An Introduction to Hyperbolic Geometry

Prof. Abhijit Pal

Department of Mathematics and Statistics

Indian Institute of Technology – Kanpur

Lecture – 29

Geodesic Structures on Surfaces: From Euclidean and Spherical to Hyperbolic Geometries

Hello! We have previously explored geodesics in both the hyperbolic plane and the Euclidean plane. In general, geodesics represent length-minimizing paths. In these two types of planes, the nature of geodesics is relatively straightforward. For instance, in the hyperbolic plane, specifically when using the upper half-plane model, geodesics manifest as vertical lines and semicircles, with their centers lying on the real axis. In contrast, in the Euclidean plane, geodesics are represented by straight lines.

Now, when we consider a surface endowed with a metric, identifying geodesics as length-minimizing paths becomes much more complex. Take the sphere, for example. One can prove that an arc of a great circle serves as a geodesic. On the sphere, if you select two points, particularly antipodal points, you can draw two arcs of a great circle, each of which qualifies as a geodesic.

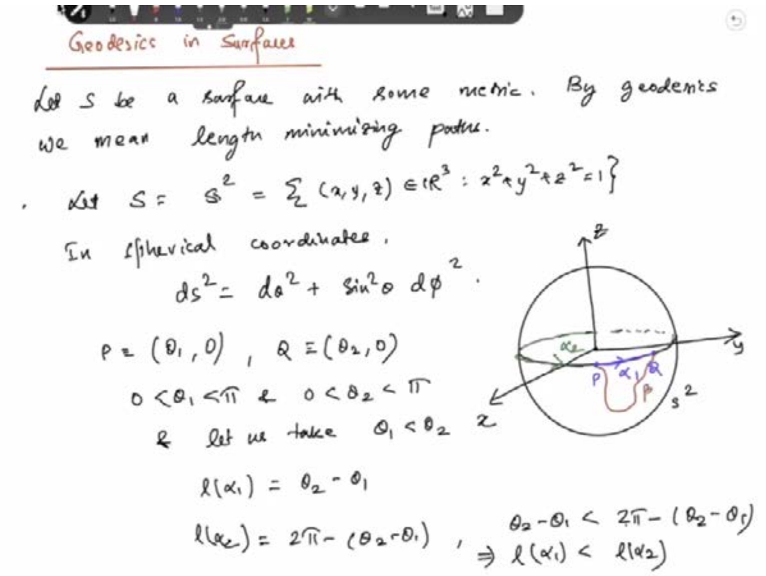
Consequently, the sphere does not qualify as a unique geodesic space. Unlike the hyperbolic and Euclidean planes, which both exhibit unique geodesics between any two points, the unit sphere allows for multiple geodesic paths. However, life on the sphere remains relatively simple; the nature of geodesics here is quite manageable.

Now, let's shift our focus to the torus or any closed orientable surface, particularly one of genus g greater than or equal to 2. Such surfaces admit a hyperbolic metric. In this context, the geodesics on these surfaces, whether they be toroidal or hyperbolic, become significantly more intricate. In the upcoming discussion, we will delve deeper into these complexities.

Let us consider a surface, denoting it as S, equipped with a certain metric. When we refer to geodesics, we mean length-minimizing paths. To illustrate this, we will first take the surface to be the sphere S^2, which is the unit sphere in R^3. In spherical coordinates, the Riemannian metric can be expressed as follows:

$$ds^2 = d\theta^2 + \sin^2\theta \, d\phi^2.$$

(Refer Slide Time: 02:50)



Now, let’s identify two points on this unit sphere and determine the length-minimizing path between them. Here is a depiction of the sphere: consider the vertical axis as the z-axis, the horizontal axis as the x-axis, and another horizontal axis as the y-axis. The intersection of the unit sphere S^2 with the xy-plane defines what we refer to as the meridian circle.

