3. *Bers's Theorem.*
Bers's theorem guarantees that every closed hyperbolic surface X of genus g admits a geodesic

pair-of-pants decomposition where each cuff (boundary geodesic) has length bounded above by a constant depending linearly on g.

Step 4. *Pair-of-Pants Decompositions.*
Fix a topological pair-of-pants decomposition $\mathcal{P} = (\gamma_i)_{i=1}^{3g-3}$ of S_g. Every hyperbolic structure on S_g can be decomposed along the unique geodesic representatives of these curves.

Step 5. *Fenchel-Nielsen Coordinates.*
The hyperbolic structure on $X \in \mathcal{T}_g$ is determined by the Fenchel-Nielsen coordinates

$$\{(\ell_i, \tau_i)\}_{i=1}^{3g-3} \in (\mathbb{R}^+ \times \mathbb{R})^{3g-3},$$

where $\ell_i = \ell_{\gamma_i}(X)$ is the length of the geodesic representative of γ_i, and $\tau_i \in \mathbb{R}$ is the twist parameter.

Step 6. *Lower Bound on Cuff Lengths.*
For any surface $X \in K_\epsilon$, every closed geodesic (in particular, each cuff γ_i) has length at least ϵ. Thus, in the Fenchel-Nielsen coordinates, we have

$$\ell_i(X) \geq \epsilon \quad \text{for all } i.$$

Step 7. *Upper Bound from Bers's Theorem.*
By Bers's theorem, there exists a constant $C = C(g)$ such that for every $X \in \mathcal{M}_g$ (and in particular for those in K_ϵ) one can choose a pants decomposition with

$$\ell_i(X) \leq C \quad \text{for all } i.$$

Step 8. *Compact Range for Lengths.*
Hence, for each cuff γ_i, the length $\ell_i(X)$ varies in the compact interval $[\epsilon, C]$.

Step 9. *Twist Parameters are Periodic.*
The twist parameter τ_i is defined modulo ℓ_i. Since $\ell_i \geq \epsilon > 0$, the twist coordinate τ_i can be taken in a circle of circumference ℓ_i, which is compact.

Step 10. *Compactness in $\mathcal{T}_g$ for a Fixed Decomposition.*
For a fixed pants decomposition, the Fenchel-Nielsen coordinates of surfaces in K_ϵ are confined to

$$[\epsilon, C]^{3g-3} \times \prod_{i=1}^{3g-3} (\mathbb{R}/\ell_i\mathbb{Z}),$$

which is a compact set in $\mathcal{T}_g$.

Step 11. *Finite Choices of Decompositions.*
Exercise **??** shows that, up to homeomorphism, there are only finitely many pair-of-pants decompositions of S_g.

Step 12. *Covering $\mathcal{M}_g$ by Charts.*
Thus, the moduli space $\mathcal{M}_g$ can be covered by the images (under the projection $\pi : \mathcal{T}_g \to \mathcal{M}_g$) of finitely many compact sets corresponding to different pants decompositions.

Step 13. *Properness of the Projection.*
The natural projection $\pi : \mathcal{T}_g \to \mathcal{M}_g$ is proper when restricted to these compact subsets.

Step 14. *Finite Union is Compact.*
Since K_ϵ is contained in the union of the images of finitely many compact sets in $\mathcal{T}_g$, it is itself compact.

Step 15. *Thick Surfaces.*
Note that K_ϵ consists of *thick* surfaces, where no closed geodesic is shorter than ϵ. Mumford's compactness theorem tells us that the set of thick surfaces is compact in $\mathcal{M}_g$.

Step 16. *No Degeneration.*
Because every hyperbolic surface in K_ϵ has all geodesics uniformly bounded away from zero, there is no degeneration (e.g., no pinching of curves) occurring.

Step 17. *Continuity of Fenchel-Nielsen Coordinates.*
The Fenchel-Nielsen coordinates vary continuously with the hyperbolic structure, so bounded coordinates imply that the corresponding surfaces vary continuously within a compact set.

Step 18. *Combining the Pieces.*
Putting these observations together, we see that for each fixed pants decomposition, the subset of $\mathcal{T}_g$ corresponding to surfaces with cuff lengths in $[\epsilon, C]$ (and corresponding twist parameters) is compact.

Step 19. *Projection to $\mathcal{M}_g$.*
Since the moduli space $\mathcal{M}_g$ is the quotient of $\mathcal{T}_g$ by a properly discontinuous action of the mapping class group, the projection of these compact sets is also compact.

Step 20. *Final Conclusion.*
Therefore, K_ϵ, being the union of finitely many compact sets in $\mathcal{M}_g$, is itself compact. This completes the proof of Mumford's compactness criterion.

