What is a Markoff Number?

Michael Lane

July 2, 2019

Abstract

The continued fraction expansion of a real number $\theta > 0$ yields a sequence of rational numbers $\{p_n/q_n\}$, called the convergents of the expansion, which are the "best rational approximations" of θ. In particular, p_n/q_n is closer to θ than any other rational with smaller (or equal) denominator. A Markoff number gives an explicit bound on the error $|\theta - p_n/q_n|$. We will calculate these numbers and illustrate their surprising connection to geodesics in the hyperbolic plane.

References

[1] T. Bedford, M. Keane, and C. Series. *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces.* Oxford University Press, Oxford.

[2] A. Markoff. *Sur les formes binaires indefinies, I.* Math. Ann., 15:281-309, 1879.

[3] C. Series. *The Geometry of Markoff Numbers.* Math. Intell., 7(3):20-29, 1985.

What is a Markoff Number?

Michael Lane

July 2, 2019

Outline

- Continued Fractions
- Basics of Hyperbolic Geometry
- Cutting and Characteristic Sequences
- Markoff’s Quadratic Form
- Diophantine Equations

Continued Fractions

- Fix $\theta > 0$. A *continued fraction* of θ is an expression of the form

$$\theta = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}}$$

  where the $a_0, a_1, a_2, \ldots$ are positive integers.
- This can be denoted more concisely as

$$\theta = [a_0, a_1, a_2, \ldots].$$

- θ is rational if and only if its continued fraction terminates.
- Continued fractions are unique up to the ambiguity $[\ldots, 2] = [\ldots, 1, 1]$.

Continued Fractions

- A *convergent* of the continued fraction for θ is a rational number ${}^{p_n}/_{q_n}$, where

$$\frac{p_n}{q_n} = [a_0, a_1, \dots, a_n]$$

and p_n, q_n are coprime and positive for all *n.*

Rational Approximations

- The convergent ${}^{p_n}/_{q_n}$ is a *best rational approximation* for θ in the sense that it is closer to θ than every other rational number with denominator smaller than (or equal to) q_n.
- This statement can be strengthened significantly.
- In fact, there exists c, depending on θ but not on *n*, such that

$$\left|\theta - \frac{p_n}{q_n}\right| < \frac{c}{q_n^2}$$

for infinitely many *n*.

Rational Approximations

- The number *c* can always be taken as $1/\sqrt{5}$.
- This bound is tight for the Golden Ratio

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \ddots}} = \frac{1+\sqrt{5}}{2} \approx 1.61803$$

- We can sometimes take *c* to be smaller. Indeed, define

$$\nu(\theta) = \inf\left\{c \colon \left|\theta - \frac{p}{q}\right| < \frac{c}{q^2} \text{ for infinitely many } q\right\}$$

Markoff Numbers

- The *Markoff numbers* are a sequence of reals $\{\nu_i\} \subset \left(\frac{1}{3}, \frac{1}{\sqrt{5}}\right]$, decreasing to 1/3, such that if $\theta > 0$ satisfies $\nu(\theta) > 1/3$, then $\nu(\theta) = \nu_i$ for some *i.*
- The elements of $\nu^{-1}(\nu_i)$ are called the *Markoff irrationalities* for ν_i.

Hyperbolic Geometry

- Hyperbolic geometry has methods for measuring distance and area in the complex plane, different from normal Euclidean geometry.
- Notably, angles are measured in the same manner in both geometries.
- Let $\mathbb{H} \subset \mathbb{C}$ denote the upper half plane, $\mathbb{H} = \{z \in \mathbb{C}: \operatorname{Im} z > 0\}$.
- The length of a smooth curve γ in $\mathbb{H}$ is given by

