Continued fractions and geodesics on the modular surface

Chris Johnson
Clemson University

September 18, 2013

Outline

The modular surface

Continued fractions

Symbolic coding

References

Some hyperbolic geometry

Recall that the hyperbolic plane is modelled by the upper half-plane,

$$\mathbb{H} = \left\{ x + iy \,\middle|\, y > 0 \right\},$$

together with the metric

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

Some hyperbolic geometry

Recall that the hyperbolic plane is modelled by the upper half-plane,

$$\mathbb{H} = \left\{ x + iy \,\middle|\, y > 0 \right\},$$

together with the metric

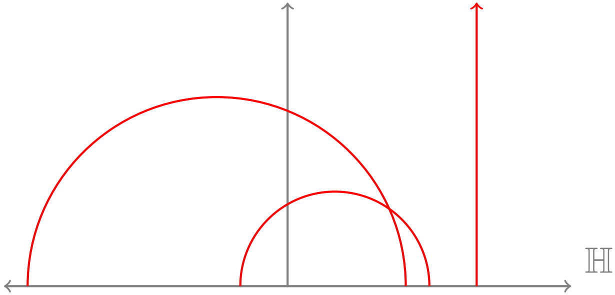
$$ds = \frac{\sqrt{dx^2 + dy^2}}{y}.$$

This just means that the arclength of a curve $\gamma(t) = x(t) + iy(t)$, $a \leq t \leq b$, is given by

$$\text{Length of } \gamma = \int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} dt.$$

Some hyperbolic geometry

Using the hyperbolic metric, geodesics in $\mathbb{H}$ are vertical lines and semi-circles orthogonal to the real axis.



Möbius Transformations

The (orientation-preserving) isometries of $\mathbb{H}$ are precisely the maps

$$z \mapsto \frac{az+b}{cz+d}$$

where $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$.

This group of isometries is represented by the group of 2×2 real matrices with determinant 1:

$$\mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R}).$$

Hyperbolic surfaces

Every discrete subgroup G of $\mathrm{PSL}(2,\mathbb{R})$ acts freely and properly discontinuously on $\mathbb{H}$ – i.e., the quotient $\mathbb{H}/G$ is a surface. This surface inherits a hyperbolic metric from $\mathbb{H}$, and all hyperbolic surfaces arise in this way.

One particularly important such subgroup is the *modular group*, $\Gamma := \mathrm{PSL}(2,\mathbb{Z})$. The quotient $\mathbb{H}/\Gamma$ is called the *modular surface*.

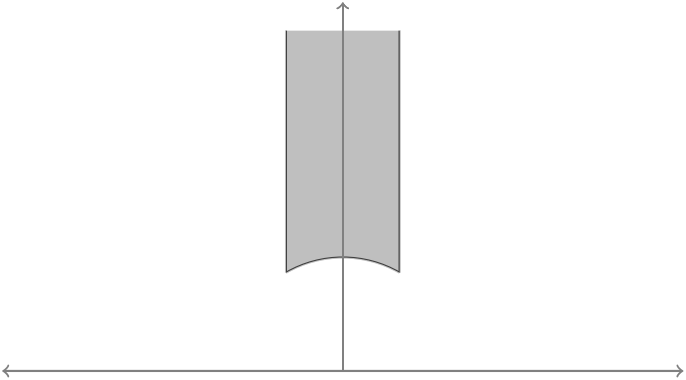
Hyperbolic surfaces

Every discrete subgroup G of $\mathrm{PSL}(2, \mathbb{R})$ acts freely and properly discontinuously on $\mathbb{H}$ – i.e., the quotient $\mathbb{H}/G$ is a surface. This surface inherits a hyperbolic metric from $\mathbb{H}$, and all hyperbolic surfaces arise in this way.

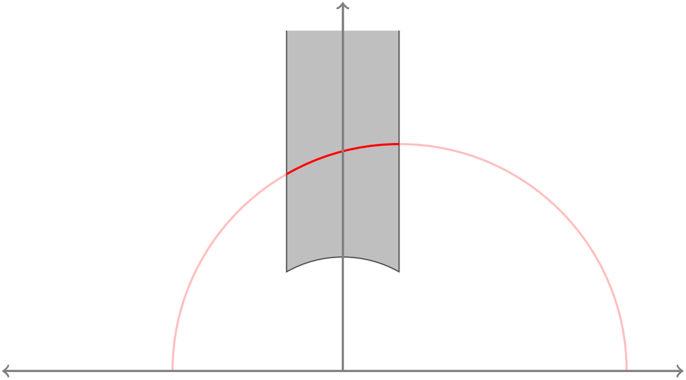
One particularly important such subgroup is the *modular group*, $\Gamma := \mathrm{PSL}(2, \mathbb{Z})$. The quotient $\mathbb{H}/\Gamma$ is called the *modular surface*.

