(*November 6, 2015*)

Fundamental domains for $SL_2(\mathbb{Z})$ and Γ_θ

Paul Garrett garrett@math.umn.edu http://www.math.umn.edu/~garrett/

[This document is http://www.math.umn.edu/~garrett/m/mfms/notes_2015-16/11_fund_dmn.pdf]

The real line $\mathbb{R}$, being a *group*, is a *homogeneous space* in the sense that (of course) it acts *transitively* on itself. The upper half-plane $\mathfrak{H}$ is not a group itself, but *is* acted upon by $SL_2(\mathbb{R})$ (2-by-2 real matrices with determinant 1) acting by *linear fractional transformations*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) \;=\; \frac{az+b}{cz+d}$$

We will see that the simplest quotient of the upper half-plane, $SL_2(\mathbb{Z})\backslash\mathfrak{H}$, is *topologically* a sphere with a point missing. However, in the $SL_2(\mathbb{R})$-invariant geometry, the missing point is *infinitely far away*, so the shape is not a *round* sphere, but stretched out like a *raindrop*.

---

1. *$\mathfrak{H}$ as homogeneous space for $SL_2(\mathbb{R})$*

The group $\mathbb{R}$ acting on $\mathbb{R}$ or $\mathbb{R}/\mathbb{Z}$ is not very different from the circle itself. Circles $\mathbb{R}/\mathbb{Z}$ are groups themselves, since $\mathbb{Z}$ is *normal* in $\mathbb{R}$, unavoidable since $\mathbb{R}$ is *abelian*. In contrast, the upper half-plane

$$\mathfrak{H} = \{z = x + iy \;:\; y > 0\} \subset \mathbb{C}$$

does *not* have any reasonable group structure itself. Luckily, the group

$$G \;=\; SL_2(\mathbb{R}) \;=\; \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \;:\; \text{ real matrices with } \; ad - bc = 1\}$$

acts on $\mathfrak{H}$ with the *linear fractional transformation action*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) \;=\; \frac{az+b}{cz+d}$$

[1.0.1] Claim: The group $SL_2(\mathbb{R})$ stabilizes $\mathfrak{H}$ and acts transitively on it. [1] In particular,

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{y} & 0 \\ 0 & \frac{1}{\sqrt{y}} \end{bmatrix}(i) = x + iy \qquad (\text{for } x \in \mathbb{R},\, y > 0)$$

Further, for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R})$ and $z \in \mathfrak{H}$

$$\text{Im } g(z) = \frac{\text{Im } z}{|cz+d|^2}$$

[1] In fact, *every* holomorphic automorphism of $\mathfrak{H}$ is given by an element of $SL_2(\mathbb{R})$. This follows from *Schwarz' lemma* (on the disk), which allows us to deduce that an automorphism of the disk fixing 0 is a rotation.

Proof: The first formula is clear. The second formula would imply that the upper half-plane is stabilized. Compute directly:

$$2i \cdot \mathrm{Im}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z)\right) = \frac{az+b}{cz+d} - \frac{a\overline{z}+b}{c\overline{z}+d} = \frac{(az+b)(c\overline{z}+d) - (a\overline{z}+b)(cz+d)}{|cz+d|^2}$$

$$= \frac{adz - bc\overline{z} - bcz + ad\overline{z}}{|cz+d|^2} = \frac{z-\overline{z}}{|cz+d|^2}$$

since $ad - bc = 1$. ///

[1.0.2] Remark: The extra information about how the imaginary part transforms will be useful in determining a *fundamental domain* just below.