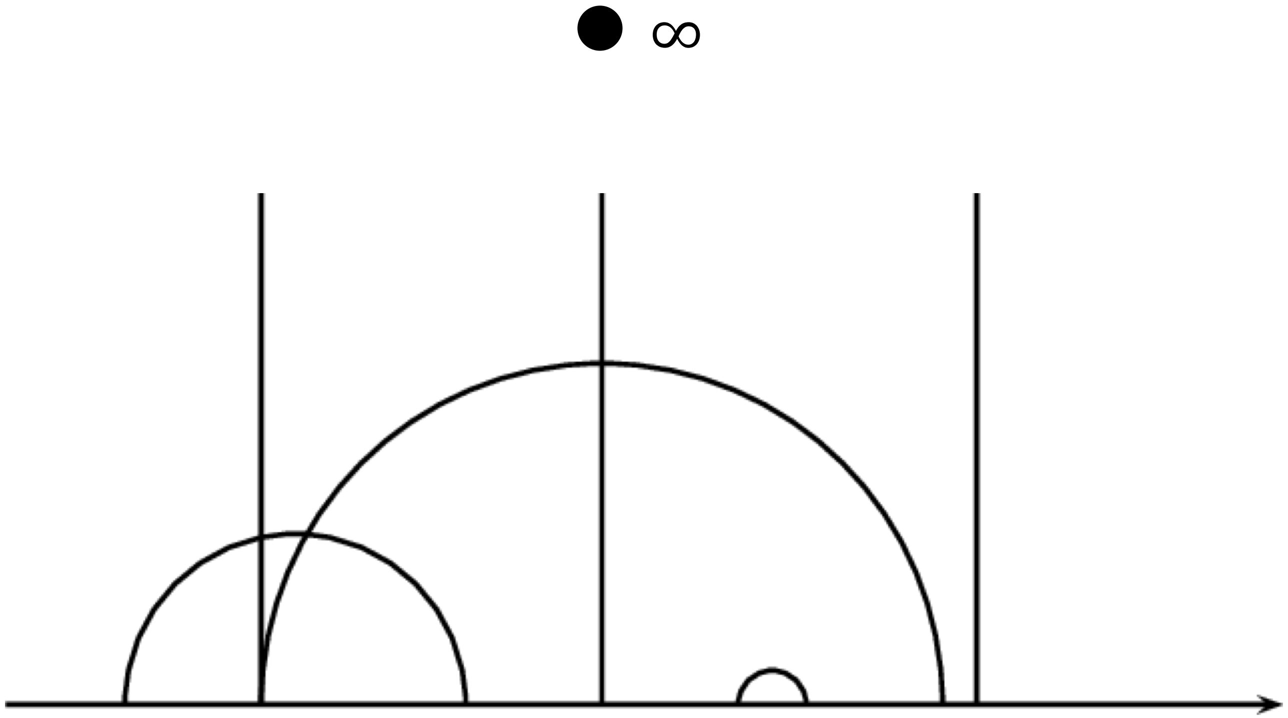
$$\int_\gamma \frac{|dz|}{\operatorname{Im} z} = \int_{\mathbb{R}} \frac{|\gamma'(t)|\, dt}{\operatorname{Im} \gamma(t)}$$

- We can define a metric ρ on $\mathbb{H}$ that agrees with this definition of length.

Hyperbolic Geometry

- Any curve between two fixed points of minimal (hyperbolic) length is called a *geodesic*.
- Geodesics are arcs of semicircles that meet the boundary $\partial\mathbb{H}$ at right angles.
    - (Lines are also circles.)
- The boundary $\partial\mathbb{H}$ includes the real axis and the point at infinity. Vertical rays meet ∞ at right angles.
    - Think of $\partial\mathbb{H}$ also as a circle, because it is one.

$$\int_\gamma \frac{|dz|}{\operatorname{Im} z} = \int_\mathbb{R} \frac{|\gamma'(t)|\, dt}{\operatorname{Im} \gamma(t)}$$



Möbius transformations

- The *Möbius* or *fractional linear transformations* are bijections from $\mathbb{H}$ onto $\mathbb{H}$ of the form

$$M(z) = \frac{az + b}{cz + d}$$

  where $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$.
- Möbius transformations are all *conformal automorphisms*, which preserve angles in $\mathbb{H}$.
- Möbius transformations send circles to circles.
- Mobius transformations *act on* $\mathbb{H}$ *isometrically;* that is, for all *M* and $z, w \in \mathbb{H}$

$$\rho(z, w) = \rho\big(M(z), M(w)\big).$$

$\mathrm{SL}(2, \mathbb{R})$

- Möbius transformations form a group $\mathcal{M}$.
- $\mathcal{M}$ is nearly isomorphic to the matrix group

$$\mathrm{SL}(2, \mathbb{R}) := \left\{ M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : M \text{ real and } \det M = 1 \right\}.$$

by the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \frac{az + b}{cz + d}.$$

- The kernel of this surjection is merely $\pm I$.

$\mathrm{SL}(2, \mathbb{Z})$

- Define the subgroup of $\mathrm{SL}(2, \mathbb{R})$

$$\mathrm{SL}(2, \mathbb{Z}) := \left\{ M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : M \text{ integral and } \det M = 1 \right\}.$$

- Like $\mathrm{SL}(2, \mathbb{R})$, the group $\mathrm{SL}(2, \mathbb{Z})$ acts isometrically on $\mathbb{H}$ by way of the corresponding Möbius transformation.

$\mathrm{SL}(2,\mathbb{Z})$ and Continued Fractions

- Now that we have a little hyperbolic geometry under our belts, let us return to continued fractions.
- The matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

  generate $\mathrm{SL}(2,\mathbb{Z})$.
  - The proof is similar to the Euclidean algorithm for calculating GCDs, which is itself similar to the standard algorithm for calculating continued fractions.
- These matrices correspond to the Möbius transformations

$$P(z) = z + 1 \quad \text{and} \quad Q(z) = -\frac{1}{z}.$$

$\mathrm{SL}(2, \mathbb{Z})$ and Continued Fractions

$$P(z) = z + 1 \quad \text{and} \quad Q(z) = -\frac{1}{z}$$

- Let us examine the effect of these generators on continued fraction expansions.

$$P^{-a_0}[a_0, a_1, a_2, \dots] = [a_1, a_2, \dots]^{-1}$$

$$QP^{-a_0}[a_0, a_1, a_2, \dots] = -[a_1, a_2, \dots]$$

$$P^{a_1}QP^{-a_0}[a_0, a_1, a_2, \dots] = -[a_2, a_3, \dots]^{-1}$$

- Allowing some liberties, a continued fraction is then equivalent to the infinite word

$$P^{a_0}QP^{-a_1}QP^{a_2} \cdots \sim [a_0, a_1, a_2, \dots]$$

$\mathrm{SL}(2,\mathbb{Z})$ and Continued Fractions

- We have just proved the following theorem:

The tails of the continued fractions of θ_1 and θ_2 coincide if and only if there exists $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $M(\theta_1) = \theta_2$.

