Simple continued fractions and the geodesic flow on the modular surface

Robert Hines

September 23, 2017

Euclidean algorithm and simple continued fractions

The integers are a Euclidean domain, with respect to the usual absolut value for instance. For a pair $a, b \neq 0$ we have

$$\begin{aligned} a &= ba_0 + r_0,\ 0 \le |r_0| < |b| \\ b &= r_0 a_1 + r_1,\ 0 \le |r_1| < |r_0| \\ r_0 &= r_1 a_2 + r_2,\ 0 \le |r_2| < |r_1| \\ &\dots \end{aligned}$$

or written in matrices

$$\begin{aligned} \begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r_0 \end{pmatrix} \\ &= \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r_0 \\ r_1 \end{pmatrix} \\ &= \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \\ &\dots \end{aligned}$$

which if $(a, b) = 1$ gives

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thinking of this as an algorithm on rationals (dividing by $b, r_0, r_1, \dots$ in the first array or acting by fractional linear transformations in the second) we obtain an expression for a/b as a continued fraction

$$\frac{a}{b} = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\dots \frac{1}{a_n}}}} =: [a_0; a_1, \dots, a_n].$$

Extending this to irrational numbers $\xi = \lfloor \xi \rfloor + \{\xi\} = a_0 + \xi_0$ gives a dynamical system

$$T : [0, 1) \to [0, 1),\ \xi \mapsto \{1/\xi\}$$

and infinite sequences

$$\xi_n = T^n \xi_0,\ a_{n+1} = \left\lfloor \frac{1}{T^n \xi_0} \right\rfloor,\ \begin{pmatrix} p_n & q_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \dots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

One can verify the following properties by induction or from the definitions:

$$q_n/q_{n-1} > 1 \ (n \geq 2),\ q_n \geq 2^{\frac{n-1}{2}} \ (n \geq 0),\ q_n\xi - p_n = (-1)^n x_0 \cdot \ldots \cdot \xi_n = \frac{(-1)^n \xi_n}{q_{n-1}\xi_n + q_n},$$

$$\frac{p_n}{q_n} = a_0 - \sum_{k=1}^{n} \frac{(-1)^k}{q_k q_{k-1}},\ \xi = \frac{p_n + p_{n-1}\xi_n}{q_n + q_{n-1}\xi_n},\ \frac{1}{q_{n+2}} \leq |q_n\xi - p_n| \leq \frac{1}{q_{n+1}}.$$

Hence for irrational ξ we have convergence

$$\xi = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \ldots}} =: [a_0; a_1, a_2, \ldots].$$

The branches of T^{-1} are all surjective and we have bijections

$$\mathbb{R} \setminus \mathbb{Q} \cong \mathbb{Z} \times \mathbb{N}^{\mathbb{N}},\ \mathbb{Q} = \{[a_0; a_1, \ldots, a_n] : n \geq 0,\ a_n \neq 1 \text{ if } n \geq 1\}.$$

T is the left shift on these sequences, $T([0; a_1, a_2, \ldots]) = [0; a_2, a_3, \ldots]$.

Invertible extension and invariant measure

The map T is not invertible, but we can construct an invertible extension on $\mathcal{G} = (-\infty, 0) \times (0, 1)$

$$\widetilde{T} : \mathcal{G} \to \mathcal{G},\ \widetilde{T}(\eta, \xi) = (1/\eta - \lfloor 1/\xi \rfloor, 1/\xi - \lfloor 1/\xi \rfloor),$$

defined piecewise by Möbius transformations acting diagonally

$$\widetilde{T}(\eta, \xi) = (g \cdot \eta, g \cdot \xi),\ \widetilde{T}^{-1}(\eta, \xi) = (h \cdot \eta, h \cdot \xi),$$

$$g = \begin{pmatrix} -\lfloor 1/\xi \rfloor & 1 \\ 1 & 0 \end{pmatrix},\ h = \begin{pmatrix} 0 & 1 \\ 1 & \lfloor -\eta \rfloor \end{pmatrix}.$$

We will think of $\mathcal{G} = (-\infty, 0) \times (0, 1)$ as a space of geodesics in hyperbolic 2-space, $\mathbb{H}^2$, and the action of $\widetilde{T}$ takes this space to itself piecewise by isometries $\mathrm{Isom}(\mathbb{H}^2) \cong PGL_2(\mathbb{R})$, where $PGL_2(\mathbb{R})$ acts on the upper half-plane by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d} \text{ or } \frac{a\bar{z} + b}{c\bar{z} + d}$$

according as the determinant $ad - bc$ is positive or negative.