Next, let us select two points on this meridian circle, which we will label P and Q. In terms of spherical coordinates, we can express P as $(\theta_1, 0)$ and Q as $(\theta_2, 0)$, where θ_1 and θ_2 are angles that lie within the interval $(0, \pi)$, specifically $0 < \theta_1 < \pi$ and $0 < \theta_2 < \pi$. For the sake of clarity, we will also assume that $\theta_1 < \theta_2$.

Now, observe that there are two arcs of the great circle that connect points P and Q. Each arc represents a potential geodesic on the sphere, and this leads us to explore their properties further.

Let’s consider two paths on the sphere: one path, which we will denote as α_1, and the other, α_2, which is the complementary arc of the great circle. Now, let’s observe that the length of α_1, with respect to the Riemannian metric, is given by $\theta_2 - \theta_1$. This can be computed using the formula for the metric:

$$ds^2 = d\theta^2 + \sin^2\theta \, d\phi^2.$$

To calculate the length of α_1, we note that $d\varphi = 0$ since φ remains constant along this path. Therefore, the length of α_1 can be expressed as:

$$\text{Length of } \alpha_1 = \int_{\theta_1}^{\theta_2} d\theta = \theta_2 - \theta_1.$$

Now, let's determine the length of α_2. This length can be calculated as $2\pi - (\theta_2 - \theta_1)$, which simplifies to:

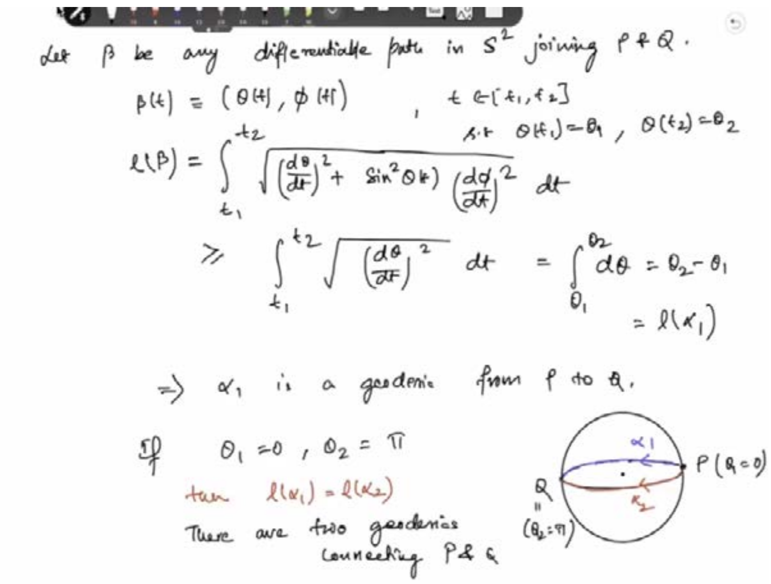
$$\text{Length of } \alpha_2 = 2\pi - \theta_2 - \theta_1.$$

Given that we have chosen θ_1 and θ_2 within the interval $(0, \pi)$ and with $\theta_1 < \theta_2$, it follows that $\theta_2 - \theta_1$ will always be less than $2\pi - \theta_2 - \theta_1$. Therefore, we can conclude that:

$$\text{Length of } \alpha_1 < \text{Length of } \alpha_2.$$

This comparison pertains solely to the two paths α_1 and α_2. Now, let us consider any other path, denoted as β, that connects points P and Q on the sphere S^2.

(Refer Slide Time: 09:23)



Let's consider any differentiable path β on the sphere S^2 that connects points P and Q. We can express this path in terms of spherical coordinates as $\beta(t) = (\theta(t), \phi(t))$, where t varies over some interval. Specifically, let t range from t_1 to t_2, such that $\theta(t_1) = \theta_1$ and $\theta(t_2) = \theta_2$. The length of the path β can be calculated using the following integral:

$$\text{Length of } \beta = \int_{t_1}^{t_2} \sqrt{\left(\frac{d\theta}{dt}\right)^2 + \sin^2(\theta(t)) \left(\frac{d\phi}{dt}\right)^2} \, dt.$$

