The Proof of Markov's Theorems

Sebastian Vögele

April 2020

In this lecture we introduce ideal arcs and prove a correspondence between them and simple closed geodesics. We then use this to give a geometric proof of both versions of Markov's theorem. We follow closely the proof laid out in [1, chapters 14 and 15].

1 Ideal arcs

Definition 1.1 *An ideal arc in a punctured hyperbolic torus is a simple geodesic with endpoints in the cusp.*

The following lemmas should be more or less clear from our previous discussions, so they are given without proof.

Lemma 1.2 *The midpoint of an ideal arc is well-defined, that is, for any choice of a horocycle at the cusp, (which gives a horocycle at each end of the ideal arc,) the midpoint of the ideal arc is the same. This means in particular that bisecting an ideal arc is a well-defined condition.*

Lemma 1.3 *Every edge in a ideal triangulation is an ideal arc. Conversely, every ideal arc is an edge of an ideal triangulation. Moreover, all ideal triangulations containing a fixed ideal arc are related to each other by flips of the other edges.*

Next, we come to the main result of this section.

Theorem 1.4 *For every ideal arc c, there is a unique simple closed geodesic g that does not intersect c. Conversely, for every simple closed geodesic g on a punctured hyperbolic torus, there is a unique ideal arc c that does not intersect g. If c is the edge of an ideal triangulation, g bisects both other edges.*

Proof. Since we didn't introduce the notions from Riemannian geometry that we would need to prove this theorem rigorously, this proof only aims to give an intuitive understanding of how it can be shown.

Let c be an ideal arc and cut the punctured torus along c to get a hyperbolic cylinder with a cusp on each component of the boundary (see figure 1). The simple closed geodesics on this cylinder correspond exactly to the simple closed geodesics on the punctured torus that don't intersect c. Due to the negative curvature, there is a unique simple closed geodesic on the cylinder.

For the converse, let g be a simple closed geodesic on the punctured hyperbolic torus and cut the torus along g to get a hyperbolic cylinder with a cusp in the interior (see figure 1). Similarly to before, the ideal arcs that don't intersect g correspond to simple geodesics on the cylinder with endpoints in the cusp. The negative curvature can be used again to show that there is a unique such geodesic.

If c is the edge of an ideal triangulation, rotating around the midpoint of c by 180° leaves the triangulation invariant. The geodesic segments in each triangle of the triangulation that bisect the other edges are sent to each other under this rotation, so they connect smoothly. This means, their union is a closed geodesic that doesn't intersect c. By uniqueness of g, this geodesic is g. □

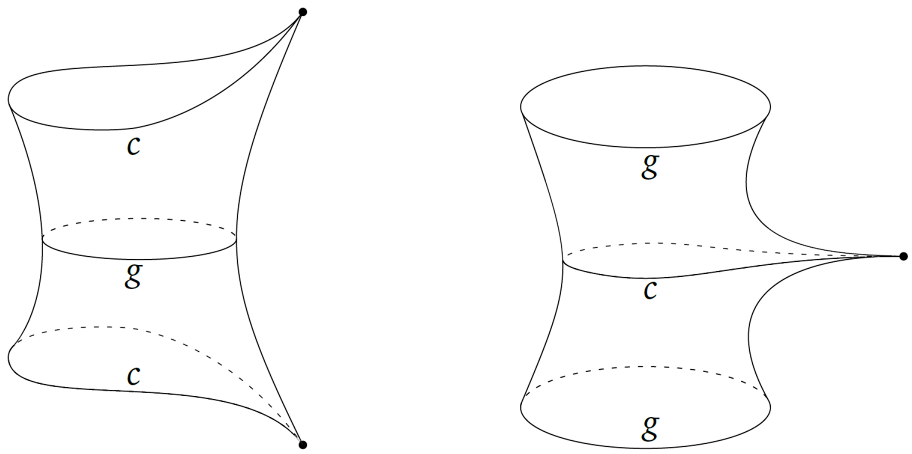

Figure 1: The cylinders received by cutting the punctured torus along c (left) respectively g (right). Note that the cusps are actually at infinity. Source: [1, figure 15]

2 The quadratic forms version

We first recall quadratic forms version of Markov's theorem.