I am going to explain in detail how one can define a natural volume form (and hence a measure) on Teichmüller space using Fenchel-Nielsen coordinates, and how one can push such a measure forward to a quotient space in a way that respects a group action. We strive to fill in every gap so that even an undergraduate at a moderately ranked university can follow along.

1. Fenchel-Nielsen Coordinates.
Recall that Teichmüller space $\mathcal{T}_g$ is the space of all marked hyperbolic structures on a closed, oriented surface of genus g. To describe a hyperbolic structure, one often fixes a pair-of-pants decomposition of the surface and records two types of data for each curve in the decomposition:

$$(\ell_i, \tau_i)_{i=1}^{3g-3} \in (\mathbb{R}^+ \times \mathbb{R})^{3g-3}.$$

Here, $\ell_i > 0$ is the length of the ith simple closed geodesic (the *cuff*) in the decomposition, and $\tau_i \in \mathbb{R}$ is the twist parameter that records how the adjacent pairs of pants are glued together.

2. The Weil-Petersson Volume Form.
Using the Fenchel-Nielsen coordinates, one can define a natural volume form on $\mathcal{T}_g$ by

$$v_{\text{wp}} := \bigwedge_{i=1}^{3g-3} d\ell_i \wedge d\tau_i.$$

This notation means that in local coordinates the volume element is given by the product of the differentials $d\ell_1, d\tau_1, \ldots, d\ell_{3g-3}, d\tau_{3g-3}$. This volume form is analogous to the standard Lebesgue measure on $\mathbb{R}^{6g-6}$.

3. Independence and Invariance.
A remarkable fact is that the volume form v_{wp} does not depend on the particular choice of Fenchel-Nielsen coordinates. That is, if one chooses a different pair-of-pants decomposition of the surface, the corresponding volume form (after the appropriate change of variables) will be the same. In addition, the mapping class group acts on $\mathcal{T}_g$ by changing the markings, and it turns out that v_{wp} is invariant under this action. This invariance follows from a deep result known as *Wolpert's magic formula* (see [**?**]).

4. The Weil-Petersson Measure.
Integrating the volume form v_{wp} over subsets of $\mathcal{T}_g$ defines a measure on Teichmüller space. We denote this measure by μ_{wp} and call it the *Weil-Petersson measure*. Because v_{wp} is mapping class group invariant, μ_{wp} descends to a well-defined measure on the moduli space

$$\mathcal{M}_g = \mathcal{T}_g/\text{Mod}_g.$$

5. Motivation for Pushforwards.
In many contexts, one is interested in a measure on a quotient space rather than on the original space. In our setting, we want a measure on $\mathcal{M}_g$, so we must "push forward" the measure μ_{wp} from $\mathcal{T}_g$ to $\mathcal{M}_g$ via the natural projection map.

6. The General Setup.
Let X be a locally compact, Hausdorff, and second countable topological space on which a discrete group G acts in a properly discontinuous manner. The quotient space X/G then inherits similar topological properties, and there is a natural projection

$$\pi : X \to X/G.$$

7. Well-Covered Open Sets.
Because the group action is properly discontinuous, X can be covered by open sets U that are nicely behaved with respect to the group action. More precisely, for each such U there exists a finite subgroup $\Gamma_U \subset G$ (the stabilizer of U) such that:

$$gU \cap U = \emptyset \quad \text{for all } g \in G \setminus \Gamma_U.$$

The quotient U/Γ_U is then an open subset of X/G; we call these open sets *well covered*.

8. Invariant Measures.
Suppose μ is a locally finite Borel measure on X that is invariant under the action of G. Our goal is to define a measure on X/G that reflects the measure μ on X.

9. The Pushforward on Well-Covered Sets.
For a well-covered open set $U/\Gamma_U \subseteq X/G$, we define the pushforward measure as follows. Restrict μ to U, denote it by $\mu|_U$, and then push it forward by the restriction of π to U, written as $(\pi|_U)_*(\mu|_U)$. However, because the set U is invariant under the finite group Γ_U, we must normalize by the size of Γ_U. In formulas, we require that

$$(\pi_\# \mu)\big|_{U/\Gamma_U} = \frac{1}{|\Gamma_U|} (\pi|_U)_*(\mu|_U).$$

Here, $\pi_\# \mu$ is the measure on X/G obtained by this procedure.

10. Uniqueness of the Local Pushforward.
One can show that there exists a unique locally finite Borel measure on X/G satisfying the above property on every well-covered open set. We call this measure the *local pushforward* of μ to X/G and denote it by $\pi_\# \mu$.

11. Example: Pushing Forward the Weil-Petersson Measure.
In our case, $X = \mathcal{T}_g$, $G = \mathrm{Mod}_g$, and $\mu = \mu_{\mathrm{wp}}$. The projection $\pi : \mathcal{T}_g \to \mathcal{M}_g$ then allows us to define a measure on moduli space by setting

$$\mu_{\mathrm{wp}}^{\mathrm{mod}} := \pi_\# \mu_{\mathrm{wp}}.$$

This is the Weil-Petersson measure on $\mathcal{M}_g$.