[1.0.3] Claim: The isotropy group in $SL_2(\mathbb{R})$ of the point $i \in H$ is the *special orthogonal group*[2]

$$SO(2) = \{g \in SL_2(\mathbb{R}) : \ g^\top \cdot g = 1_2\} = \left\{\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} : \ \theta \in \mathbb{R}\right\}$$

Proof: For real a, b, c, d, the equation $(ai+b)/(ci+d) = i$ gives $ai + b = -c + id$, so $a = d$ and $c = -b$. The determinant condition $ad - bc = 1$ gives $a^2 + b^2 = 1$, which we can reparametrized via trigonometric functions as indicated. ///

[1.0.4] Corollary: We have an isomorphism of $SL_2(\mathbb{R})$-spaces

$$SL_2(\mathbb{R})/SO(2) \approx \mathfrak{H} \quad \text{via} \quad g \cdot SO(2) \to g(i)$$

That is, that map respects the action of $SL_2(\mathbb{R})$, in the sense that

$$g \cdot (h \cdot SO(2)) \ \longrightarrow \ g(h(i))$$

Proof: This is because of *associativity*:

$$g \cdot (h \cdot SO(2)) \ = \ (gh) \cdot SO(2) \ \longrightarrow \ (gh)(i) \ = \ g(h(i))$$

giving the result. ///

[1.0.5] Remark: Proving associativity of the linear fractional transformation action directly is pretty ugly. Instead, use the fact that linear fractional transformation action *descends* from *linear* action of $GL_n(\mathbb{C})$ on $\mathbb{CP}^1$, via the inclusion $\mathbb{C} \to \mathbb{CP}^1$ by

$$z \ \longrightarrow \ \begin{pmatrix} z \\ 1 \end{pmatrix} \cdot \mathbb{C}^\times \ \in \ (\mathbb{C}^2 - \{0\})/\mathbb{C}^\times \ = \ \mathbb{CP}^1$$

[2] The choice of corner in which to put the $-\sin\theta$ does not matter much in the larger scheme of things, and often the opposite choice is made, but there are some reasons one might make the present choice. Still, it doesn't really matter.

2. *Fundamental domain for* $\Gamma = SL_2(\mathbb{Z})$ *on* $\mathfrak{H}$

The simplest beginning choice of discrete subgroup Γ of $G = SL_2(\mathbb{R})$ is

$$\Gamma \;=\; SL_2(\mathbb{Z}) = \{\text{2-by-2 integer matrices with determinant 1}\}$$

Both for use below and to show that $SL_2(\mathbb{Z})$ is a large group, note:

[2.0.1] Claim: Given *relatively prime* integers c, d, there are integers a, b such that $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$.

Proof: For *any* integers c, d, there are integers m, n such that

$$\text{greatest common divisor } c, d \;=\; m \cdot c + n \cdot d$$

Here the greatest common divisor is 1, so take $a = n$, $b = -m$, and then $ad - bc = 1$. ///

To be able to draw a picture of the quotient, we take an archaic [3] approach which nevertheless succeeds in this case. First, we find a *fundamental domain* for Γ on $\mathfrak{H}$, meaning to find a *nice* set of representatives for the quotient. Second, see how the edges of the fundamental domain are glued together when mapped to the quotient $\Gamma\backslash\mathfrak{H}$.

[2.0.2] Claim: Every Γ-orbit in $\mathfrak{H}$ has a representative in

$$\overline{F} \;=\; \{z \in \mathfrak{H} : |z| \geq 1,\ |\mathrm{Re}(z)| \leq \frac{1}{2}\}$$

More precisely, each orbit has a *unique* representative in the *standard fundamental domain*

$$F \;=\; \{z \in \mathfrak{H} : |z| > 1,\ -\frac{1}{2} \leq \mathrm{Re}(z) < \frac{1}{2}\} \cup \{z \in \mathfrak{H} : |z| = 1, \mathrm{Re}(z) \leq 0\}$$

[2.0.3] Remark: The fundamental domain is illustrated in the picture *[... iou ...]*...

Proof: From above, for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$

$$\mathrm{Im}\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{\mathrm{Im}\ z}{|cz+d|^2}$$

The set of complex numbers $cz + d$ is a subset of the lattice $\mathbb{Z} \cdot z + \mathbb{Z} \subset \mathbb{C}$. Since it is a discrete *subgroup*, it has (at least one) smallest (in absolute value) non-zero element.

Thus, $\inf |cz + d| = \min |cz + d| > 0$, taking the infimum or minimum over *relatively prime* c, d, which we have observed are exactly the lower rows of elements of Γ. Then

$$\sup \frac{1}{|cz+d|} = \max \frac{1}{|cz+d|} < \infty$$

[3] The approach which succeeds in general is a descendant of the present argument. As usual, the modernized argument is successful because it discards many (adroitly chosen) details, which in hindsight were inessential.