Theorem 2.1 *Let (a, b, c) be a Markov triple, $p_1, p_2 \in \mathbb{Z}$ such that $p_2 b - p_1 a = c$, $x_0 = \frac{p_2}{a} + \frac{b}{ac} - \frac{3}{2}$, $r = \sqrt{\frac{9}{4} - \frac{1}{c^2}}$ and $f(p, q) = p^2 - 2x_0 pq + (x_0^2 - r^2)q^2$, then*

$$M(f) := \inf_{(p,q)\in\mathbb{Z}^2\setminus\{0\}} \frac{|f(p, q)|}{\sqrt{|\det f|}} = \frac{1}{r}$$

and the infimum is attained.

Conversely, if f is an indefinite binary quadratic form with $M(f) > \frac{2}{3}$, then it is a multiple of a form received by this construction.

We have already showed using the correspondence between indefinite quadratic forms and geodesics that this theorem is equivalent to the following theorem.

Theorem 2.2 *In the notation of 2.1, the geodesic g on the hyperbolic plane with endpoints $x_0 \pm r$ satisfies*

$$M(g) := \inf_{h \text{ ford circle}} d(g, h) = -\log r$$

and the infimum is attained.

Conversely, if g is a geodesic on the hyperbolic plane such that

$$M(g) > -\log\frac{3}{2}$$

then there is a Markov triple (a, b, c) such that g is $GL_2(\mathbb{Z})$-related to the geodesic with endpoints $x_0 \pm r$.

This theorem in turn is a direct corollary of the following proposition.

Proposition 2.3 *Let g be a complete geodesic on the hyperbolic plane and $\pi : \mathbb{H} \to \mathbb{T}$ the projection to the modular torus, then the following are equivalent:*

1. *$\pi(g)$ is a simple closed geodesic.*
2. *There is a Markov triple (a, b, c) such that in the notation of 2.1, g is $GL_2(\mathbb{Z})$-related to the geodesic with endpoints $x_0 \pm r$.*
3. *$M(g) > -\log\frac{3}{2}$*

Moreover, if these conditions are satisfied, then

$$M(g) = -\log r$$

and the infimum is attained.

Proof. "1. $\Rightarrow$ 2.": Let c be the ideal arc that doesn't intersect $\pi(g)$ and a, b ideal arcs such that a, b and c form a triangulation of the modular torus. We refer to the ideal arcs as well as their weights with a, b and c. Last lecture we saw that the weights (a, b, c) of this triangulation are a Markov triple.

All lifts of triangles in this triangulation to the hyperbolic plane are $GL_2(\mathbb{Z})$-related to each other and g bisects a lift of the two edges a and b (actually, g bisects infinitely many such lifts). Therefore we only need to show that the geodesic with endpoints $x_0 \pm r$ bisects the two edges a and b of one of the lifted triangles to see that these geodesics are $GL_2(\mathbb{Z})$-related.

Consider the decorated ideal triangle on $\mathbb{H}$ with vertices

$$\nu_1 = \frac{p_1}{b},\ \nu_2 = \frac{p_2}{a},\ \nu_3 = \infty$$

and Ford circles $h(p_1, b)$, $h(p_2, a)$ and $h(1, 0)$ as horocycle at the endpoints. Using the formulas for the weights of this triangle, we see that they are exactly a, b and c. Moreover, since $\nu_1 < \nu_2$ and $\nu_3 = \infty$, we can use a proposition form last time, which shows that the geodesic bisecting the edges corresponding to a and b has endpoints $x_0 \pm r$.

"2. $\Rightarrow$ 3.": We show that $M(g) = -\log r$ and that the infimum is attained in this step. Since $r < \frac{3}{2}$, this also shows 3.