Let $\mathbb{H}^2$ have coordinates (x, y) with area $\frac{dxdy}{y^2}$. The top of the geodesic (η, ξ) (semi-circle from η to ξ with center on the rea line $x = 0$) has coordinates $\left(\frac{\xi+\eta}{2}, \frac{\xi-\eta}{2}\right)$, so that hyperbolic area becomes $d\tilde{\mu}(\eta, \xi) := \frac{d\xi d\eta}{(\xi-\eta)^2}$ in these coordinates. This gives an isometry invariant measure on our space of geodesics, and since $\widetilde{T}$ is a bijection defined piecewise by isometries, $\tilde{\mu}$ is $\widetilde{T}$-invariant. Pushing forward to the second coordinate gives a T-invariant measure on $(0, 1)$

$$d\mu(\xi) = d\xi \int_{-\infty}^{-1} \frac{d\eta}{(\xi - \eta)^2} = \frac{d\xi}{1 + \xi}$$

which we will normalize to a probability measure by dividing by $\log 2$. This measure was known to Gauss, although there is no indication of how he arrived at it. His mastery of quadratic forms (represented by the geodesics above) and knowledge of hyperbolic geometry could make something like the above plausible.

Direct proof of ergodicity

For fixed $a_1, \ldots, a_n \in \mathbb{N}$, we have the cylinder set

$$\Delta_n = \{\psi(t) = \frac{p_n + p_{n-1}t}{q_n + q_{n-1}t} : 0 \leq t < 1\}$$

which is the half-open interval between p_n/q_n and $\frac{p_n+p_{n-1}}{q_n+q_{n-1}}$ (oriented depending on the parity of n). These cylinder sets generate the Borel σ-algebra. If λ is Lebesgue measure, then we have (bar indicating conditional probability)

$$\lambda(T^{-n}[s,t) \mid \Delta_n) = \frac{\psi(t) - \psi(s)}{\psi(1) - \psi(0)} = (t-s)\frac{q_n(q_n + q_{n-1})}{(q_n + q_{n-1}s)(q_n + q_{n-1}t)} = (t-s)C,$$

where $1/2 \leq C \leq 2$. Hence there exists (a different) $C > 0$ such that

$$\frac{1}{C}\mu(A) \leq \mu(T^{-n}A \mid \Delta_n) \leq C\mu(A)$$

for measurable A.

Considering T-invariant sets of positive measure, we have

$$\begin{aligned} T^{-1}A = A &\Rightarrow \frac{1}{C}\mu(A) \leq \mu(A \mid \Delta_n) \\ \mu(A) > 0 &\Rightarrow \frac{1}{C}\mu(\Delta_n) \leq \mu(\Delta_n \mid A) \\ &\Rightarrow \frac{1}{C}\mu(B) \leq \mu(B \mid A) \text{ for any measurable} B \\ A^c = B &\Rightarrow \mu(A^c) = 0,\ \mu(A) = 1. \end{aligned}$$

Hence μ is ergodic.

Consequences of ergodicity

We can apply the following ergodic theorem to various functions to obtain almost everywhere statistics for continued fractions.

Theorem. *Supose (X, T, μ) is a measure preserving system and $f \in L^1$, then the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{n-1} f \circ T^n = f^*$$

exists almost everywhere and

$$\int_X f d\mu = \int_X f^* d\mu$$

(i.e. f^* is the conditional expectation of f with respect to the σ-algebra of T-invariant sets). In particular, if the system is ergodic then f^* is constant almost everywhere,

$$f^* = \int_X f d\mu.$$

For instance:

- If f is the indicator of the interval $(\frac{1}{k+1}, \frac{1}{k})$, we get

$$\mathbb{P}(a_n = k) = \lim_{N\to\infty} \frac{1}{N} |\{i : a_i = k,\ 1 \le i \le N\}| = \frac{1}{\log 2} \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{d\xi}{1+\xi} = \frac{1}{\log 2} \log\left(\frac{(k+1)^2}{k(k+2)}\right),$$

k	1	2	3	4
$\mathbb{P}(a_n = k)$	41.58%	16.99%	9.31%	5.89%

- Taking f to be $\sum_k \log k \cdot \mathbf{1}_{(\frac{1}{k+1}, \frac{1}{k})}$, we get

$$\lim_{N\to\infty} \left(\prod_{n=1}^{N} a_n\right)^{1/N} = \exp\left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \log a_n\right) = \exp\left(\frac{1}{\log 2} \sum_k \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{\log k}{1+\xi} d\xi\right)$$
$$= \prod_k \left(\frac{(k+1)^2}{k(k+2)}\right)^{\frac{\log k}{\log 2}} = 2.6854520010\ldots$$