Why care about this space? It parametrizes all complex tori!

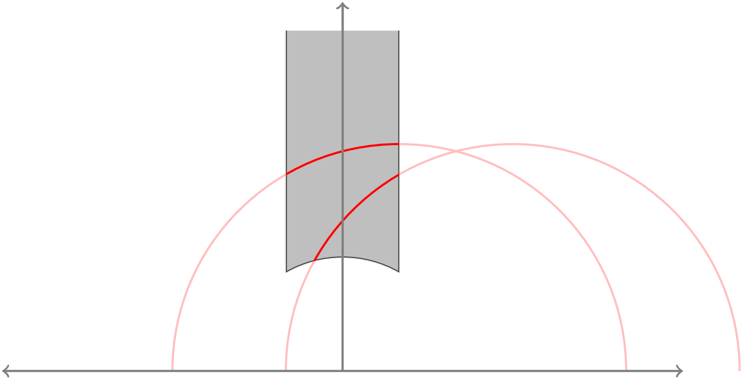
The modular surface



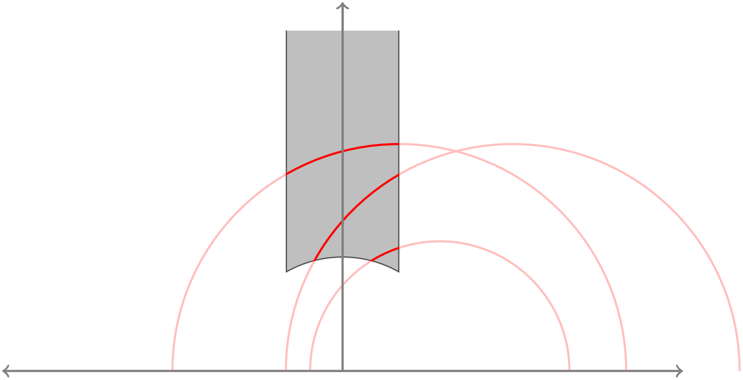
The modular surface



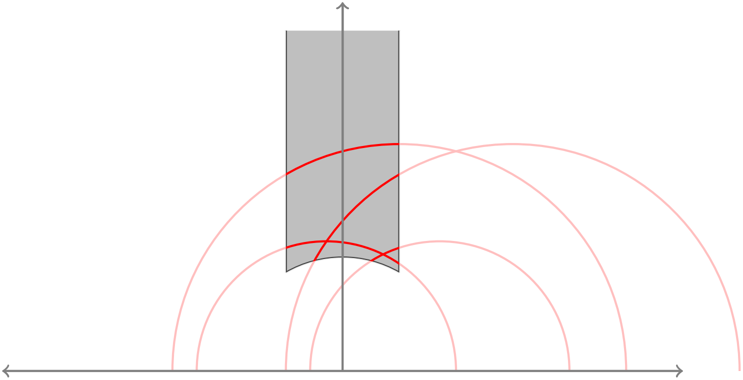
The modular surface



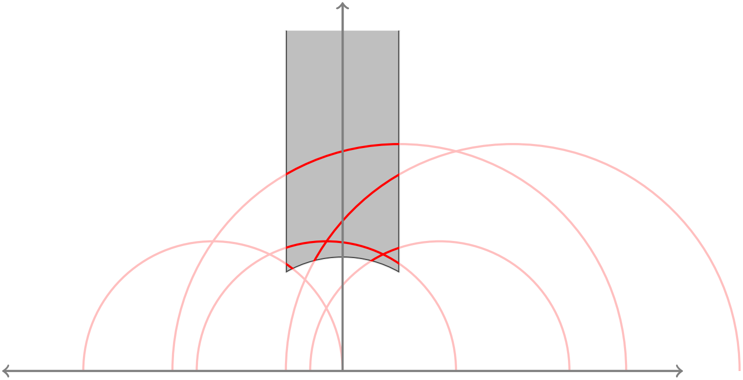
The modular surface



The modular surface



The modular surface



Questions

These geodesics may wind around the surface in complicated ways. How can we determine the behavior of a given geodesic on the surface?