Thus, for fixed $z \in \mathfrak{H}$,

$$\sup \mathrm{Im} \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z) \;=\; \sup \frac{\mathrm{Im}\, z}{|cz+d|^2} = \max \frac{\mathrm{Im}\, z}{|cz+d|^2} < \infty$$

Thus, in each Γ-orbit there is (at least one) point z assuming the maximum value of $\mathrm{Im}\, z$ on that orbit.

Since $\mathrm{Im} \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z) = \mathrm{Im}\, z/|cz+d|^2$, for z giving maximal $\mathrm{Im}\, z$ in its orbit, it must be that

$$|cz+d| \;\geq\; 1$$

for all c, d relatively prime. Thus, for example, for $d = 0$ there is the *inversion*

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} (z) \;=\; -1/z$$

Thus, $|1 \cdot z + 0| \geq 1$, so for $\mathrm{Im}\, z$ maximal in its Γ-orbit, $|z| \geq 1$.

We can adjust any $z \in \mathfrak{H}$ by

$$\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} (z) \;=\; z + n \quad (\text{for } n \in \mathbb{Z})$$

to normalize $-1/2 \leq \mathrm{Re}(z) < 1/2$.

So take $|z| \geq 1$ and $|\mathrm{Re}(z)| \leq 1/2$ and show that $|cz+d| \geq 1$ for *all* c, d. Break z into its real and imaginary parts $z = x + iy$. Then

$$|cz+d|^2 \;=\; (cx+d)^2 + c^2y^2 \;=\; c^2(x^2+y^2) + 2cdx + d^2 \;\geq\; c^2(x^2+y^2) - |cd| + d^2$$

$$\geq\; c^2(|z|^2 - \frac{1}{4}) + \frac{c^2}{4} - |cd| + d^2 \;\geq\; c^2(|z|^2 - \frac{1}{4})$$

Thus, for $|c| \geq 2$, we have $|cz+d| > 1$ when $|z| \geq 1$ and $|x| \leq 1/2$.

For $c = 0$, necessarily $d = \pm 1$, and the only corresponding elements of Γ are

$$\begin{bmatrix} \pm 1 & n \\ 0 & \pm 1 \end{bmatrix}$$

The only z's with $|z| \geq 1$ and $|x| \leq 1/2$ that can be mapped to each other by such group elements are $-\frac{1}{2} + iy$ and $\frac{1}{2} + iy$. We whimsically keep the former as our chosen representative.

For $c = \pm 1$,

$$|cz+d|^2 \;=\; 2xd + d^2 + |z|^2 \geq -|d| + d^2 + 1 \;\geq\; 1 \qquad\qquad (\text{for } d \in \mathbb{Z})$$

In fact, for $|x| < 1/2$, there is a *strict* inequality

$$2xd + d^2 + |z|^2 \;>\; -|d| + d^2 + 1 \;\geq\; 1$$

so $|cz+d| > 1$. When $|x| = 1/2$, still $-|d| + d^2 + 1 > 1$, *except* for $d = 0, \pm 1$.

Thus, first without worrying about strictness of the inequalities, $|cz+d| \geq 1$ for $|z| \geq 1$ and $|x| \leq 1/2$, and the set $\overline{F}$ contains (at least one) representative for every orbit. What remains is to eliminate duplicates.

We have already observed that the only duplicates for $|z| > 1$ have $|x| = 1/2$, and $z \to z + 1$ maps the $x = -1/2$ line to the $x = 1/2$ line.