- With $f_M = \sum_{k\le M} k \cdot \mathbf{1}_{(\frac{1}{k+1}, \frac{1}{k})}$ and taking $M \to \infty$ we get

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{\infty} a_n = \infty.$$

With a little more work we can show

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = \frac{1}{\log 2} \cdot \frac{\pi^2}{12},\ \lim_{n\to\infty} \frac{1}{n} \log |\xi - p_n/q_n| = -\frac{1}{\log 2} \cdot \frac{\pi^2}{6}.$$

We have

$$\begin{pmatrix} \xi \\ 1 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} 1 \\ T^n\xi \end{pmatrix},\ T^n\xi = (-1)^{n-1} \frac{q_n\xi - p_n}{q_{n-1}\xi - p_{n-1}},$$

so that

$$\prod_{k=0}^{n-1} T^k\xi = (-1)^n (q_{n-1}\xi - p_{n-1}) = |q_{n-1}\xi - p_{n-1}|.$$

Hence, from the list of properties in the first section, we have

$$\frac{1}{q_{n+1}} \le \prod_{k=0}^{n-1} T^k\xi \le \frac{1}{q_n}.$$

Taking logarithms and letting $n \to \infty$ we get

$$-\lim_{n\to\infty} \frac{1}{n} \log q_{n+1} \le \lim_{n\to\infty} \sum_{k=0}^{n-1} \log T^k\xi \le -\lim_{n\to\infty} \frac{1}{q_n} \log q_n,$$

so that

$$\lim_{n\to\infty} \frac{1}{n} \log q_n = -\frac{1}{\log 2} \int_0^1 \frac{\log \xi}{1+\xi} d\xi = \frac{1}{\log 2} \sum_{k=0}^{\infty} (-1)^{k+1} \int_0^1 \xi^k \log \xi\ d\xi$$
$$= \frac{1}{\log 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2} = \frac{1}{\log 2} \cdot \frac{\pi^2}{12}.$$

Since $\frac{1}{q_n q_{n+2}} \leq |\xi - \frac{p_n}{q_n}| \leq \frac{1}{q_n q_{n+1}}$, the above gives

$$\lim_{n\to\infty} \frac{1}{n} \log|\xi - p_n/q_n| = -\frac{1}{\log 2} \cdot \frac{\pi^2}{6}.$$

Finally, a result on the normalized error $\theta_n(\xi) = q_n|q_n\xi - p_n|$ (assuming $\widetilde{T}$ is ergodic):

Proposition. *For μ almost every ξ, we have*

$$\lim_{N\to\infty} \frac{1}{N}|\{1 \leq n \leq N : \theta_n(\xi) \leq t\}| = \begin{cases} \frac{t}{\log 2} & 0 \leq t \leq 1/2 \\ 1 + \frac{1-t+\log t}{\log 2} & 1/2 \leq t \leq 1 \\ 1 & t \geq 1 \end{cases}.$$

Proof. Note that

$$\widetilde{T}^n(-\infty, \xi) = (-q_n/q_{n-1}, T^n\xi) = (-[a_n; a_{n-1}, \dots, a_1], [0; a_{n+1}, a_{n+2}, \dots])$$
$$\theta_n(\xi) = \frac{1}{1/T^n\xi + q_{n-1}/q_n} = \frac{1}{[a_{n+1}; a_{n+2}, \dots] + [0; a_n, \dots, a_1]},$$

so that $\theta_n(\xi) \leq t$ iff $\frac{1}{1/\xi' - 1/\eta'} \leq t$ where $(\eta', \xi') = \widetilde{T}^n(-\infty, \xi)$. Let $\mathcal{G}(c) = \{(\eta, \xi) \in \mathcal{G} : 1/\xi - 1/\eta \geq c\}$ Then for $\epsilon > 0$ and n large, we have

$$\widetilde{T}^n(\eta, \xi) \in \mathcal{G}(1/t + \epsilon) \Rightarrow \widetilde{T}^n(-\infty, \xi) \in \mathcal{G}(1/t) \Rightarrow \widetilde{T}^n(\eta, \xi) \in \mathcal{G}(1/t - \epsilon).$$