Now, it's important to note that this length is always greater than or equal to the integral of the first term alone, which simplifies to:

$$\int_{t_1}^{t_2} \left|\frac{d\phi}{dt}\right| \, dt.$$

This latter integral can be rewritten in terms of the angular difference:

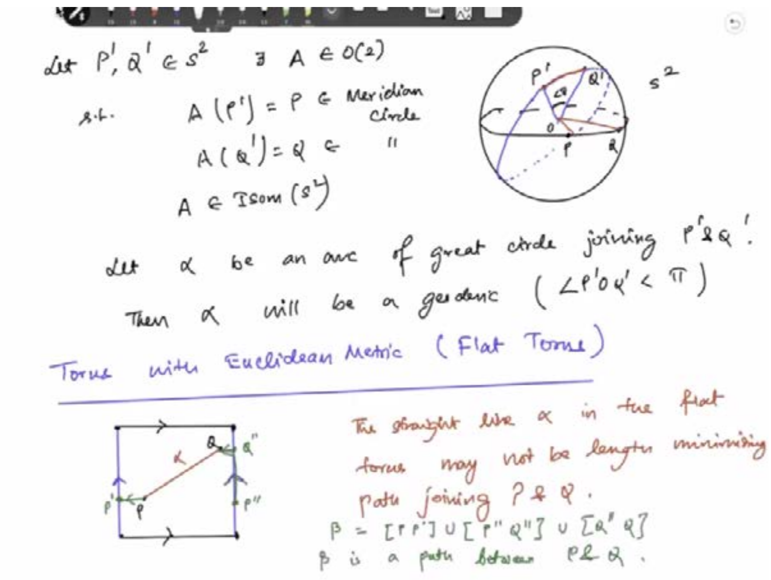
$$\int_{\theta_1}^{\theta_2} \left(\frac{d\theta}{dt}\right)^2 \, dt = \theta_2 - \theta_1,$$

which represents the length of the path α_1.

Thus, we conclude that the length of β is greater than or equal to the length of α_1. This implies that α_1 is indeed a length-minimizing path, qualifying it as a geodesic from point P to point Q.

Now, let's take a closer look at the specific case where we set $\theta_1 = 0$ and $\theta_2 = \pi$. Allow me to illustrate this with a diagram. Suppose point P corresponds to $\theta_1 = 0$ and point Q corresponds to $\theta_2 = \pi$. In this scenario, we have two possible geodesics connecting P and Q: one along α_1 and another which we can denote as α_2.

(Refer Slide Time: 14:34)



In fact, both paths α_1 and α_2 yield equal lengths, demonstrating that there are two distinct geodesics connecting these two points. Therefore, this pair of points P and Q is not representative of a unique geodesic space. Here, we specifically selected P and Q to lie on the meridian circle of the sphere.

Now, let's consider another pair of points, P' and Q', on the sphere S^2. We know that there exists an orthogonal transformation capable of mapping P' to a point P on the meridian circle and Q' to a point Q on the same circle. To clarify, if P' and Q' are points in S^2, then there exists an orthogonal matrix A that belongs to the group O(2) such that:

$$A(P') = P,$$

where P lies on the meridian circle, and

$$A(Q') = Q,$$

where Q also lies on the meridian circle. This transformation is an isometric action, meaning that the orthogonal transformation A acts on S^2 preserving distances. Therefore, the distance between P' and Q' will be the same as the distance between P and Q.

Now, we can visualize a great circle passing through P' and Q'. When we apply the orthogonal transformation A, it transforms the blue great circle into the black meridian circle. Consequently, the arc of this great circle connecting P' and Q' becomes a geodesic. Let's denote this arc as α, which is the segment of the great circle that connects P' and Q'.

We also assume that the angle formed between the points O, P', and O, Q' is less than π. This condition is crucial because if this angle is indeed less than π, then α will qualify as a geodesic. Thus, we have established that in this scenario, the geodesics between these points are segments of a great circle.

Now, let us transition our discussion to the torus.