Now consider $|z| = 1$. For $|x| < 1/2$, the only cases where $|cz + d| = 1$ are with $c = \pm 1$ and $d = 0$, which correspondes to matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{bmatrix} * & \pm 1 \\ \mp 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & \pm 1 \\ \mp 1 & 0 \end{bmatrix} \qquad \text{(for some } n \in \mathbb{Z})$$

For $|z| = 1$, the inversion $z \to -1/z$ maps $z = x + iy$ to

$$-\frac{1}{z} = -\overline{z}/|z|^2 = -\overline{z} = -x + iy$$

Thus, for $|x| < 1/2$, the only one among these products that maps z back to the fundamental domain is exactly the inversion $z \to -1/z$. This inversion identifies the two arcs

$$\{|z| = 1 \text{ and } -\frac{1}{2} \le x \le 0\} \quad \{|z| = 1 \text{ and } 0 \le x \le \frac{1}{2}\}$$

Thus, we should include only one or the other of these two arcs in the strict fundamental domain.

Last, with $|z| = 1$ and $|x| = 1/2$, there are exactly four group elements modulo $\pm 1_2$ (the center $\{\pm 1_2\}$ acts trivially) that map z to the closure of the fundamental region. These are: the identity, one of the translations $z \to z \pm 1$, the inversion $z \to -1/z$, and the *composite* of the translation and the inversion. That is, in addition to the identity,

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{map } -\frac{1}{2} + \frac{i\sqrt{3}}{2} \text{ to the boundary of } \overline{F}$$

and

$$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad \text{map } \frac{1}{2} + \frac{i\sqrt{3}}{2} \text{ to the boundary of } \overline{F}$$

Thus, in the quotient $\Gamma\backslash\mathfrak{H}$, the identification of the sides $x = \pm 1$ creates a (topological) cyclinder, and the identification of the two arcs on the bottom closes the bottom of the cylinder. Thus, topologically, we have a cylinder closed at one end, which is a disk. But the non-euclidean geometry [4] (if we were to pay more attention to details) suggests that the *top* of the cylinder is infinitely far away, and the radius of the cylinder goes to 0 as one goes toward the open top end, so it is more accurate to think of the quotient $\Gamma\backslash\mathfrak{H}$ as a raindrop shape. ///

[... iou ...] pictures

---

3. *Inversion and translation generate $SL_2(\mathbb{Z})$*

[3.0.1] Claim: The inversion (long Weyl element) $w = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and translations $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ with $n \in \mathbb{Z}$ *generate* $\Gamma = SL_2(\mathbb{Z})$.

Proof: Again use the fact that $\mathbb{Z} \cdot z + \mathbb{Z}$ is a *lattice* in $\mathbb{C}$. In particular, there is *no* infinite sequence of decreasing sizes $|c_1 z + d_1| > |c_2 z + d_2| > \dots$ with integers c_j, d_j. Thus, there is no infinite *increasing* sequence of heights

$$\frac{y}{|c_1 z + d_1|^2} < \frac{y}{|c_2 z + d_2|^2} < \dots$$

[4] We will look at the non-euclidean geometry of the upper half-plane and other examples a little later.

Since $\mathrm{Im}\,(\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z)) = \dfrac{y}{|cz+d|^2}$, this implies that there is *no* infinite increasing sequence

$$\mathrm{Im}\,(\gamma_1 z) \;<\; \mathrm{Im}\,(\gamma_2 z) \;<\; \dots \qquad\qquad (\text{for } \gamma_j \in \Gamma)$$

This promises that the following procedure does eventually put every point $z \in \mathfrak{H}$ inside the standard fundamental domain for Γ.

Given $z \in \mathfrak{H}$, translate z to z_1 satisfying $|\mathrm{Re}\,(z_1)| \leq \frac{1}{2}$. If $|z_1| \geq 1$, stop: z_1 is in the fundamental domain. If $|z_1| < 1$, apply the inversion, noting

$$\mathrm{Im}\,(\frac{-1}{z_1}) \;=\; \frac{\mathrm{Im}\,(z_1)}{|z_1|^2} \;>\; \mathrm{Im}\,(z_1) \qquad\qquad (\text{since } |z_1| < 1)$$

Continue: translate $-1/z_1$ back to z_2 in the strip. If $|z_2| \geq 1$, stop. If $|z_2| < 1$, invert. Translate back to z_3 in the strip, and so on. The sequence $\mathrm{Im}\,(z_1) < \mathrm{Im}\,(z_2) < \dots$ must be *finite*, so the process terminates after finitely many steps.