Are there any periodic geodesics? Are there any dense geodesics?

If both possibilities occur, is one more likely than the other? If we pick a random geodesic, is it more likely to be periodic, dense, or something else?

Emil Artin studied these questions by cleverly encoding geodesics using continued fractions.

What is a continued fraction?

A *continued fraction* is a way to represent a real number as a fraction of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ddots}}}$$

where $a_i \in \mathbb{Z}$ and $a_j > 0$ for $j > 0$.

Every real number has a continued fraction expansion, which is finite if and only if the number is rational.

$$\pi = 3.14159...$$

$$\begin{aligned}\pi &= 3.14159...\\ &= 3 + r_1\end{aligned}$$

$$\begin{aligned}\pi &= 3.14159... \\ &= 3 + r_1 \\ &= 3 + \cfrac{1}{\left(\frac{1}{r_1}\right)}\end{aligned}$$

$$\begin{aligned}\pi &= 3.14159...\\ &= 3 + r_1\\ &= 3 + \frac{1}{\left(\frac{1}{r_1}\right)}\\ &= 3 + \frac{1}{7 + r_2}\end{aligned}$$

$$\begin{aligned}
\pi &= 3.14159\ldots \\
&= 3 + r_1 \\
&= 3 + \cfrac{1}{\left(\frac{1}{r_1}\right)} \\
&= 3 + \cfrac{1}{7 + r_2} \\
&= 3 + \cfrac{1}{7 + \cfrac{1}{\left(\frac{1}{r_2}\right)}}
\end{aligned}$$

$$\begin{aligned}
\pi &= 3.14159... \\
&= 3 + r_1 \\
&= 3 + \cfrac{1}{\left(\frac{1}{r_1}\right)} \\
&= 3 + \cfrac{1}{7 + r_2} \\
&= 3 + \cfrac{1}{7 + \cfrac{1}{\left(\frac{1}{r_2}\right)}} \\
&= 3 + \cfrac{1}{7 + \cfrac{1}{15 + r_3}}
\end{aligned}$$

Partial quotients

The a_i in $a_0 + 1/(a_1 + 1/(a_2 + \cdots))$ are called the *partial quotients* of the number.

The Gauss map, $T : (0, 1) \to (0, 1)$ defined by

$$T(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor,$$

is related to the partial quotients by

$$a_{n\geq 1} = \left\lfloor \frac{1}{T^{n-1}(x)} \right\rfloor.$$

(Note $a_0 = \lfloor x \rfloor$.)

Distribution of partial quotients

Theorem

For almost every real number, the frequency an integer n appears as a partial quotient (i.e., how often some $a_i = n$) is given by

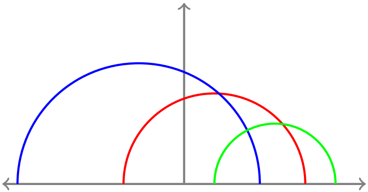
$$\frac{2\log(1+n) - \log(n) - \log(2+n)}{\log(2)}.$$

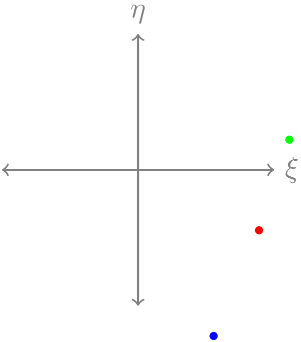
Theorem

Fix any finite sequence of natural numbers $(b_1, b_2, ..., b_m) \in \mathbb{N}^m$. Almost every real number contains $b_1, ..., b_m$ as consecutive partial quotients of its continued fraction expansion.

The space of geodesics

Each geodesic on $\mathbb{H}/\Gamma$ lifts (not uniquely!) to a geodesic on $\mathbb{H}$. Applying an element of $\mathrm{PSL}(2,\mathbb{Z})$, we may assume that this geodesic is a semicircle with endpoints $\xi > \eta$ on the real axis.





We think of (ξ, η) as coordinates for the space of all geodesics on the surface.

Coding geodesics

This gives the space of all geodesics a topology: we'll say two geodesics are "close" if their corresponding (ξ, η) coordinates are "close."