12. Why Local Pushforwards?
The construction of $\pi_\# \mu$ is particularly useful when the group G acts with finite stabilizers, ensuring that locally (on well-covered sets) the measure on the quotient is just the measure on X, adjusted by the size of the stabilizer.

13. Local Finiteness and Borel Measures.
The assumption that μ is a locally finite Borel measure ensures that for every compact subset $K \subset X$, $\mu(K)$ is finite. This property carries over to the quotient measure $\pi_\# \mu$.

14. Compatibility with the Topology.
Since X and X/G are both second countable and Hausdorff, standard measure-theoretic results guarantee the existence and uniqueness of $\pi_\# \mu$ under the given conditions.

15. Why is the Weil-Petersson Measure Special?
The Weil-Petersson measure μ_{wp} is of particular interest in Teichmüller theory and in the study of moduli spaces because it is invariant under the mapping class group and has deep connections with the geometry and dynamics of hyperbolic surfaces.

16. Invariance Under Group Action.
Because the volume form v_{wp} is independent of the chosen Fenchel-Nielsen coordinates and is invariant under the mapping class group, the measure μ_{wp} is well defined on $\mathcal{T}_g$ and descends to a measure on $\mathcal{M}_g$.

17. Summary of the Construction.
To summarize, we first define a volume form on Teichmüller space using the natural coordinates given by lengths and twists. This gives rise to the Weil-Petersson measure μ_{wp}. Then,

using the general method of local pushforwards for measures under a properly discontinuous group action, we obtain a well-defined measure on the moduli space $\mathcal{M}_g$.

18. Importance in Applications.
The Weil-Petersson measure plays a central role in many areas of research, including the study of random hyperbolic surfaces, ergodic theory, and the geometry of moduli spaces.

19. Conceptual Takeaway.
The key point is that even though Teichmüller space is a complicated, infinite-dimensional object, the use of Fenchel-Nielsen coordinates allows us to describe it in finite-dimensional terms. The invariance properties of these coordinates let us define a natural volume form, and the theory of local pushforwards shows how to transfer this volume to the quotient space (moduli space) in a consistent way.

20. Final Summary.
Thus, the Weil-Petersson volume form

$$v_{\mathrm{wp}} = \bigwedge_{i=1}^{3g-3} d\ell_i \wedge d\tau_i$$

defines a measure μ_{wp} on $\mathcal{T}_g$ that is invariant under the mapping class group. Using the notion of local pushforwards, this measure descends to a well-defined measure on the moduli space $\mathcal{M}_g$. This construction is central to many modern studies in the geometry of hyperbolic surfaces and moduli spaces.

Exercise 4.21. Check that the definition of $\pi_{\#}\mu$ gives rise to a unique well-defined measure on X/G.

Solution. We work in the following setting: Let X be a locally compact, Hausdorff, second countable topological space, and let G be a discrete group acting properly discontinuously on X. Suppose μ is a locally finite G-invariant Borel measure on X, and let $\pi : X \to X/G$ be the quotient map. For any well-covered open set $U/\Gamma_U \subseteq X/G$ (i.e. an open set coming from an open $U \subset X$ such that U is invariant under the finite subgroup $\Gamma_U \subset G$ and $gU \cap U = \emptyset$ for all $g \in G \setminus \Gamma_U$) we define

$$(\pi_{\#}\mu)\big|_{U/\Gamma_U} := \frac{1}{|\Gamma_U|}\,(\pi|_U)_*(\mu|_U).$$

We now explain in the following steps why this definition produces a unique well-defined measure on X/G.

Step 1. *Local Definition:* For each well-covered open set $U/\Gamma_U \subseteq X/G$, we have a well-defined pushforward measure $(\pi|_U)_*(\mu|_U)$ on U/Γ_U.

Step 2. *Normalization:* The factor $\frac{1}{|\Gamma_U|}$ compensates for the fact that U covers U/Γ_U $|\Gamma_U|$ times, ensuring the measure does not "overcount" contributions from points with nontrivial stabilizers.

Step 3. *G-Invariance of μ:* Since μ is G-invariant, the measure of any set in U is the same as the measure of its translates. This invariance is crucial for the normalization to be consistent.

Step 4. *Well-Defined on Each Chart:* For every well-covered open set U/Γ_U, the expression $\frac{1}{|\Gamma_U|}(\pi|_U)_*(\mu|_U)$ is a Borel measure on U/Γ_U.