Thus, given $\gamma \in \Gamma$, take z in the *interior* of the fundamental domain, and let δ be a finite product of inversions and integer translations so that $\delta^{-1}\gamma z$ is back in the fundamental domain. Since z is in the interior, $\delta^{-1}\gamma = \pm 1_2$. Since $w^2 = -1_2$, necessarily γ is expressible in terms of inversions and integer translations. ///

[3.0.2] Remark: The number of steps require to move a given $z \in \mathfrak{H}$ into the fundamental domain is not simple to describe. This complication is visible in pictures of the tiling of the upper half-plane by images of the fundamental domain.

---

4. *Re-enabling the action of* $SL_2(\mathbb{R})$

Taking the quotient $\Gamma\backslash\mathfrak{H}$ of $\mathfrak{H}$ by $\Gamma = SL_2(\mathbb{Z})$ obstructs the group action of $SL_2(\mathbb{R})$, since Γ is far from being a *normal* subgroup of G.

That is, while the upper half-plane $\mathfrak{H}$ *is* a homogeneous space for $SL_2(\mathbb{R})$, being

$$\mathfrak{H} \approx SL_2(\mathbb{R})/SO(2) \quad (SO(2) \text{ the isotropy group of } i \in \mathfrak{H})$$

but the group $SL_2(\mathbb{R})$ no longer acts on

$$\Gamma\backslash\mathfrak{H} \;\approx\; \Gamma\backslash SL_2(\mathbb{R})/SO(2)$$

because since the Γ gets in the way: $SL_2(\mathbb{R})$ normalizes *no* such Γ. Thus, the $SL_2(\mathbb{R})$-homogeneity is difficult to see or use in this form.

We *could* have $SL_2(\mathbb{R})$ act on the *right* on $SL_2(\mathbb{Z})\backslash\mathfrak{H} \approx SL_2(\mathbb{Z})\backslash SL_2(\mathbb{R})/SO(2)$ if the $SO(2)$ weren't in the way. Indeed, recovery of the action of $SL_2(\mathbb{R})$ is a powerful argument in favor of giving up the (otherwise appealing) complex structure on $\mathfrak{H}$.

---

5. *Fundamental domain for* Γ_θ *and* $\Gamma(2)$

The determination of the standard fundamental domain F for $\Gamma(1) = SL_2(\mathbb{Z})$ allows explicit determination of fundamental domains for finite-index *subgroups* such as the *principal congruence subgroups*

$$\Gamma(N) \;=\; \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \;:\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod N\}$$

by choosing coset representatives γ_i for $\Gamma(N)$ in $\Gamma(1)$, and then [5]

$$\text{fundamental domain for } \Gamma(N) \;=\; \bigcup_i \gamma_i F$$

It is useful that $\Gamma(N)$ is exactly the kernel of the group homomorphism

$$SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/N) \qquad \text{by} \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \to \begin{pmatrix} a \bmod N & b \bmod N \\ c \bmod N & d \bmod N \end{pmatrix}$$

so is *normal* in $\Gamma(1)$.

For the important special choice [6]

$$\Gamma_\theta \;=\; \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \;:\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod 2 \text{ or } \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \bmod 2\}$$

$$= \quad \Gamma(2) \quad \cup \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \Gamma(2)$$

the coset-representative oriented choice of fundamental domain can be adjusted to prove the corollary that Γ_θ is generated by $z \to -1/z$ and $z \to z+2$, as below.

[5.0.1] Remark: The following assertion holds without assuming p is prime, but all we need at the moment is $p = 2$, in any case. Further, the surjectivity of $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/2)$ is easy to observe directly, since, for example, the elements

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$$

surject to $SL_2(\mathbb{Z}/2)$.

[5.0.2] Claim: For p prime, the natural map

$$SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/p) \quad \text{ is surjective}$$

Proof: Let q be the quotient map $\mathbb{Z} \to \mathbb{Z}/p$. First, given u, v not both 0 in $\mathbb{Z}/p$, we will find *relatively prime* c, d in $SL_2(\mathbb{Z})$ such that $qc = u$ and $qd = v$.