The measure of $\mathcal{G}(c)$ with respect to $\frac{d\xi d\eta}{(\xi-\eta)^2 \log 2}$ for $c \geq 1$ is

$$\begin{cases} \frac{1}{\log 2}\left(1 - \frac{1}{c} + \log 2 - \log c\right) & 1 \leq c \leq 2 \\ \frac{1}{c \log 2} & c \geq 2 \end{cases},$$

which gives the result when $t = 1/c$. $\square$

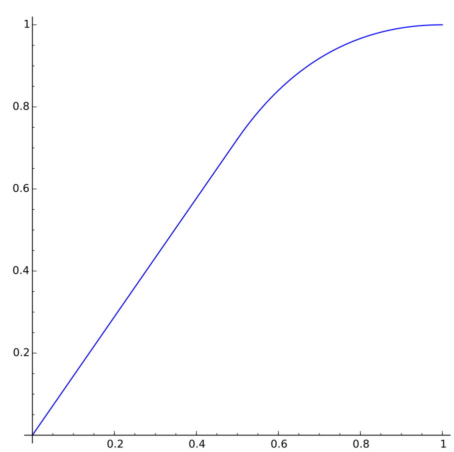

Figure 1: Distribution of the normalized error for almost every ξ.

Geodesic flow on the modular surface and continued fractions

The group $SL_2(\mathbb{R})$ acts transitively on the upper half-plane by fractional linear transformations, and the stabilizer of $z = i$ is $SO_2(\mathbb{R})$. Hence, as a homogeneous space, $\mathbb{H}^2 \cong SL_2(\mathbb{R})/SO_2(\mathbb{R})$. Moreover $SL_2(\mathbb{R})$ acts as orientation perserving isometries and in fact $PSL_2(\mathbb{R}) \cong \mathrm{Isom}^+(\mathbb{H}^2)$.

We can identify $PSL_2(\mathbb{R})$ with $T^1(\mathbb{H}^2)$, the unit tangent bundle, as follows. Let $(z, v) \in T^1(\mathbb{H}^2) \subseteq \mathbb{H}^2 \times \mathbb{C}$, where $z = x+yi$ is in the upper half-plane and v has norm 1 in the hyperbolic metric, i.e. if $v = v_1 + v_2 i$, then $\langle v, v\rangle_z := v_1 v_2/y^2 = 1$. Define the derivative action of $SL_2(\mathbb{R})$ on $T^1(\mathbb{H}^2)$ by

$$g \cdot (z, v) = (g(z), g'(z)v) = \left(\frac{az + b}{cz + d}, \frac{v}{(cz + d)^2}\right).$$

One can verify that this action is isometric and transitive with kernel ± 1, so that we get the identification $PSL_2(\mathbb{R}) \cong T^1(\mathbb{H}^2)$.

Under the NAK decomposition, with

$$p = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} \in NA, \ \ k = \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \in K,$$

we have $pk(i, i) = (x + yi, yie^{i\theta})$.

Associating the identity with the point $(z, v) = (i, i) \in T^1(\mathbb{H}^2)$, the *geodesic flow* $\Phi_t : T^1(\mathbb{H}^2) \to T^1(\mathbb{H}^2)$ is right multiplication by $\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$. It takes a point (z, v) along a unit speed geodesic in the direction v for time t, as can be checked along the imaginary axis and extended via the isometric action.

The group $\Gamma = SL_2(\mathbb{Z})$ is a discrete subgroup of $SL_2(\mathbb{R})$ and the quotient $\Gamma\backslash SL_2(\mathbb{R})$ has finite volume, equivalently, the quotient $\Gamma\backslash\mathbb{H}^2$ has finite hyperbolic area (namely $\pi/3$). We identify $\Gamma\backslash SL_2(\mathbb{R})$ with the unit tangent bundle of the modular surface.