That is to say, if $\theta_1 = [a_0, a_1, \dots]$ and $\theta_2 = [b_0, b_1, \dots]$, then there exist $r, N \in \mathbb{N}$ such that $a_{n+r} = b_n$ for all $n \geq N$ if and only if there exists $M \in \mathrm{SL}(2,\mathbb{Z})$ such that $M(\theta_1) = \theta_2$.

Cutting Sequences

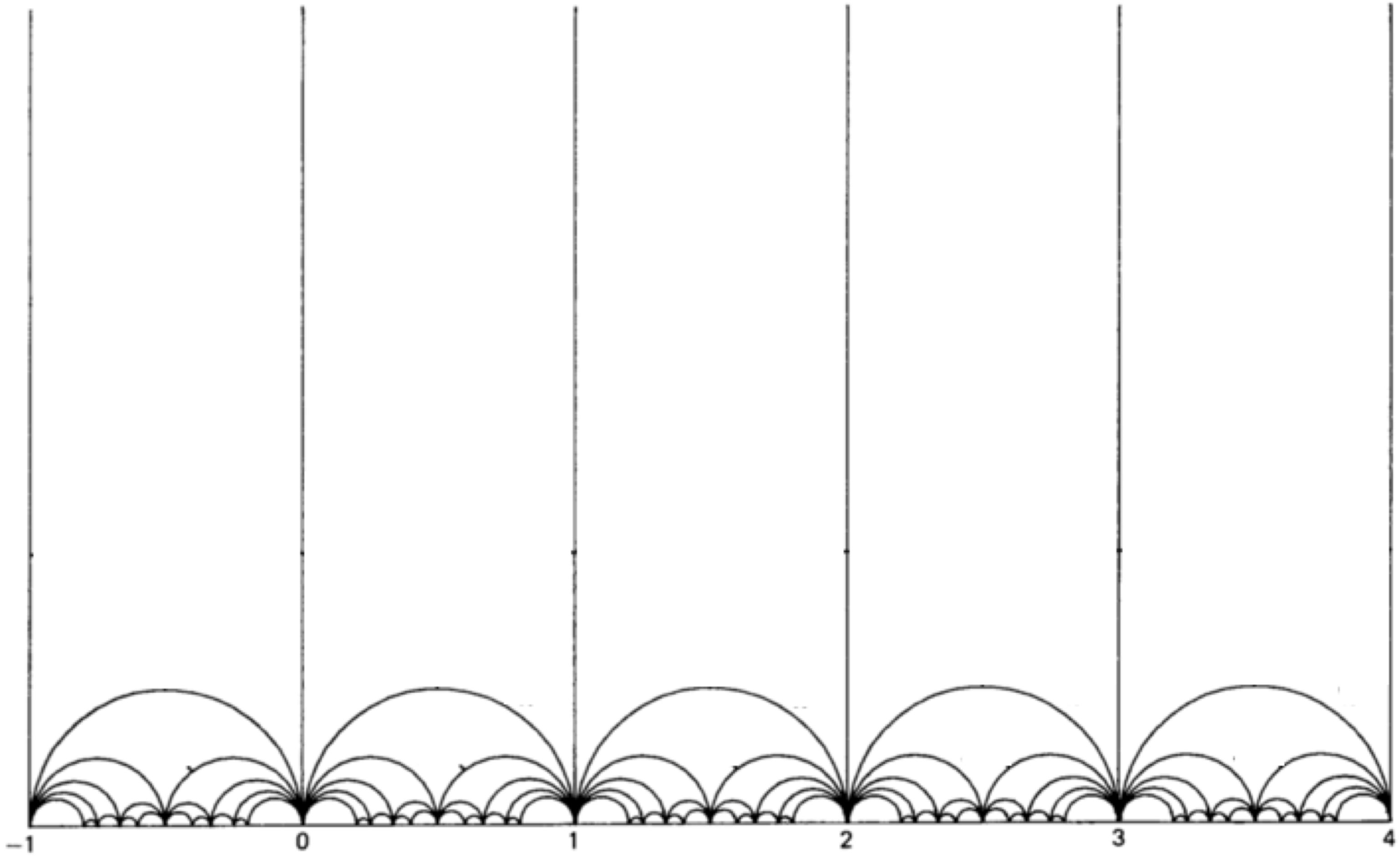
- We wish to formalize this philosophy of associating continued fractions with infinite words of generators of $\mathrm{SL}(2, \mathbb{Z})$.
- This formalization is achieved by way of *cutting sequences.*
- In the interest of time, we must brush many details under the rug.