For $v \notin p\mathbb{Z}$, there is $0 \neq d \in R$ such that $qd = v$. Consider the conditions on $c \in R$

$$c = u \bmod p \qquad \text{and} \qquad c = 1 \bmod d$$

As $d \notin p\mathbb{Z}$, by the maximality of the ideal $p\mathbb{Z}$ there are $x \in \mathbb{Z}$ and $pm \in p\mathbb{Z}$ such that $xd + pm = 1$. Let $c = xdu + pm$. From $xd + pm = 1$, $xd = 1 \bmod pm$ and $pm = 1 \bmod d$, so this expression for c satisfies the two congruences conditions. In particular, $qc = u$, and since $c = 1 \bmod d$ it must be that $\gcd(c, d) = 1$.

---

[5] Since $\mathfrak{H} = \bigcup_{\gamma \in \Gamma(1)} \gamma \overline{F}$, for representatives γ_i with $\Gamma(1) = \bigcup_i \Gamma(N)\gamma_i$,

$$\mathfrak{H} \;=\; \bigcup_{\gamma \in \Gamma(1)} \gamma\overline{F} \;=\; \bigcup_{\gamma \in \bigcup_i \Gamma(N)\gamma_i} \gamma\overline{F} \;=\; \bigcup_{\gamma \in \Gamma(N)} \bigcup_i \gamma\gamma_i\overline{F} \;=\; \bigcup_{\gamma \in \Gamma(N)} \gamma\Big(\bigcup_i \gamma_i\overline{F}\Big)$$

[6] This subgroup Γ_θ is important because it appears in sums-of-squares problems, the simplest application of *theta series* to seemingly elementary number-theory problems.

For $v = 0$ in $\mathbb{Z}/p$, necessarily $u \neq 0$, and we reverse the roles of c, d in the previous paragraph.

Thus, there are coprime c, d in $\mathbb{Z}$ whose images mod p are u, v. For integers s, t there exist a, b such that $\gcd(s,t) = as - bt$. The coprimality of c, d implies that there are a, b in R such that $ad - bc = 1$. That is, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} * & * \\ u & v \end{pmatrix} \mod p$$

Further adjustment to accommodate the *upper* row is more straightforward: Given $\begin{pmatrix} r & s \\ u & v \end{pmatrix}$ in $SL_2(\mathbb{Z}/p)$, and letting $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ also denote its image in $SL_2(\mathbb{Z}/p)$,

$$\begin{pmatrix} r & s \\ u & v \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} r & s \\ u & v \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} r & s \\ u & v \end{pmatrix} \begin{pmatrix} v & -b \\ -u & a \end{pmatrix} = \begin{pmatrix} rv - su & * \\ uv - vu & * \end{pmatrix} = \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix}$$

The right-hand side is in $SL_2(\mathbb{Z}/p)$, so, in fact, it must be of the form $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$, and

$$\begin{pmatrix} r & s \\ u & v \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod p$$

So

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} r & s \\ u & v \end{pmatrix} \mod p$$

giving the surjectivity. ///

[5.0.3] Claim: $\#SL_2(\mathbb{Z}/p) = (p^2 - 1)p$ for prime p.

Proof: First, count $GL_2(\mathbb{Z}/p)$. This is the number of ordered bases for the vector space $(\mathbb{Z}/p)^2$ over $\mathbb{Z}/p$, since an element of $GL_2(\mathbb{Z}/p)$ sends one basis to another, is transitive on ordered bases, and $g \in GL_2(\mathbb{Z}/p)$ fixes a basis v_1, v_2 only for $g = 1_2$.

The first basis element v_1 can be any non-zero vector in $(\mathbb{Z}/p)^2$, giving $p^2 - 1$ choices. For each such choice, the second basis element can be anything not on the $\mathbb{Z}/p$-line spanned by v_1, giving $p^2 - p$ choices. Thus, $\#GL_2(\mathbb{Z}/p) = (p^2 - 1)(p^2 - p)$.