We now want to relate continued fractions (or its invertible extension) with the geodesic flow on the modular surface. There is a slight complication because continued fractions are defined over $GL_2(\mathbb{Z})$. Consider the following subsets of $T^1(\mathbb{H}^2)$

$$\begin{aligned} \mathcal{C}^+ &= \{(z, v) \in \overrightarrow{\eta\xi} : (\eta, \xi) \in \mathcal{G},\ z \in i\mathbb{R}\}, \\ \mathcal{C}^- &= \{(z, v) \in \overrightarrow{(-\eta)(-\xi)} : (\eta, \xi) \in \mathcal{G},\ z \in i\mathbb{R}\}, \\ \mathcal{C} &= \mathcal{C}^+ \cup \mathcal{C}^-, \end{aligned}$$

i.e. those points and directions of the intersection of elements of $\mathcal{G}$ with the imaginary axis (and their reflected images). Let $\pi : T^1(\mathbb{H}^2) \to T^1(\Gamma\backslash\mathbb{H}^2)$ be the natural projection.

Proposition. *If $(z, v) = (\eta, \xi) \in \pi(\mathcal{C}^+)$, the next return of the geodesic flow to $\pi(\mathcal{C})$ is in $\pi(\mathcal{C}^-)$ with coordinates $-\widetilde{T}(\eta, \xi)$, and similarly for $(z, v) = (-\eta, -\xi) \in \pi(\mathcal{C}^-)$, the next return of the geodesic flow to $\pi(\mathcal{C})$ is in $\pi(\mathcal{C}^+)$ with coordinates $\widetilde{T}(\eta, \xi)$. In other words, the map*

$$S : \mathcal{G} \cup -\mathcal{G},\ \ S(\eta, \xi) = -\widetilde{T}(\eta, \xi),\ \ S(-\eta, -\xi) = \widetilde{T}(\eta, \xi),$$

is the first return of the geodesic flow to the cross section $\pi(\mathcal{C})$.

Proof. Proof by picture. Applying $-1/z$ to $(\eta, \xi) \in (-\infty, -1) \times (0, 1)$ gives $(-1/\eta, -1/\xi) \in (0, 1) \times (-\infty, -1)$. Following the geodesic flow to the next intersection with $\pi(\mathcal{C})$ corresponds to translation by $\lfloor 1/\xi\rfloor$ where we end up in $-\mathcal{G}$ with coordinates $(\lfloor 1/\xi\rfloor - 1/\eta, \lfloor 1/\xi\rfloor - 1/\xi) = -\widetilde{T}(\eta, \xi)$. Similarly for the other case. $\square$

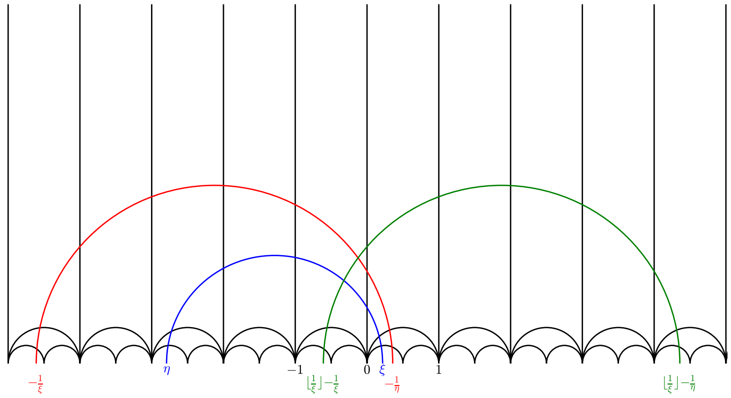

Figure 2: Visualizing S as first return of the geodesic flow to $\pi(\mathcal{C})$.

To complete the picture of how continued fractions fits into the geodesic flow, we need to consider the *return time* $r(\eta, \xi)$, $r(-\eta, -\xi)$ to the cross section. Here is a more general construction, the *special flow* under a function.

Proposition. *Suppose (X, T, μ) is a measure preserving system, and $f : X \to (0, \infty)$ is μ-measurable. Let $X_f = \{(x, s) : 0 \leq s < f(x)\}$ and for each $t \geq 0$ let $\phi_t : X_f \to X_f$ be defined by*

$$\phi_t(x, s) = \left(T^n x, s + t - \sum_{k=0}^{n-1} f(T^k x) \right)$$

where n is the least non-negative integer such that $0 \leq s + t < \sum_{k=0}^{n} f(T^k x)$. Let $\mu_f = \mu \times \lambda$ restricted to X_f.