Now, let's consider a torus equipped with a Euclidean metric, which we refer to as a flat torus. I'll illustrate this flat torus by visualizing it as a square where opposite sides are identified. In this representation, one of the blue vertical sides is connected to the other blue vertical side, creating the characteristic shape of a flat torus.

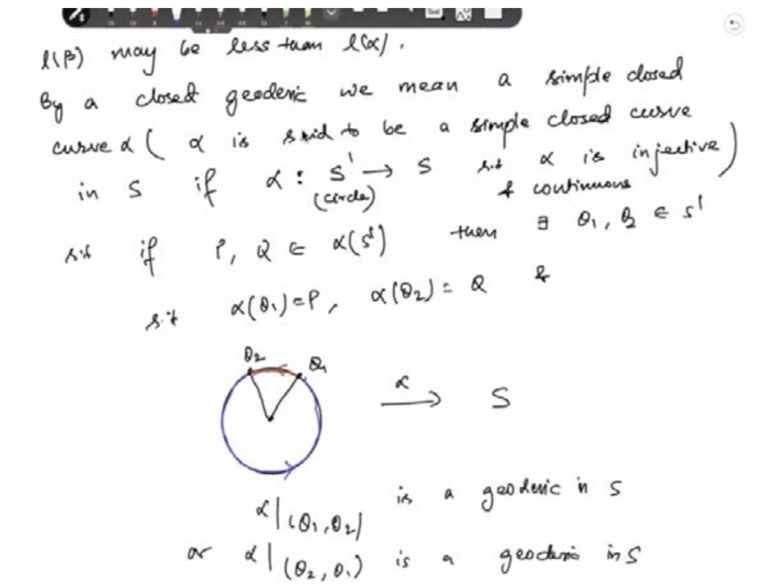
Next, let’s choose two points on the torus, which we will call P and Q. Our objective here is to find a length-minimizing path that connects these two points. One might think that simply joining P and Q with a straight line would suffice. We can denote this straight line as α. However, it’s important to note that this line segment α, while a direct connection, may not actually be the shortest path between P and Q in the context of the flat torus.

Why is this the case? We can explore an alternative path that also connects P and Q through a different route. Consider a horizontal segment, represented here in green. This green horizontal line segment will intersect the left edge of the torus at a point we will call P'. Importantly, P' is identified with another point, which we can denote as P''.

From P'', we can move vertically to a point Q'' that lies directly above Q at the same level. Now, we can connect Q to Q'' with another horizontal segment. Let’s denote this entire path as β, which consists of the horizontal segment from P to P' (where P' is identified with P''), followed by the vertical segment from P'' to Q'', and finally, the horizontal segment connecting Q'' to Q.

So, in summary, β serves as another path between the points P and Q on the flat torus. This exploration demonstrates that the straight line α may not be the geodesic we are looking for, as β could potentially represent a shorter, length-minimizing path.

(Refer Slide Time: 23:50)



It may occur that the length of path β is less than that of path α. Consequently, geodesics on this torus can be a bit complex. However, when it comes to finding closed geodesics, the process becomes much simpler.