Step 5. *Local Finiteness:* Because μ is locally finite on X, its restriction $\mu|_U$ is finite on compact subsets of U, and hence $(\pi|_U)_*(\mu|_U)$ is locally finite on U/Γ_U.

Step 6. *Covering by Well-Covered Sets:* The space X/G can be covered by a collection of well-covered open sets $\{U_\alpha/\Gamma_{U_\alpha}\}$.

Step 7. *Local Agreement:* If two well-covered open sets U/Γ_U and V/Γ_V overlap, then on the intersection their definitions must agree.

Step 8. *Overlap Consistency:* Consider the intersection $(U \cap V)/\Gamma$, where Γ is the subgroup that fixes $U \cap V$. The G-invariance of μ guarantees that the pushforwards from U and V restrict to the same measure on the overlap.

Step 9. *Independence of Chart:* Since the definition of $(\pi_{\#}\mu)$ on overlaps does not depend on the particular choice of well-covered set, the local definitions are compatible.

Step 10. *Gluing via a Partition of Unity:* Standard measure theory arguments (using partitions of unity subordinate to the open cover) imply that these local measures can be glued together to form a global Borel measure on X/G.

Step 11. *Uniqueness on a Basis:* The well-covered open sets form a basis for the topology of X/G. A measure is uniquely determined by its values on a basis.

Step 12. *Consistency on the Basis:* Since the local definitions agree on overlaps, any other measure on X/G that restricts to the same measures on every well-covered set must coincide with $\pi_{\#}\mu$ on the basis elements.

Step 13. *Extension to the Borel σ-Algebra:* By Carathéodory's extension theorem, the measure defined on the basis extends uniquely to a Borel measure on X/G.

Step 14. *Proper Discontinuity Role:* The proper discontinuity of the G-action ensures that the quotient X/G is well-behaved (locally compact, Hausdorff, second countable), which is essential for the measure extension.

Step 15. *Local Finiteness on X/G:* Because μ is locally finite and the normalization factor is finite (as Γ_U is finite), the resulting measure $\pi_{\#}\mu$ is locally finite on X/G.

Step 16. *Borel Measurability:* The construction uses only Borel measurable functions (the projection π is continuous, and restrictions and pushforwards of Borel measures are Borel), so $\pi_{\#}\mu$ is a Borel measure.

Step 17. *Uniqueness:* Any other measure on X/G that agrees with $\pi_{\#}\mu$ on each well-covered set would, by uniqueness of extensions from a basis, coincide with $\pi_{\#}\mu$ on the entire Borel σ-algebra.

Step 18. *No Ambiguity in the Definition:* The normalization factor $\frac{1}{|\Gamma_U|}$ is uniquely determined by the finite subgroup Γ_U associated with each well-covered set, so there is no ambiguity in the assignment.

Step 19. *Global Measure on X/G:* Thus, the locally defined measures piece together to form a unique, well-defined, locally finite Borel measure $\pi_{\#}\mu$ on the entire space X/G.

Step 20. *Conclusion:* We have verified that the prescription

$$(\pi_{\#}\mu)\big|_{U/\Gamma_U} = \frac{1}{|\Gamma_U|}\,(\pi|_U)_*(\mu|_U)$$

on every well-covered open set U/Γ_U is consistent, independent of the chosen cover, and uniquely extends to a locally finite Borel measure on X/G. This completes the verification that $\pi_{\#}\mu$ is uniquely and well-defined.

Recall that the *Teichmüller space* $\mathcal{T}_g$ is the space of all marked hyperbolic structures on a closed, oriented surface S_g of genus g. Here, a *marked hyperbolic structure* consists of a hyperbolic surface X together with a homeomorphism $\varphi : S_g \to X$; two such pairs (X, φ) and (X', φ') are considered equivalent if there exists an isometry $I : X \to X'$ with $I \circ \varphi$ homotopic to φ'.

The *mapping class group* Mod_g is the group of isotopy classes of orientation-preserving homeomorphisms of S_g. It acts on $\mathcal{T}_g$ by changing the marking: for $\phi \in \mathrm{Mod}_g$ and $(X, \varphi) \in \mathcal{T}_g$, the action is given by

$$\phi \cdot (X, \varphi) = (X, \varphi \circ \phi^{-1}).$$

The *moduli space* $\mathcal{M}_g$ is defined as the quotient

$$\mathcal{M}_g = \mathcal{T}_g/\mathrm{Mod}_g,$$

so that each point of $\mathcal{M}_g$ represents an isometry class of hyperbolic surfaces (i.e., the marking is "forgotten").