The $\Gamma(2)$ Lattice

- $\text{I will not define } \Gamma(2)$, because its definition is not important.
- $\Gamma(2)$ is some finite-index subgroup of $\mathrm{SL}(2, \mathbb{Z})$

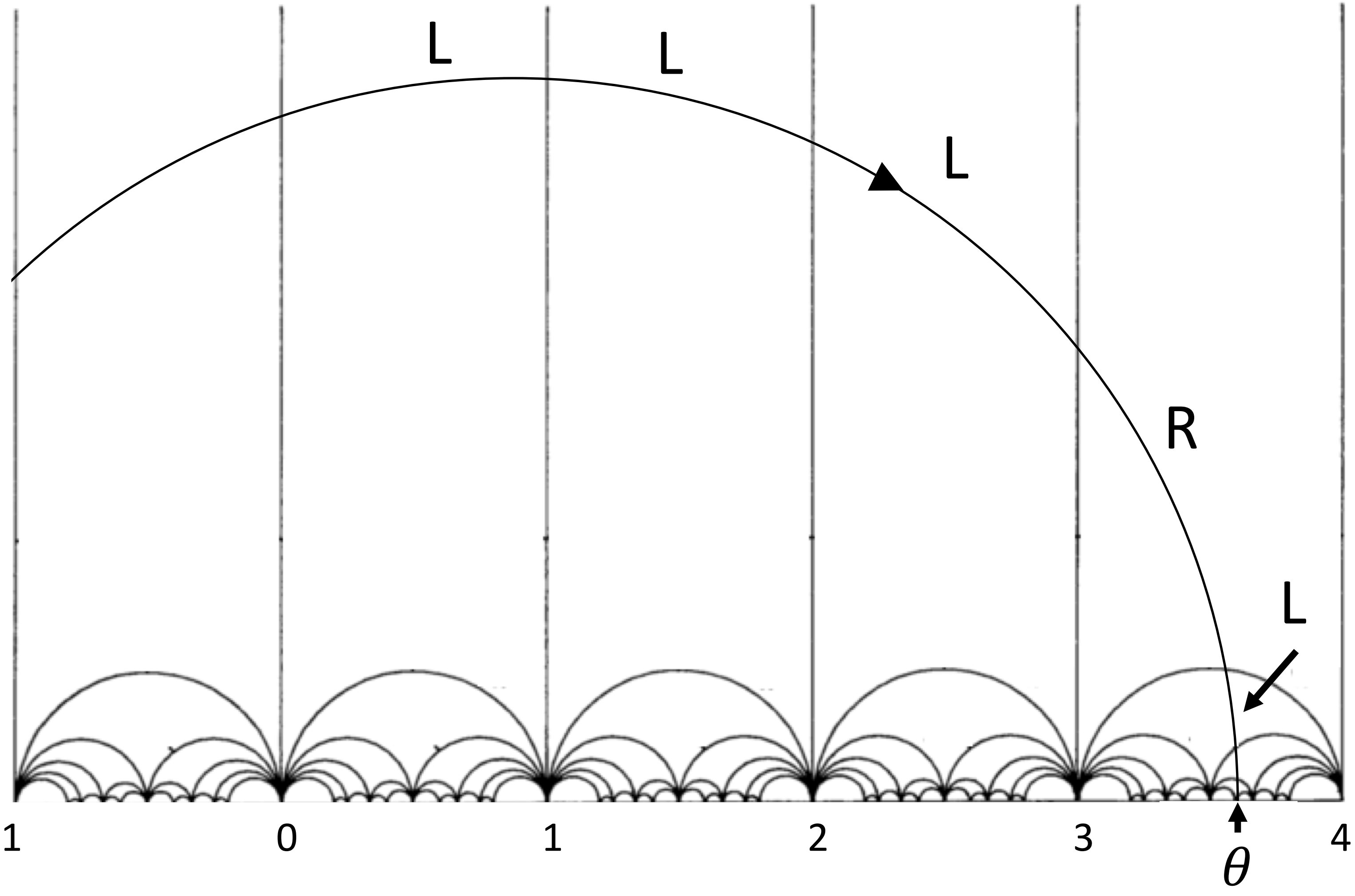
The $\Gamma(2)$ Lattice

- Do not confuse with a Euclidean lattice!
- $\mathbb{H}$ can be partitioned into non-overlapping triangles with vertices on $\partial\mathbb{H}$.
- The triangles' boundaries are geodesics.
- $\Gamma(2)$ acts transitively and freely on the triangles.
- These are the only salient features of $\Gamma(2)$.
- The continued fraction of $\theta > 0$ can be calculated by recording how the corresponding geodesic intersects these triangles.



The $\Gamma(2)$ Cutting Sequence

- Fix $\theta > 0$.
- Plot a geodesic.
- Record the direction of the "odd vertex out" starting from $i\mathbb{R}$
- We get a cutting sequence $L^{a_0}R^{a_1}L^{a_2}\dots$