- *μ_f is ϕ_t-invariant for all $t \geq 0$,*
- *$\mu_f(X_f) < \infty \Leftrightarrow f \in L^1(X, \mu)$,*
- *If $\mu_f(X_f) < \infty$, then (X, T, μ) is ergodic $\Leftrightarrow$ the flow $\{\phi_t\}_{t \geq 0}$ is ergodic.*

Let $r : \mathcal{G} \cup -\mathcal{G} \to (0, \infty)$ be the return time of the associated $(z, v) \in \pi(\mathcal{C})$ to $\pi(\mathcal{C})$, $X_r = (\mathcal{G} \cup -\mathcal{G})_r$, $S(\eta, \xi)$ as above, and ϕ the special flow associated to r. Let $\Sigma = SL_2(\mathbb{Z}) \backslash SL_2(\mathbb{R}) / SO_2(\mathbb{R})$ be the modular surface, $T^1(\Sigma) = SL_2(\mathbb{Z}) \backslash SL_2(\mathbb{R})$ its unit tangent bundle, and Φ the geodesic flow (right multiplication by $\mathrm{diag}(e^{t/2}, e^{-t/2})$). We have

Proposition. *The following diagram commutes*

$$\begin{array}{ccc} X_r & \xrightarrow{\phi_t} & X_r \\ \downarrow & & \downarrow \\ T^1(\Sigma) & \xrightarrow{\Phi_t} & T^1(\Sigma) \end{array}$$

where the arrows on the left and right are $((y, x), s) \mapsto \Phi_s(z, v)$. Moreover, the measure μ_r for the special flow under the return time (constructed from Gauss measure and Lebesgue measure) is the pullback (up to a multiplicative constant) of the Φ-invariant measure $\frac{d\theta dx dy}{y^2}$ on the unit tangent bundle (Haar measure on $PSL_2(\mathbb{R})$).

Proof. That the above commutes follows from our earlier discussion, so we now consider the measures involved. We have two coordinate systems on $PSL_2(\mathbb{R}) \cong T^1(\mathbb{H}^2)$, thinking of a point $z = x + yi$ and direction $v = iye^{i\theta}$ at z,

$$g(x,y,\theta) = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} \begin{pmatrix} \cos(\theta/2) & \sin(\theta/2) \\ -\sin(\theta/2) & \cos(\theta/2) \end{pmatrix},$$

discussed above, and another obtained by associating to (z, v) the geodesic (η, ξ) it determines and the distance/time one must go from the "top" of the geodesic

$$g(\eta,\xi,t) = \frac{1}{\sqrt{|\eta-\xi|}} \begin{pmatrix} \max\{\xi,\eta\} & \min\{\xi,\eta\} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\epsilon} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix},$$

with $\epsilon = 0$ if $\xi > \eta$ and $\epsilon = 1$ otherwise. One can think of this as taking the geodsic $\overrightarrow{0\infty}$ with marked point i, mapping it to the geodesic $\overrightarrow{\eta\xi}$ with marked point at the top, and flowing for the required time. We have an invariant measure on each of these

$$d\mu dt = \frac{d\eta d\xi}{(\xi-\eta)^2}dt, \ \frac{dxdy}{y^2}d\theta$$

and we would like to show that these are the same (perhaps up to a constant). Assume $\xi > \eta$ (the other case being similar). We have

$$\frac{1}{\sqrt{|\eta-\xi|}} \begin{pmatrix} \xi & \eta \\ 1 & 1 \end{pmatrix} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \cdot (i,i) = (x+yi, iye^{i\theta})$$

where

$$x = \frac{\xi e^t + \eta e^{-t}}{e^t + e^{-t}}, \ y = \frac{\xi - \eta}{e^t + e^{-t}}, \ \theta = \arctan\left(\frac{1}{\sinh t}\right).$$

Some computation shows

$$\frac{dxdyd\theta}{y^2} = \left(\frac{e^t + e^{-t}}{\xi - \eta}\right)^2 \begin{vmatrix} \frac{e^t}{e^t+e^{-t}} & \frac{e^{-t}}{e^t+e^{-t}} & * \\ \frac{1}{e^t+e^{-t}} & \frac{-1}{e^t+e^{-t}} & * \\ 0 & 0 & \frac{-1}{\cosh t} \end{vmatrix} = 2\frac{d\eta d\xi dt}{(\xi-\eta)^2},$$

as desired. ☐

We can compute the return time from the above,

$$r(\pm\eta, \pm\xi) = \frac{1}{2}\log\left(\frac{1 - \eta\lfloor 1/\xi \rfloor}{1 - \xi\lfloor 1/\xi \rfloor}\right),$$

and its integral must be

$$\int_{\mathcal{G}\cup-\mathcal{G}} r(\eta,\xi)\frac{d\xi d\eta}{(\xi-\eta)^2} = \frac{\pi^2}{3}$$

since the total volume of $T^1(\Sigma)$ is $2\pi^2/3$. (I couldn't compute this integral, and neither could Mathematica, but it agrees numerically within the estimated error.)