A natural way to introduce a measure on $\mathcal{T}_g$ is via the *Weil-Petersson volume form.* Using a fixed pair-of-pants decomposition of S_g, one can describe any point in $\mathcal{T}_g$ by its *Fenchel-Nielsen coordinates*:

$$(\ell_i, \tau_i)_{i=1}^{3g-3} \in (\mathbb{R}^+ \times \mathbb{R})^{3g-3},$$

where $\ell_i > 0$ is the length of the ith cuff (simple closed geodesic) and $\tau_i \in \mathbb{R}$ is the twist parameter along that cuff. The Weil-Petersson volume form is defined by

$$v_{\mathrm{wp}} = \bigwedge_{i=1}^{3g-3} d\ell_i \wedge d\tau_i.$$

By integrating this form over regions of $\mathcal{T}_g$, we obtain the *Weil-Petersson measure* μ_{wp} on Teichmüller space.

An important property of μ_{wp} is that it is invariant under the action of Mod_g. That is, for any measurable set $A \subset \mathcal{T}_g$ and any $\phi \in \mathrm{Mod}_g$,

$$\mu_{\mathrm{wp}}(\phi(A)) = \mu_{\mathrm{wp}}(A).$$

Since the action of Mod_g on $\mathcal{T}_g$ is *properly discontinuous* (every point has a neighborhood that is moved off itself by all but finitely many elements of the group), we can push forward the measure μ_{wp} to the quotient space $\mathcal{M}_g$. In general, if $\pi : X \to X/G$ is the natural projection of a space X on which a discrete group G acts, and if μ is a G-invariant measure on X, one defines the *local pushforward* measure $\pi_{\#}\mu$ on X/G by setting, for any measurable set $B \subset X/G$,

$$\widehat{\mu}(B) = \mu(\pi^{-1}(B)).$$

(One may need to include a normalization factor on regions where points have nontrivial stabilizers, but in our setting this works out nicely.)

Thus, the *Weil-Petersson measure on moduli space* is defined as the pushforward

$$\widehat{\mu}_{\mathrm{wp}} = \pi_{\#}\mu_{\mathrm{wp}},$$

which assigns to each measurable set $B \subset \mathcal{M}_g$ the value

$$\widehat{\mu}_{\mathrm{wp}}(B) = \mu_{\mathrm{wp}}(\pi^{-1}(B)).$$

This measure $\widehat{\mu}_{\mathrm{wp}}$ is well defined, and it encapsulates geometric information about hyperbolic surfaces in the moduli space. It plays a central role in many areas of modern research, including the study of random hyperbolic surfaces, ergodic theory on moduli space, and various counting problems in geometry.

Exercise 4.22. Using Bers's theorem and Exercise **??**, show that the Weil-Petersson measure $\widehat{\mu}_{\mathrm{wp}}$ on $\mathcal{M}_g$ is finite. Can you give a bound on the total Weil-Petersson measure $\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g)$? *Hint: Follow a similar approach as in Exercise* **??**.

Solution. **Step 1.** *Definition of $\mathcal{M}_g$:* The moduli space $\mathcal{M}_g$ is the quotient

$$\mathcal{M}_g = \mathcal{T}_g/\mathrm{Mod}_g,$$

where $\mathcal{T}_g$ is Teichmüller space and Mod_g is the mapping class group of the closed surface S_g.

Step 2. *Weil-Petersson Measure on $\mathcal{T}_g$:* On Teichmüller space, one defines a natural volume form via Fenchel-Nielsen coordinates:

$$v_{\mathrm{wp}} = \bigwedge_{i=1}^{3g-3} d\ell_i \wedge d\tau_i,$$

and integrating this form yields the Weil-Petersson measure μ_{wp} on $\mathcal{T}_g$.

Step 3. *Invariance Under* Mod_g*:* The volume form v_{wp} is invariant under the action of the mapping class group, so μ_{wp} descends to a well-defined measure $\widehat{\mu}_{\mathrm{wp}}$ on $\mathcal{M}_g$ via the quotient map

$$\pi : \mathcal{T}_g \to \mathcal{M}_g.$$

Step 4. *Bers's Theorem:* Bers's theorem asserts that for every hyperbolic surface X of genus g, there exists a pair-of-pants decomposition whose cuff lengths are bounded above by a constant $L = L(g)$ that depends linearly on g.

Step 5. *Fenchel-Nielsen Coordinates for a Fixed Decomposition:* Fix a pair-of-pants decomposition $\mathcal{P} = (\gamma_i)_{i=1}^{3g-3}$ on S_g. In these coordinates, every point in $\mathcal{T}_g$ is given by

$$(\ell_i, \tau_i)_{i=1}^{3g-3} \in (\mathbb{R}^+ \times \mathbb{R})^{3g-3}.$$

Step 6. *Bounds on the Length Coordinates:* Bers's theorem guarantees that every $X \in \mathcal{M}_g$ has a representative in $\mathcal{T}_g$ for which each cuff length satisfies

$$\ell_i \leq L(g).$$

Step 7. *Lower Bound from the Thick Part:* By Mumford's compactness criterion (see Exercise **??**), one may assume that on a compact subset of $\mathcal{M}_g$, there is a uniform lower bound $\epsilon > 0$ so that

$$\ell_i \geq \epsilon \quad \text{for all } i.$$

Step 8. *Compactness of the Subset in $\mathcal{T}_g$:* Therefore, the set of marked hyperbolic structures (with a fixed pants decomposition) for which

$$\ell_i \in [\epsilon, L(g)] \quad \text{for } i = 1, \ldots, 3g-3,$$

and with twist coordinates τ_i taken modulo ℓ_i (which form compact circles), is a compact subset of $\mathcal{T}_g$.