If we move a circle around with an element of $\mathrm{PSL}(2, \mathbb{Z})$, then the endpoints move in the same way.

Since we can move things around by $\mathrm{PSL}(2, \mathbb{Z})$, we can restrict our attention to the strip

$$\left\{(\xi, \eta) \,\middle|\, \xi \geq 1,\ -1 \leq \eta \leq 0\right\}.$$

Coding geodesics

To show that a geodesic on $\mathbb{H}/\Gamma$ fills the surface, we just need to show the $\mathrm{PSL}(2, \mathbb{Z})$-orbit of the point in the $\xi\eta$-plane is dense. (If the geodesic gets close to every point on the plane, then it must get close to every circle.)

To actually show this, use continued fractions.

Coding geodesics

Write down the continued fraction expansions of ξ and $-\eta$,

$$\xi = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}} \qquad -\eta = \cfrac{1}{a_{-1} + \cfrac{1}{a_{-2} + \cfrac{1}{a_{-3} + \ddots}}}$$

Put the partial quotients into a bi-infinite sequence:

$$(\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots)$$

Coding geodesics

From the bi-infinite sequence we construct a sequence of endpoints for circles.

$$\xi_n = a_n + \cfrac{1}{a_{n+1} + \cfrac{1}{a_{n+2} + \ddots}} \qquad -\eta_n = \cfrac{1}{a_{n-1} + \cfrac{1}{a_{n-2} + \ddots}}$$

These endpoints are related to the original ξ and η by the relations

$$\begin{aligned}\xi &= \frac{P_n \xi_n + P_{n-1}}{Q_n \eta_n + Q_{n-1}} \\ \eta &= \frac{P_n \eta_n + P_{n-1}}{Q_n \eta_n + Q_{n-1}}\end{aligned}$$

where P_n, Q_n are integers satisfying $P_n Q_{n-1} - P_{n-1} Q_n = (-1)^n$.

This means each (ξ_{2n}, η_{2n}) is in the $\mathrm{SL}(2, \mathbb{Z})$ orbit of (ξ, η). So to see if $\mathrm{SL}(2, \mathbb{Z}) \cdot (\xi, \eta)$ gets close to some given (x, y), we only need to find some (ξ_{2n}, η_{2n}) that's close!

Coding geodesics

The key observation: If two numbers are sufficiently close together, the first n partial quotients of their continued fractions must agree.

However, we know that almost every continued fraction contains any given finite sequence of consecutive partial quotients. This means that, almost surely, some (ξ_{2n}, η_{2n}) gets arbitrarily close to any given (x, y).

Coding geodesics

The key observation: If two numbers are sufficiently close together, the first *n* partial quotients of their continued fractions must agree.

However, we know that almost every continued fraction contains any given finite sequence of consecutive partial quotients. This means that, almost surely, some (ξ_{2n}, η_{2n}) gets arbitrarily close to any given (x, y).

Hence almost every geodesic on $\mathbb{H}/\Gamma$ is dense.

Other Properties

Other information, besides density of orbits, can be read off the continued fractions.

- A geodesic is periodic if and only if $(..., a_{-2}, a_{-1}, a_0, a_1, a_2, ...)$ is periodic.
- A geodesic asymptotically approaches a periodic trajectory if and only if the sequence is eventually periodic.
- A geodesic approaches the cusp of $\mathbb{H}/\Gamma$ if and only if at least one side of the sequence is finite (i.e., one of ξ or η is a rational number.)

An application

These ideas provide an alternative proof, using Masur's criterion, that almost every geodesic on a flat torus is uniquely ergodic (in particular, uniformly distributed).

In other words, if you shoot a billiard in a random direction on a rectangular billiard table, then with probability 1 you'll pick a direction so that the billiard ball visits every region of the table. Furthermore, if two regions of the table have the same area, the billiard will spend the same amount of time inside of those regions.

References I

Thanks to Brandon Edwards of Oregon State University for telling me about Artin's original paper, and providing an English translation.

- Emil Artin, *Ein mechanisches System mit quasiergodischen Bahnen*, Abh. Math. Sem. Univ. Hamburg **3** (1924), no. 1, 170–175. MR 3069425
- Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London Ltd., London, 2011. MR 2723325 (2012d:37016)
- Caroline Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31** (1985), no. 1, 69–80. MR 810563 (87c:58094)