The $\Gamma(2)$ Cutting Sequence

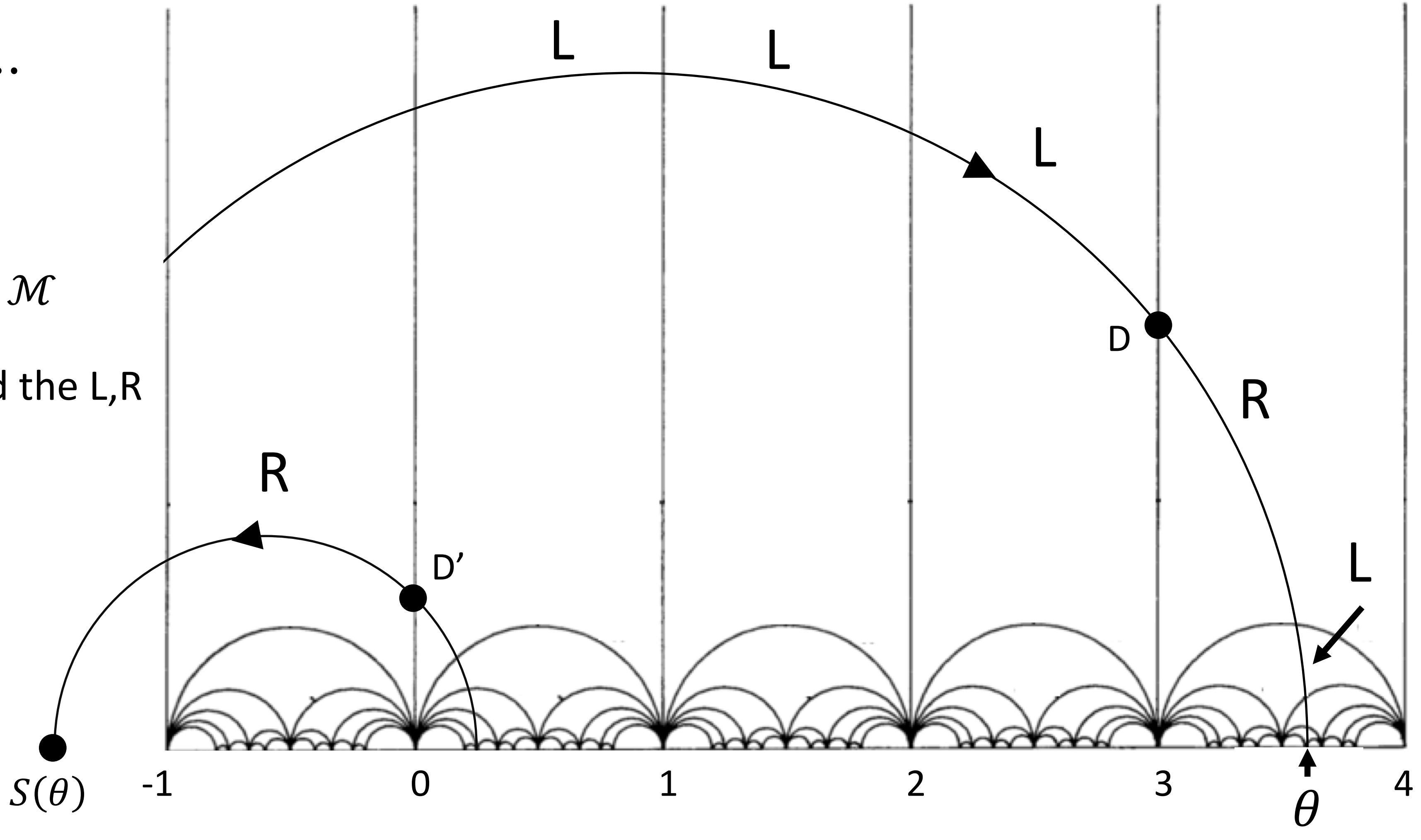
$$L^{a_0} R^{a_1} L^{a_2} \cdots$$

- We see $a_0 = \lfloor \theta \rfloor$
- The map

$$S(z) = -\frac{1}{z - a_0} \in \mathcal{M}$$

preserves the lattice and the L,R labels.

- $S = QP^{-a_0}$
- D maps to D'



Characteristic Sequences

- Consider a finite word in *a*, *b*:

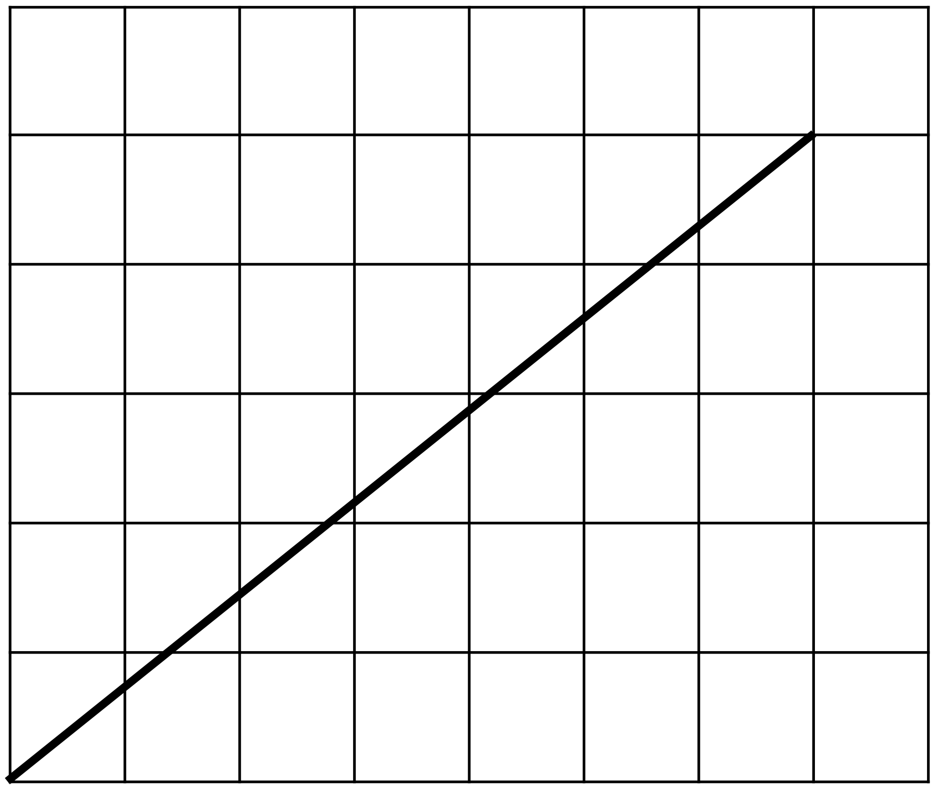
$$ababab^2abab^2$$

- Notice that the *a* are by themselves and that the *b* exponents differ by at most 1.
- Such a word is called *almost constant*.
- We can *derive* an almost constant word to get a new word.

Step	Word	Recoding		Derived Seq
0	ab^2abab^2abab	$ab \to a$	$b \to b$	1
1	aba^2ba^2	$a \to a$	$ba^2 \to b$	2
2	ab^2	$ab^2 \to a$		2
3	a			

Characteristic Sequences

- Almost constant sequences that can be derived indefinitely are called *characteristic.*
- This is another technique for calculating continued fractions!
- Plot a line of slope $0 < \theta < 1$
- Write an *a* whenever it crosses the horizontal lines, and a *b* for the vertical. Write *ab* for lattice points.
- The derived sequence is the continued fraction!