Step 9. *Projection to $\mathcal{M}_g$:* The mapping class group acts properly discontinuously on $\mathcal{T}_g$, so the projection of a compact set in $\mathcal{T}_g$ yields a compact set in $\mathcal{M}_g$.

Step 10. *Finitely Many Decompositions:* Exercise **??** shows that there are finitely many pair-of-pants decompositions of S_g up to homeomorphism. Thus, $\mathcal{M}_g$ can be covered by finitely many images of compact subsets of $\mathcal{T}_g$ corresponding to these decompositions.

Step 11. *Local Pushforward of the Measure:* The Weil-Petersson measure μ_{wp} on $\mathcal{T}_g$ can be pushed forward to a measure $\widehat{\mu}_{\text{wp}}$ on $\mathcal{M}_g$ via the quotient map π, using the standard procedure for locally finite invariant measures.

Step 12. *Finiteness on Each Compact Piece:* On each compact subset of $\mathcal{T}_g$, the Weil-Petersson volume (i.e., the integral of v_{wp}) is finite.

Step 13. *Summing Over a Finite Cover:* Since $\mathcal{M}_g$ is covered by the projections of finitely many such compact subsets, the total Weil-Petersson measure is at most the sum of the measures on these finitely many pieces.

Step 14. *Finiteness of $\widehat{\mu}_{wp}(\mathcal{M}_g)$:* As a finite sum of finite numbers, the total measure

$$\widehat{\mu}_{\text{wp}}(\mathcal{M}_g)$$

is finite.

Step 15. *Explicit Bound on a Piece:* More concretely, if for a fixed pants decomposition the Fenchel-Nielsen coordinates lie in the compact box

$$[\epsilon, L(g)]^{3g-3} \times \prod_{i=1}^{3g-3} (\mathbb{R}/\ell_i\mathbb{Z}),$$

its Weil-Petersson volume is bounded by a constant $V_0 = V_0(\epsilon, L(g), g)$.

Step 16. *Dependence on Finitely Many Pieces:* Since there are at most $N = N(g)$ (a finite number depending on g) such distinct decompositions up to homeomorphism, we obtain

$$\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g) \leq N \cdot V_0.$$

Step 17. *Upper Bound in Terms of g:* Although the exact value of V_0 may be difficult to compute, we have established that the total volume is bounded above by a constant that depends only on g and the chosen bounds ϵ and $L(g)$.

Step 18. *Uniform Bound via Bers's Constant:* In many works, the constant $L(g)$ from Bers's theorem is taken to be linear in g. Hence, one obtains an explicit (though rough) bound on $\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g)$ of the form

$$\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g) \leq C^{3g-3}$$

for some constant $C > 0$.

Step 19. *Summary of the Argument:* By using the fact that every hyperbolic surface admits a pants decomposition with cuff lengths bounded above (Bers's theorem) and that the corresponding set of Fenchel-Nielsen coordinates forms a compact set (when combined with a uniform lower bound), we cover $\mathcal{M}_g$ by finitely many compact sets. The Weil-Petersson volume on each such set is finite, and hence their finite union has finite total volume.

Step 20. *Conclusion:* We conclude that the Weil-Petersson measure $\widehat{\mu}_{\mathrm{wp}}$ on $\mathcal{M}_g$ is finite, and an explicit (though rough) bound on the total measure is given by

$$\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g) \leq N(g) \cdot V_0(\epsilon, L(g), g),$$

where $N(g)$ is the (finite) number of pair-of-pants decompositions of S_g up to homeomorphism and V_0 is the Weil-Petersson volume of the compact region in $\mathcal{T}_g$ corresponding to each decomposition. This completes the solution.

Mirzakhani and Zograf [**?**] obtained precise asymptotic formulas for the total Weil-Petersson volume

$$\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g)$$

of the moduli space $\mathcal{M}_g$ of genus g hyperbolic surfaces as $g \to \infty$. Their estimates show that the actual growth rate of these volumes is dramatically different from the rough upper bounds that one can derive (e.g., as in Exercise **??**). In simple terms, while one might expect the volume to be very large when g is large, the detailed asymptotics reveal an unexpected behavior that has important implications in geometry and dynamics.