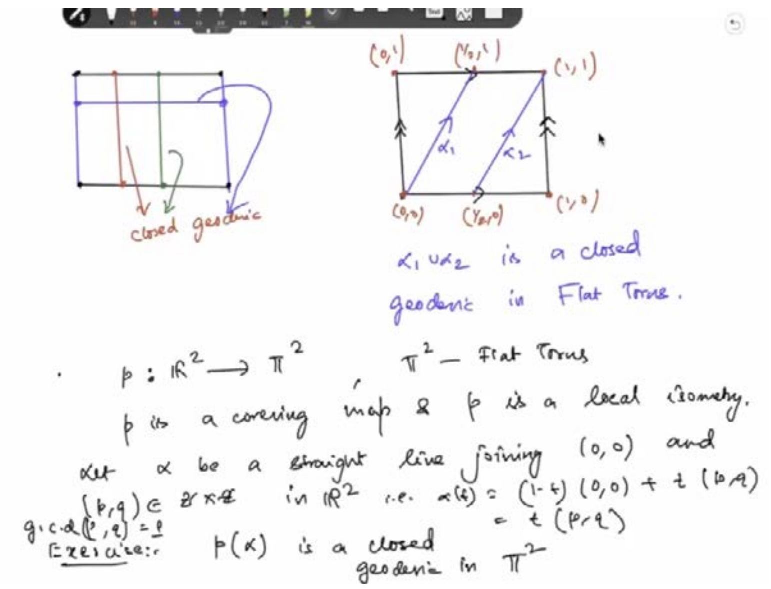
Now, what exactly do we mean by a closed geodesic? A closed geodesic is defined as a simple closed curve. To clarify, let's first define what a simple closed curve is. A curve α is said to be a simple closed curve on a surface S if it can be represented as a map from a circle to the surface. For α to qualify as a simple closed curve, it must be injective, meaning that it does not intersect itself, and it must also be continuous.

When we refer to a closed geodesic, we are talking about a simple closed curve α such that if points P and Q lie on α, then they are part of the same simple closed curve. This means there exist angles θ_1 and θ_2 within the circle S^1 such that $\alpha(\theta_1) = P$ and $\alpha(\theta_2) = Q$.

Now, let's visualize this: we have our points θ_1 and θ_2 on the circle, and this is our mapping α. If we restrict the map α to the arc between θ_1 and θ_2, we have a geodesic in the surface S. This implies that either the portion of the curve from θ_1 to θ_2 is a geodesic in S, or conversely, the segment from θ_2 back to θ_1 is also a geodesic in S.

In essence, this is what we mean by a closed geodesic on a surface. In our context, we are considering the surface S to be a flat torus.

(Refer Slide Time: 29:30)



Here we have our flat torus, and our aim is to identify closed geodesics. To start, let's consider an example: if I take two points and connect them with a vertical line, that line constitutes a closed geodesic. In fact, any vertical line I choose will also form a closed geodesic. Similarly, if I draw a horizontal line, that too will be a closed geodesic.

Now, let's explore another example of a closed geodesic on the torus. Imagine this flat torus where I designate a point as the origin. We can label this point as $(\frac{1}{2}, 0)$. Then we have the points (1, 0), (1, 1), $(\frac{1}{2}, 1)$, and (0, 1) on the torus. If I take a straight line connecting the origin (0, 0) to $(\frac{1}{2}, 1)$, this point $(\frac{1}{2}, 1)$ is identified with $(\frac{1}{2}, 0)$. Next, if I draw a straight line from $(\frac{1}{2}, 0)$ to (1, 1), the two blue parallel lines formed will indeed create a closed geodesic on the flat torus.

In fact, we can completely characterize the closed geodesics on a flat torus. How is this achieved? We start with a covering map $P: R^2 \rightarrow \text{flat torus}$, which acts as both a covering map and a local isometry. If I draw a straight line through the origin and extend it to some point (P, Q) where both P and Q are integer points, projecting this path down onto the flat torus will result in a closed geodesic.

Let's denote this straight line as α, which connects the origin to the point (P, Q) in R^2. In mathematical terms, we can express α(t) as a convex combination:

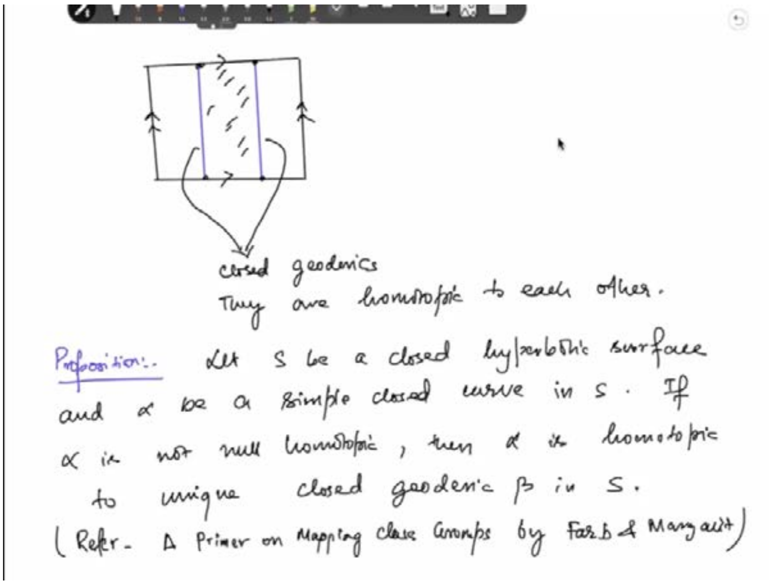
$$\alpha(t) = (1 - t)(0,0) + t(P, Q)$$

Now, it can be proven as an exercise that P(α) defines a closed geodesic in the torus T^2. However, there's an important condition we need to consider: P and Q must be coprime, meaning their greatest common divisor (GCD) must equal 1, i.e., $\text{GCD}(P, Q) = 1$.

Why is this assumption necessary? Let's illustrate with an example: suppose P and Q are both equal to 2. If I draw a straight line from the origin (0, 0) to (2, 2), it will pass through the point (1, 1). When we project this down onto the torus, the line will wind around the torus twice. The path from (0, 0) to (1, 1) generates a simple closed curve, and from (1, 1) to (2, 2), we would encounter the same simple closed curve again.

Thus, this winding leads us to a non-injective mapping, demonstrating that P(α) cannot be a unique geodesic. Therefore, it is essential to ensure that the GCD of P and Q equals 1 to maintain the integrity of our closed geodesic.

(Refer Slide Time: 37:16)

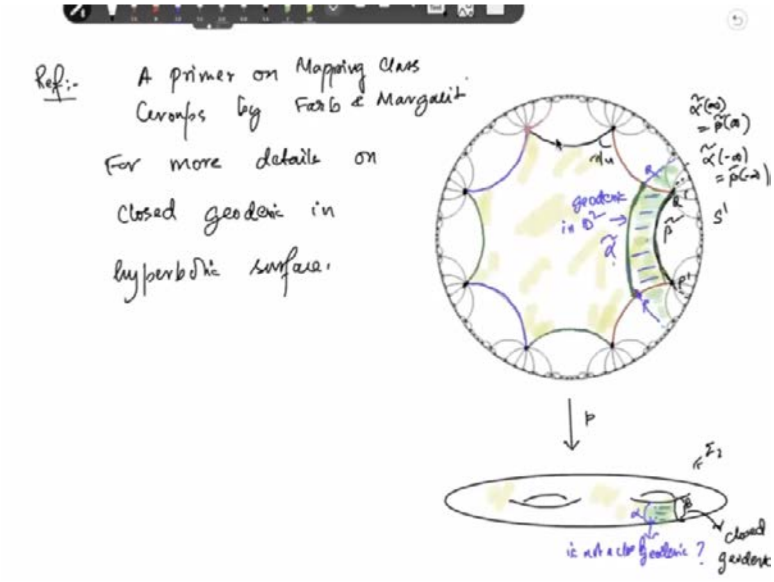


Let's take a closer look at some fascinating properties of the torus. Here is our torus, where the opposite sides are identified. If we consider two distinct vertical closed geodesics, both of these paths are indeed closed geodesics. Remarkably, these two geodesics are homotopic to each other, which means they can be continuously deformed into one another without leaving the surface of the torus. Therefore, if we take a closed geodesic within the torus, we can find another geodesic within its homotopy class.

Now, if we shift our focus to hyperbolic surfaces, we can certainly create closed geodesics as well. However, an interesting distinction arises in the homotopy class of a closed geodesic on a hyperbolic surface: we will only find that specific geodesic itself. Allow me to formalize this observation with a proposition: let S be a closed hyperbolic surface, and let α be a simple closed curve in S. If α is null homotopic, then it is homotopic to a unique closed geodesic β in S.