Mixing of the geodesic flow

Mixing of the geodesic flow on the modular surface is implied by the following "decay of matrix coefficients" theorem, applied to $\mathcal{H} = L_0^2(SL_2(\mathbb{Z})\backslash SL_2(\mathbb{R}))$ (L^2 functions f with $\int f = 0$) and $g = \mathrm{diag}(e^{t/2}, e^{-t/2})$.

Theorem (Howe-Moore)**.** *Suppose $\rho : SL_2(\mathbb{R}) \to \mathcal{U}(\mathcal{H})$ be a unitary representation with no non-zero fixed vectors, i.e. $\{v \in \mathcal{H} : \rho(G)v = v\} = \{0\}$. Then for all $v, w \in \mathcal{H}$, $\langle \rho(g)v, w\rangle \to 0$ as $g \to \infty$ (i.e. for any $\epsilon > 0$, there exists $K \subseteq G$ compact such that for $g \notin K$, $\langle \rho(g)v, w\rangle < \epsilon$).*

Proof. We first note that any $g \in SL_2(\mathbb{R})$ can be written uniquely as KA^+K (Cartain decomposition) where K is $SO_2(\mathbb{R})$ and A^+ is the collection of $\mathrm{diag}(a, a^{-1})$, $a > 0$. (Proof: diagonalize the quadratic form $\|gx\|^2$.) Let $g_n \to \infty$, $g_n = k_n a_n l_n$, and for $v, w \in \mathcal{H}$, let $\tilde{v}$, $\tilde{w}$ be weak limits of $\rho(l_n)v$, $\rho(k_n^{-1})w$. We have

$$\begin{aligned}
&\langle \rho(k_n a_n l_n)v, w\rangle - \langle \rho(a_n)\tilde{v}, \tilde{w}\rangle \\
&= \langle \rho(a_n)(\rho(l_n)v - \tilde{v}), \rho(k_n^{-1})w\rangle + \langle \rho(a_n), \rho(k_n^{-1})w - \tilde{w}\rangle \to 0.
\end{aligned}$$

Therefore we may assume $g_n = a_n = \mathrm{diag}(t_n, 1/t_n)$ with $t_n \to \infty$.

Let

$$u_s^+ = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix},\ u_s^- = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}.$$

Then

$$a_n^{-1} u_s^+ a_n = u_{s/t_n^2}^+ \to 1,\ a_n u_s^- a_n^{-1} = u_{s/t_n^2}^- \to 1.$$

Let $v, w \in \mathcal{H}$, and let E be a weak limit of $\rho(a_n)$ (diagonal argument on an orthonormal basis). We want to show that $E = 0$ (in particular,$\langle \rho(a_n)v, w\rangle \to \langle Ev, w\rangle = 0$). We have

$$\langle \rho(u_s^+)Ev, w\rangle = \lim_n \langle \rho(u_s^+)\rho(a_n)v, w\rangle = \lim_n \langle \rho(a_n)\rho(a_n^{-1}u_s^+ a_n)v, w\rangle = \langle Ev, w\rangle,$$

and similarly $\langle \rho(u_s^-)E^*v, w\rangle = \langle E^*v, w\rangle$. Also, $\langle \rho(a_n)v, \rho(a_m)v\rangle = \langle \rho(a_m^{-1})v, \rho(a_n^{-1})w\rangle$ (since A is Abelian), so that $EE^* = E^*E$, i.e. E is normal. If $E \neq 0$, then $EE^* \neq 0$ and there is some $v \in \mathcal{H}$ such that $w = E^*Ev = EE^*v \neq 0$. This w is invariant under $\{u_s^\pm\}_s$, which generates $SL_2(\mathbb{R})$, contradicting the assumption that π has no non-zero invariant vectors. Hence $E = 0$. □

References

[B] P. Billingsley, *Ergodic theory and information theory*

[K] A. Ya. Khinchin, *Continued Fractions*

[EW] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*

[M] David Witte Morris, *Introduction to arithmetic groups*

[S] W. M. Schmidt, *Diophantine approximation*