$$ab^2abab^2abab$$

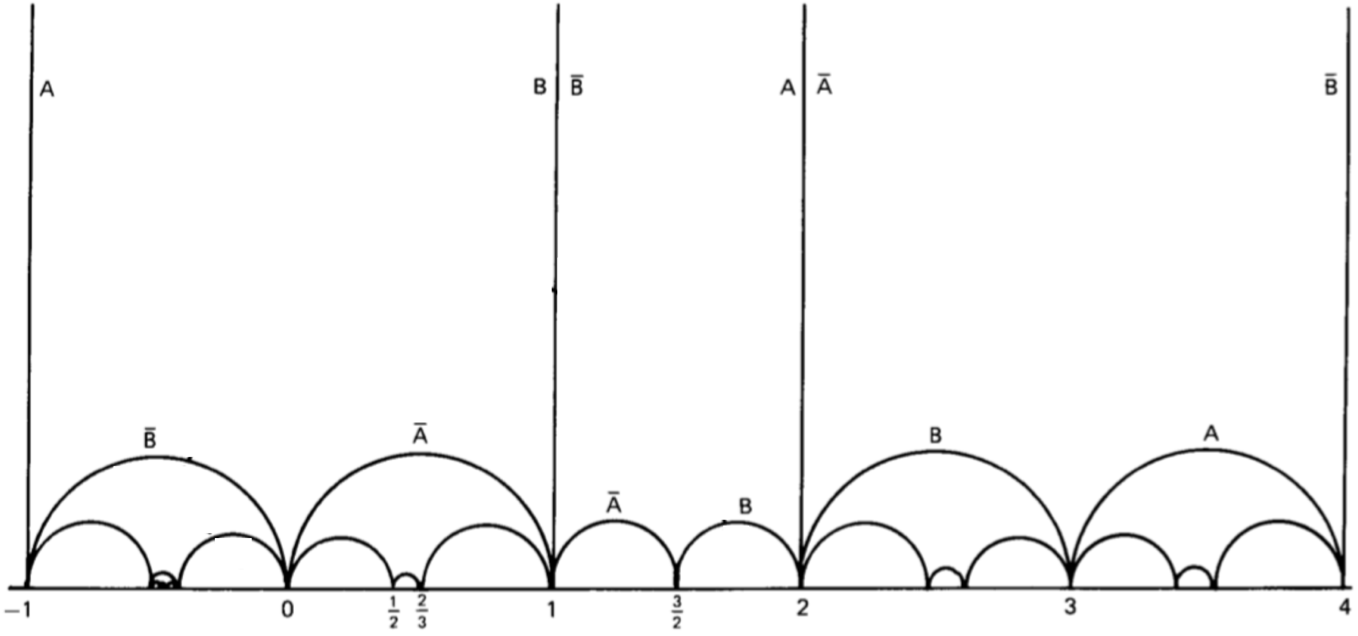
$$\theta = \frac{5}{7} = [0, 1, 2, 2]$$

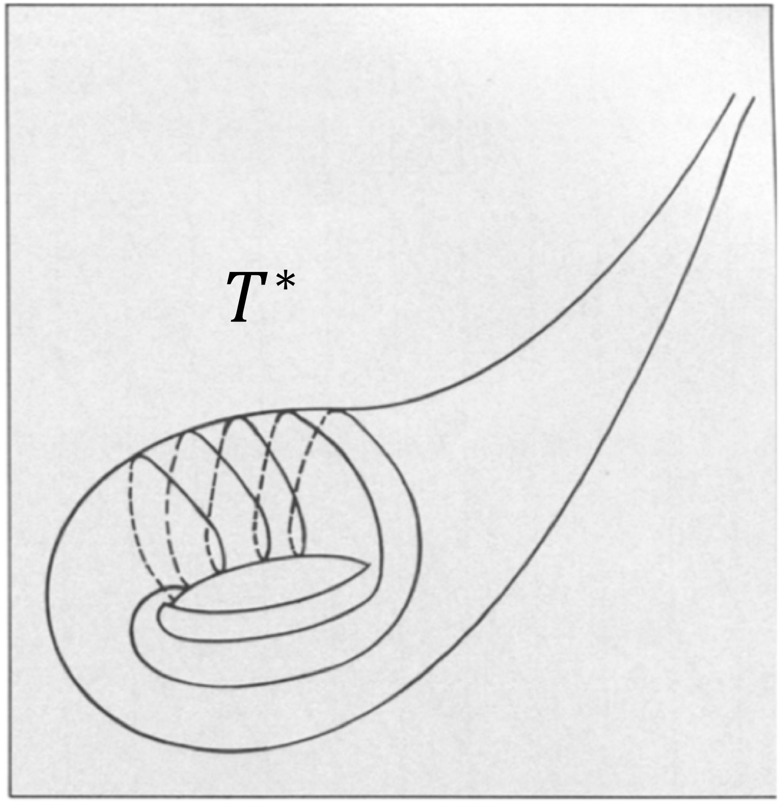
Characteristic Sequences

- Every cutting sequence of a Euclidean line is characteristic.
- Conversely, every *finite* characteristic sequence is the cutting sequence of a line.
- The situation for hyperbolic geodesics is a little more complicated, as we shall see.

Taking Quotients of $\mathbb{H}$

- We must change lattices again





A: $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ B: $\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$

Definitions for Geodesics in T^*

- A *simple* curve does not intersect itself.
- A *closed* curve is periodic.
- A *lift* of a curve $\bar{\gamma}$ in T^* is any curve γ in $\mathbb{H}$ that projects to $\bar{\gamma}$.
- Geodesics lift to geodesics.
- If a sequence of geodesics $\{\bar{\gamma}_n\}$ has a sequence of lifts $\{\gamma_n\}$ whose endpoints converge to x, y, then the projection $\bar{\gamma}$ of the geodesic γ corresponding to the endpoints x, y is called the *limit* of $\{\bar{\gamma}_n\}$.
- The collection of all limits of simple, closed geodesics is denote *X.*
- All geodesics of *X* are simple, though not necessarily closed.

Bringing it all together

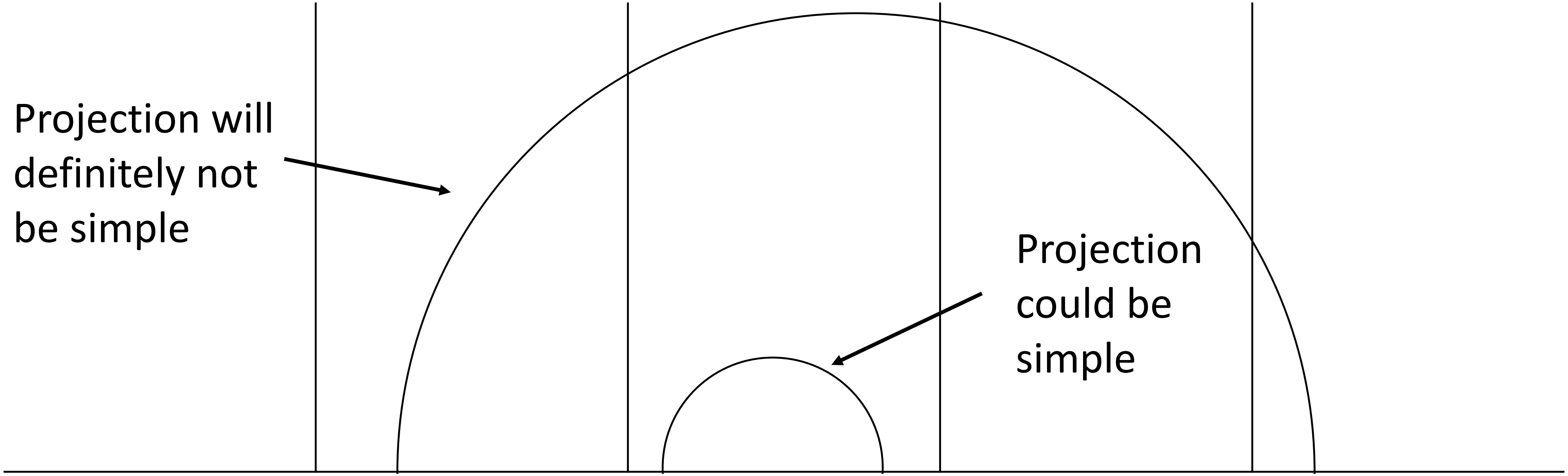
A geodesic on T^* is closed and simple if and only if its cutting sequence is periodic and characteristic.

A geodesic is in *X* if and only if its cutting sequence is characteristic.