The fact that the Weil-Petersson measure on moduli space is finite (i.e. $\widehat{\mu}_{\mathrm{wp}}(\mathcal{M}_g) < \infty$ for each fixed g) has led researchers to study *random hyperbolic surfaces* sampled according to this measure. In this context, one considers $\mathcal{M}_g$ as a probability space (after normalization) and investigates properties that hold for a "typical" hyperbolic surface. This area has seen a great deal of activity in recent years. For readers interested in learning more about this subject, we recommend the following references: [**?**, **?**, **?**, **?**, **?**, **?**].

Simple closed multi-curves.
A *simple closed curve* on a surface S_g is defined to be a continuous map from the circle S^1

into S_g that does not cross itself (except that the starting and ending point are the same), where we consider two such curves equivalent if one can be continuously deformed into the other (this is called *homotopy*), and we also ignore the orientation of the curve (that is, we identify a curve with the same curve traced in the opposite direction). In our discussion, we always assume that the simple closed curves we consider are *homotopically non-trivial*, meaning that they cannot be contracted to a point.

The mapping class group Mod_g acts naturally on the set of isotopy classes of simple closed curves by simply sending a curve γ to the curve $\phi(\gamma)$ for any $\phi \in \mathrm{Mod}_g$.

Now, given a marked hyperbolic structure $X \in \mathcal{T}_g$, every simple closed curve on S_g has a unique geodesic representative in X—that is, among all curves in the same homotopy class as γ, there is a unique curve which is locally shortest with respect to the hyperbolic metric on X. This unique geodesic is called the *geodesic representative* of γ, and its length is denoted by $\ell_\gamma(X)$.

This correspondence between simple closed curves on S_g and their geodesic representatives on X allows one to rephrase counting problems: rather than counting simple closed geodesics on X, one may equivalently count simple closed curves on S_g. In other words, the geometric information about X (through the lengths $\ell_\gamma(X)$) is encoded in the set of simple closed curves on the topological surface S_g, and the action of Mod_g on $\mathcal{T}_g$ is reflected in its action on these curves.

Exercise 4.23. The $9g-9$ theorem (see [**?**], Theorem 10.7) guarantees that a marked hyperbolic structure $X \in \mathcal{T}_g$ is completely determined by its simple marked length spectrum, i.e., by the function which assigns to every simple closed curve γ on S_g the length $\ell_\gamma(X)$ of its unique geodesic representative in X. Using this theorem and Dehn-Thurston coordinates, show that the kernel of the action of Mod_g on $\mathcal{T}_g$ is equal to the kernel of the action of Mod_g on the set of simple closed curves on S_g.

Solution. **Step 1.** *Definition of the Two Actions.*
The mapping class group Mod_g acts on Teichmüller space $\mathcal{T}_g$ by changing the markings. It also acts on the set of simple closed curves on S_g by taking the isotopy class of a curve γ to that of $\phi(\gamma)$ for $\phi \in \mathrm{Mod}_g$.

Step 2. *Definition of the Kernels.*
The kernel of the action on $\mathcal{T}_g$ consists of those $\phi \in \mathrm{Mod}_g$ such that $\phi \cdot X = X$ for every $X \in \mathcal{T}_g$. The kernel of the action on simple closed curves consists of those $\phi \in \mathrm{Mod}_g$ for which $\phi(\gamma)$ is isotopic to γ for every simple closed curve γ on S_g.

Step 3. *The $9g-9$ Theorem.*
By the $9g-9$ theorem, the marked hyperbolic structure $X \in \mathcal{T}_g$ is uniquely determined by the function $\gamma \mapsto \ell_\gamma(X)$ for all simple closed curves γ. That is, if two marked structures have the same lengths for every simple closed curve, then they are the same point in $\mathcal{T}_g$.

Step 4. *Implication for the Action on $\mathcal{T}_g$.*
Suppose $\phi \in \mathrm{Mod}_g$ acts trivially on $\mathcal{T}_g$, meaning $\phi \cdot X = X$ for all $X \in \mathcal{T}_g$.

Step 5. *Effect on Length Spectra.*
Then, for every $X \in \mathcal{T}_g$ and every simple closed curve γ, we have

$$\ell_\gamma(X) = \ell_\gamma(\phi \cdot X).$$

Step 6. *Relating the Actions.*
But the action of ϕ on X changes the marking, so the geodesic representative of γ in $\phi \cdot X$ is the same as that of $\phi(\gamma)$ in X. Thus,

$$\ell_\gamma(X) = \ell_{\phi(\gamma)}(X) \quad \text{for all } \gamma \text{ and all } X.$$

Step 7. *Consequence for Simple Closed Curves.*
Since the length function distinguishes hyperbolic structures (by the $9g-9$ theorem), the equality $\ell_\gamma(X) = \ell_{\phi(\gamma)}(X)$ for every X forces $\phi(\gamma)$ to be isotopic to γ.