Consider the triangulation of the modular torus corresponding to the Markov triple (a, b, c) as in the last step. g bisects infinitely many of the lifts of the edges a and b to the hyperbolic plane, so by a proposition from last time the signed distance from g to the Ford circles at the vertices of the corresponding ideal triangles is $-\log r$.

We claim that every other Ford circle has a larger signed distance to g. Indeed, if we take any other Ford circle, there is another lift of $\pi(g)$ that bisects the edges a and b of a lift of a triangle of the triangulation incident to the Ford circle and is therefore closer to the Ford circle than g and already at signed distance $-\log r$ from it.

"$\neg$1. $\Rightarrow$ $\neg$3.": If g is not a simple closed geodesic, we want to find a sequence of Ford circles, such that their distance to g approaches or is lower than $-\log \frac{3}{2}$.

If one of the endpoints of g is in $\mathbb{Q} \cup \{\infty\}$, then the distance of g to the ford circle at this point is $-\infty$, so $M(g) = -\infty$.

If both endpoints of g are irrational, we want to find a sequence of sorted Markov triples (a_n, b_n, c_n), such that $\pi(g)$ intersects the edges corresponding to a_n and b_n of the corresponding triangulation T_n of the modular torus at least once without intersecting c_n in between and such that $c_n \to \infty$ as $n \to \infty$. By some results we have seen in the last lecture, the distance from g to one of the Ford circles at the vertices then has to be less than $-\log \sqrt{\frac{9}{4} - \frac{1}{c_n^2}}$, which approaches $-\log \frac{3}{2}$.

We construct (a_n, b_n, c_n) and T_n inductively. $(a_0, b_0, c_0) = (1, 1, 1)$ and T_0 is chosen so that $\pi(g)$ crosses the edges a_0 and b_0 as required. This can be done because all edges have the same weight and $\pi(g)$ does not end at a vertex since the endpoints of g are irrational.

Given the triangulation T_n, we construct T_{n+1} as follows. Since $\pi(g)$ is not simply closed and has irrational endpoints, it intersects all edges of the triangulation infinitely often. Therefore we can find an intersection with the edge c_n immediately after an uninterrupted intersection with the edges a_n and b_n. If the crossings appear in the order $a_n b_n c_n$, then let T_{n+1} be the triangulation obtained from flipping the edge b_n and let b'_n be the flipped edge. The corresponding sorted Markov triple is

$$(a_{n+1}, b_{n+1}, c_{n+1}) = (a_n, c_n, b'_n)$$

which satisfies the requirements. If the crossings appear in the order $b_n a_n c_n$, we reverse the roles of a_n and b_n and get an analogous triangulation and corresponding Markov triple. □

3 The Diophantine approximation version

Again, we recall the Diophantine approximation version of Markov's theorem.

Theorem 3.1 *Let (a, b, c) be a Markov triple, $p_1, p_2 \in \mathbb{Z}$ such that $p_2 b - p_1 a = c$, $x_0 = \frac{p_2}{a} + \frac{b}{ac} - \frac{3}{2}$, $r = \sqrt{\frac{9}{4} - \frac{1}{c^2}}$ and $x = x_0 + r$, then the Lagrange number of x satisfies*

$$L(x) := \sup\{\lambda \in \mathbb{R} \,|\, \text{there are infinitely many } \frac{p}{q} \in \mathbb{Q} \text{ such that } |x - \frac{p}{q}| < \frac{1}{\lambda q^2}\} = 2r$$

and the supremum is attained.

Conversely, if $x \in \mathbb{R} \setminus \mathbb{Q}$ with $L(x) < 3$, then x is $GL_2(\mathbb{Z})$-related to a number received by this construction.

We have already seen using the correspondence between geodesics and Diophantine approximations that this theorem is equivalent to the next one.

Theorem 3.2 *In the notation of 3.1, the geodesic g on the hyperbolic plane with endpoints x and ∞ satisfies*

$$L(g) := \inf\{M \in \mathbb{R} \,|\, \text{there are infinitely many ford circles } h \text{ such that } d(h, g) < M\} = -\log r$$

and the infimum is attained.