While I won't delve into the proof of this proposition here, I highly recommend consulting the book A Primer on Mapping Class Group by Farb and Margalit for a thorough explanation. Instead, I will present some examples of closed geodesics on hyperbolic surfaces and offer a conceptual framework for understanding how to prove this proposition.

(Refer Slide Time: 41:56)



Let’s revisit how we construct a hyperbolic surface, particularly a genus 2 surface that admits a hyperbolic metric. This process begins with an 8,8 tiling of the unit disc model of the hyperbolic plane. In this model, we have a regular octagon, specifically a geodesic octagon, where the side pairings are determined by isometries of the hyperbolic plane.

In this geodesic octagon, the red geodesic corresponds to another red geodesic, the green geodesic is paired with a green one, the blue geodesic with a blue one, and similarly for the black geodesic. To form a closed curve or geodesic in this hyperbolic surface, we first select a point. This point will be identified with another point within the octagon, and we can connect them with a geodesic.

Now, the two points, denoted as P and Q, will be identified with a single point on this surface, denoted as Σ_2. Thus, in the hyperbolic plane, this indeed represents a geodesic. The real question arises when we project this down; we need to determine whether it forms a simple closed curve on the surface Σ_2.

It is essential to note that this may not always be the case. For instance, the blue geodesic is clearly homotopic to the black geodesic, indicating that they are both geodesics in the unit disc D^2. If we project down this relationship, we label the projection as P'Q'. Consequently, this projection yields a black simple closed curve on the surface.

Thus, we observe that the blue simple closed curve is homotopic to the black simple closed curve. To clarify our notation, let's name the blue curve α and its projection $\tilde{\alpha}$, while the black curve will be referred to as β and its projection as $\tilde{\beta}$. Hence, the closed curve α is homotopic to the closed curve β.

By employing the homotopy lifting property, we conclude that the projections $\tilde{\alpha}$ and $\tilde{\beta}$ are also homotopic to each other. Furthermore, this extension of $\tilde{\alpha}$ can be carried out in other directions across the tiling.

As a result, you will obtain infinitely many geodesics, and a similar process applies to the curve $\tilde{\beta}$. Thus, the curves $\tilde{\alpha}$ and $\tilde{\beta}$ can be extended into infinitely long geodesics, and when we project them down, we yield two curves, α and β, which are homotopic to each other.

Next, consider lifting this region up; this process gives rise to a green region in one tile, as well as in the adjacent tile. It turns out that the curves $\tilde{\alpha}$ and $\tilde{\beta}$ lie within a uniformly bounded neighborhood of each other, demonstrating that both are infinite geodesics. On the boundary, which is represented as the circle S^1, these curves will intersect at a single point. Thus, the endpoints of $\tilde{\alpha}$ and $\tilde{\beta}$ will coincide, meaning that $\widetilde{\alpha_\infty}$ is identical to $\widetilde{\beta_\infty}$, and $\widetilde{\alpha_{-\infty}}$ is the same as $\widetilde{\beta_{-\infty}}$.

Given that their endpoints coincide and considering that D^2 is a unique geodesic space, it follows that $\tilde{\alpha}$ and $\tilde{\beta}$ should also coincide. However, that is not the case here. Referring to the illustration, we see that the length of the segment of $\tilde{\alpha}$ is greater than the length of the segment of $\tilde{\beta}$.

When we project this down onto the surface, it becomes evident that α does not represent a closed geodesic, whereas β does qualify as a closed geodesic. This distinction underscores the proposition that if we take any curve α to be a simple closed curve that is not null homotopic, then within its homotopic class, there exists a unique closed geodesic β in the surface S.

It is worth noting that the depiction of closed geodesics can be quite intricate. For those interested in delving deeper into the topic of closed geodesics on hyperbolic surfaces, I recommend consulting the book A Primer on Mapping Class Groups by Farb and Margalit. I will conclude my discussion here.