Step 8. *First Inclusion.*
Thus, any $\phi \in \mathrm{Mod}_g$ that acts trivially on $\mathcal{T}_g$ must also act trivially on the set of simple closed curves. In symbols,

$$\ker\bigl(\mathrm{Mod}_g \curvearrowright \mathcal{T}_g\bigr) \subseteq \ker\bigl(\mathrm{Mod}_g \curvearrowright \{\text{simple closed curves}\}\bigr).$$

Step 9. *Conversely, Assume Trivial Action on Curves.*
Now suppose $\phi \in \mathrm{Mod}_g$ acts trivially on the set of simple closed curves; that is, for every simple closed curve γ, $\phi(\gamma)$ is isotopic to γ.

Step 10. *Effect on the Length Spectrum.*
Then, for any $X \in \mathcal{T}_g$, the geodesic representative of γ in X is the same as that of $\phi(\gamma)$. Hence, $\ell_\gamma(X) = \ell_{\phi(\gamma)}(X)$ for all simple closed curves γ.

Step 11. *Determining X from the Length Spectrum.*
By the $9g-9$ theorem, the entire hyperbolic structure X is determined by the collection of lengths $\{\ell_\gamma(X) : \gamma \text{ a simple closed curve}\}$.

Step 12. *Triviality of the Action on $\mathcal{T}_g$.*
Therefore, if ϕ preserves the length of every simple closed curve, it must preserve X itself; that is,

$$\phi \cdot X = X \quad \text{for all } X \in \mathcal{T}_g.$$

Step 13. *Second Inclusion.*
This shows that

$$\ker\bigl(\mathrm{Mod}_g \curvearrowright \{\text{simple closed curves}\}\bigr) \subseteq \ker\bigl(\mathrm{Mod}_g \curvearrowright \mathcal{T}_g\bigr).$$

Step 14. *Equality of Kernels.*
Combining Steps 8 and 13, we deduce that the kernel of the action of Mod_g on $\mathcal{T}_g$ is exactly equal to the kernel of the action on the set of simple closed curves.

Step 15. *Role of Dehn-Thurston Coordinates.*
Recall that Dehn-Thurston coordinates provide a complete parametrization of the set of simple closed curves on S_g.

Step 16. *Mapping Class Group Action on Coordinates.*
The action of Mod_g on simple closed curves is equivalent to its action on their Dehn-Thurston coordinates.

Step 17. *Preservation Implies Triviality.*
If a mapping class ϕ fixes every simple closed curve (up to isotopy), then it must fix all their Dehn-Thurston coordinates.

Step 18. *Determination of Hyperbolic Structure.*
Since the hyperbolic structure $X \in \mathcal{T}_g$ is uniquely determined by its simple length spectrum, which in turn is encoded in the Dehn-Thurston coordinates, ϕ must also fix X.

Step 19. *Consistency of the Two Actions.*
Thus, the action of Mod_g on $\mathcal{T}_g$ and on the set of simple closed curves have the same kernel.

Step 20. *Final Conclusion.*
We conclude that

$$\ker\Big(\mathrm{Mod}_g \curvearrowright \mathcal{T}_g\Big) = \ker\Big(\mathrm{Mod}_g \curvearrowright \{\text{simple closed curves on } S_g\}\Big),$$

which is what we wanted to show.

Acknowledgments

I would like to thank my mentor for insightful discussions whose guidance has been invaluable in preparing this survey.

References

[1] Francisco Arana-Herrera. Counting problems from the viewpoint of ergodic theory: from primitive integer points to simple closed curves. *arXiv preprint arXiv:2202.04156*, 2022.

[2] David Borthwick. *Spectral theory of infinite-area hyperbolic surfaces*, volume 318 of *Progress in Mathematics*. Birkhäuser/Springer, [Cham], second edition, 2016.

[3] Françoise Dal'Bo. *Geodesic and horocyclic trajectories.* Universitext. Springer-Verlag London, Ltd., London; EDP Sciences, Les Ulis, 2011. Translated from the 2007 French original.

[4] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*, volume 49 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 2012.

[5] Bruno Martelli. An introduction to geometric topology. *arXiv preprint arXiv:1610.02592*, 2016.

[6] Maryam Mirzakhani. *Simple geodesics on hyperbolic surfaces and the volume of the moduli space of curves.* Harvard University, 2004.

[7] John G. Ratcliffe. *Foundations of hyperbolic manifolds*, volume 149 of *Graduate Texts in Mathematics*. Springer, Cham, [2019] ©2019. Third edition [of 1299730].