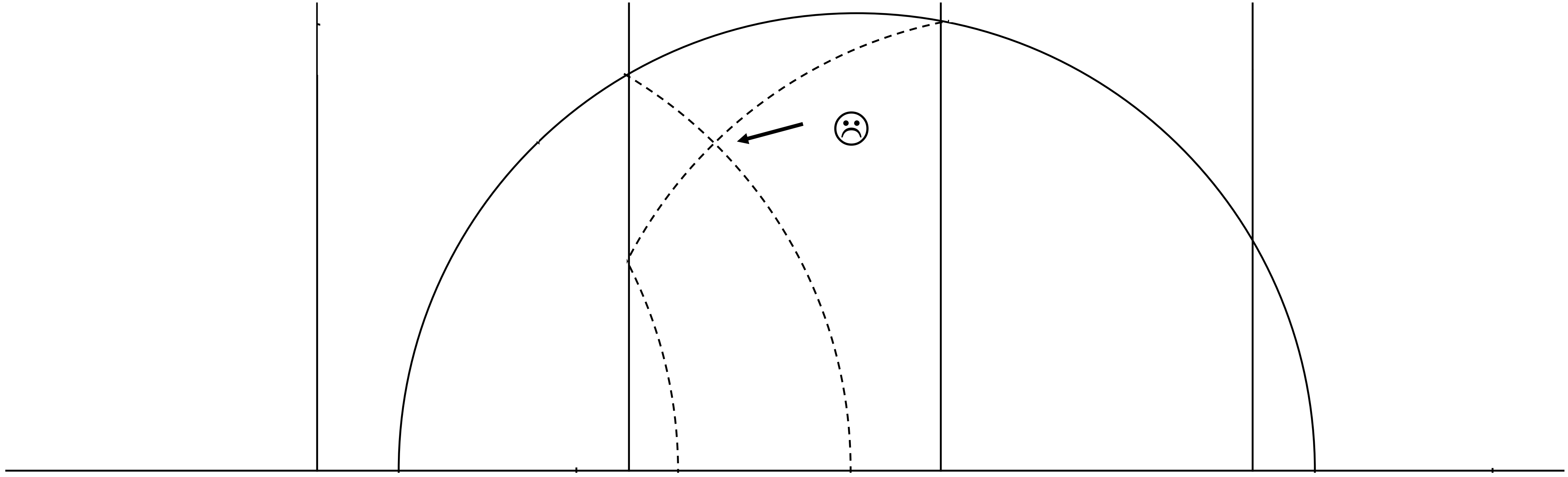
A geodesic is in *X* if and only if all of its lifts lie below the line $\operatorname{Im} z < 3/2$.

Markoff irrationalities are the endpoints of lifts of geodesics in *X*.

Motivation for $\mathbf{Im}\, z < 3/2$



Motivation for $\mathbf{Im}\, z < 3/2$



Markoff's Quadratic Form

- Determining if the $\mathrm{Im}\, z < 3/2$ condition holds is equivalent to determining if the endpoints x_1, x_2 satisfy $|x_1 - x_2| < 3$.
- Choose c_0, c_1, c_2 so that the quadratic polynomial
$$f(x) \coloneqq c_2 x^2 + c_1 x + c_0$$
has x_1, x_2 as roots. Denote
$$\Delta = c_1^2 - 4c_0 c_2.$$
- Notice that
$$|x_1 - x_2| = \frac{\sqrt{\Delta}}{f(1)}$$

Markoff's Quadratic Form

$$|x_1 - x_2| = \frac{\sqrt{\Delta}}{f(1)}$$

- Let us define the corresponding *binary* quadratic form

$$Q(x, y) = c_2x^2 + c_1xy + c_0y^2$$

- To prove that $|x_1 - x_2| < 3$, it therefore suffices to prove that

$$\inf_{x,y \in \mathbb{Z}^2 \setminus 0} \frac{Q(x, y)}{\sqrt{\Delta}} > \frac{1}{3}$$

- In fact, the infimum is $\nu(x_1) = \nu(x_2)$!

Diophantine Equations

- Consider the Diophantine equation
$$x^2 + y^2 + z^2 = 3xyz$$
- The trivial solution is
$$(1,1,1)$$
- *All* solutions can be derived from this solution by some application of
$$(x, y, z) \mapsto (z, x, y)$$
$$(x, y, z) \mapsto (x, 3xy - z, y)$$
- Let
$$x_{\pm} = \frac{1}{2} + \frac{y}{xz} \pm \frac{1}{2}\sqrt{9 - \frac{4}{z^2}}$$

Diophantine Equations

- The numbers $x_\pm$ are Markoff irrationalities and

$$\nu(x_+) = \nu(x_-) = \left(9 - \frac{4}{z^2}\right)^{-1/2} > \frac{1}{3}.$$

What just happened?

- We associated continued fractions with *cutting sequences* of geodesics in $\mathbb{H}$.
- We found that Markoff irrationalities are the limits of endpoints of closed, simple geodesics in T^*
- We found that such limits can be associated to the minima of a certain binary quadratic form.
- The minima of this form can be associated to the solution of a Diophantine equation.

References

- This presentation is based primarily on C. Series' Math Intelligencer article, *The Geometry of Markoff Numbers,* 1985.
- For a concise treatment of the basics of hyperbolic geometry, consult chapter one of *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces,* edited by T. Bedford, M. Keane, and C. Series.