Conversely, if g is a complete geodesic on the hyperbolic plane with

$$L(g) > -\log \frac{3}{2}$$

then each endpoint x of g is rational, ∞ or there is a Markov triple (a, b, c) such that x is $GL_2(\mathbb{Z})$-related to $x_0 + r$.

This theorem follows directly from 2.3 and the following proposition.

Proposition 3.3 *Let g be a complete geodesic on the hyperbolic plane and $X \subset \mathbb{R} \setminus \mathbb{Q}$ the set of the endpoints of complete geodesics that project to simple closed geodesics on the modular torus, then the following are equivalent:*

1. *The endpoints of g are contained in $\mathbb{Q} \cup \{\infty\} \cup X$.*
2. *$L(g) > -\log \frac{3}{2}$*

Moreover, if these conditions are satisfied and at least one endpoint of g is contained in X, then $L(g)$ is the minimum of $M(\tilde{g})$ for $\tilde{g}$ a geodesic that has a common endpoint with g and projects to a simple closed geodesic on the modular torus. In this case the infimum in $L(g)$ is attained.

Proof. "1. $\Rightarrow$ 2.": In this step we show the additional statement. Since by 2.3 $M(\tilde{g}) > -\log \frac{3}{2}$, 2. follows for the case where at least one endpoint of g is in X. In the other case $L(g) \geq 0$ by the explanation below.

We partition g into three segments: A segment at each endpoint and a finitely long remaining part between them. Due to its finite length, this last part can only intersect finitely many Ford circles. If an endpoint of g is in $\mathbb{Q} \cup \{\infty\}$, we choose the segment of g at this endpoint to be the part contained in the Ford circle at the endpoint. This segment then doesn't intersect any ford circle.

If an endpoint of g is in X, let $\tilde{g}$ be the geodesic with the same endpoint that projects to a simple closed geodesic on the modular torus. All Ford circles have a signed distance of at least $M(\tilde{g})$ to $\tilde{g}$. If we choose the segment of g at this endpoint to be small enough, it is arbitrarily close to $\tilde{g}$ and therefore all Ford circles intersecting it have a distance of at least $M(\tilde{g}) - \varepsilon$ for an arbitrary $\varepsilon > 0$.

Putting these estimate together (recall that if g does not intersect a horocycle, their signed distance is positive), we get that for any $\varepsilon > 0$ there are only finitely many Ford circles with distance less than $M(\tilde{g}) - \varepsilon$ from g, so $L(g) \geq M(\tilde{g})$. It remains to show that there are infinitely many Ford circles with distance $M(\tilde{g})$ from g.

In the proof of 2.3 we showed that there are infinitely many Ford circles with distance $M(\tilde{g})$ to $\tilde{g}$. Since g approaches $\tilde{g}$ near the endpoint, they intersect the same edges of the triangles of the corresponding triangulation infinitely often. By a proposition from last lecture the signed distance from g to at least one of the Ford circles in such a triangle must be less than $M(\tilde{g})$, so there are infinitely many with distance less than $M(\tilde{g})$ from g.

"$\neg 1. \Rightarrow \neg 2.$": If g has an irrational endpoint that is not the endpoint of a simple closed geodesic, we can do the same construction as in the proof of 2.3 ("$\neg 1. \Rightarrow \neg 3.$") to get Ford circles with signed distance less than $-\log\sqrt{\frac{9}{4}-\frac{1}{c_n^2}}$ from g for any $n \in \mathbb{N}$, where $c_n \to \infty$ as $n \to \infty$. This gives of infinitely many Ford circles with signed distance less than $-\log\frac{3}{2}$ to g, so $L(g) \leq -\log\frac{3}{2}$. $\square$

References

[1] B. Springborn. The hyperbolic geometry of markov's theorem on diophantine approximation and quadratic forms. *L'Enseignement Mathématique*, 63(2):333–373, 2017.