The determinant map surjects $GL_2(\mathbb{Z}/p) \to (\mathbb{Z}/p)^\times$, and has kernel $SL_2(\mathbb{Z})$, so the *index* of $SL_2(\mathbb{Z}/p)$ is $\#(\mathbb{Z}/p)^\times = p - 1$, and the cardinality is as claimed. ///

[5.0.4] Corollary: $\Gamma(2)$ has six coset representatives in $\Gamma(1)$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$$

Proof: The index is $(2^2 - 1)2 = 6$. The six listed matrices are in $SL_2(\mathbb{Z})$ and are distinct mod 2. ///

[5.0.5] Corollary: Γ_θ has three coset representatives in $\Gamma(1)$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Proof: The index is 3, since Γ_θ is index 2 above $\Gamma(2)$. The three listed matrices are in $SL_2(\mathbb{Z})$ and are not only distinct mod 2 but also do not differ mod $\Gamma(2)$ merely by multiplication by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. ///

[5.0.6] Corollary: A fundamental domain for Γ_θ is

$$F_\theta \;=\; \{z \in \mathfrak{H} \;:\; |z| \geq 1 \text{ and } |\mathrm{Re}(z)| \leq 1\}$$

Proof: With standard fundamental domain

$$F \;=\; \{z \in \mathfrak{H} \;:\; |z| \geq 1 \text{ and } |\mathrm{Re}(z) \leq \tfrac{1}{2}\}$$

for $\Gamma(1)$, the coset representatives for Γ_θ in $\Gamma(1)$ give a fundamental domain

$$F' \;=\; F \cup \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} F \cup \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} F$$

for Γ_θ. *[... iou ...]* pictures! We will symmetrize this into a more easily-describable form. With hindsight, we replace

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

by

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

The point is that $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} F$ is understandable as a translate of the inverted F.

Move the *right half* of $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} F \cup \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} F$ left by $z \to z - 2$, so that the two halves are symmetric about the imaginary axis. This produces the region claimed in the theorem. ///

[5.0.7] Corollary: Inversion $z \to -1/z$ and translation $z \to z + 2$ generate Γ_θ.

Proof: Given $z \in \mathfrak{H}$, translate z by $2\mathbb{Z}$ to $|\mathrm{Re}(z)| \leq 1$. If $|z| \geq 1$, stop. If not, invert, and then translate back to $|\mathrm{Re}(z)| \leq 1$. This produces a sequence of points $z_1, z_2, \ldots$ with

$$\mathrm{Im}(z_1) \;<\; \mathrm{Im}(z_2) \;<\; \ldots$$

As earlier, $\mathrm{Im}(z_n)$ is of the form $\mathrm{Im}(z)/|cz + d|^2$, and any such sequence must be finite. That is, inversion and translation by $1\mathbb{Z}$ eventually put z into the fundamental domain for Γ_θ.

Given $\gamma \in \Gamma_\theta$, choose z in the interior of the fundamental region, and let δ be a composition of inversions and translations by $2\mathbb{Z}$ so that $\delta^{-1}\gamma z$ is back in the fundamental domain. Then $\delta^{-1}\gamma = \pm 1_2$, so $\gamma = \pm\delta$. Since the inversion squares to -1_2, $\gamma \in \Gamma_\theta$. ///

---

6. *Theta series are modular forms*

Generation of Γ_θ by inversion $z \to -1/z$ and translation $z \to z + 2$ allows an easy proof that the simplest *theta series*

$$\theta_{2k}(z) \;=\; \sum_{m_1,\ldots,m_{2k}} e^{\pi i(m_1^2 + \ldots + m_{2k}^2)z}$$

are *modular forms* for Γ_θ. Visibly,

$$\text{coefficient of } e^{\pi i n z} \text{ in } \theta_{2k}(2k) \quad = \quad \text{number of ways to express } n = m_1^2 + \ldots + m_{2k}^2$$

Thus, behavior of the Fourier coefficients of $\theta_{2k}(z)$ bears on *sums of squares* problems, and the fact that these theta series are *modular forms*, rather than purely combinatorial objects, is the key mechanism.

For now, for simplicity, only consider $2k \in 8\mathbb{Z}$. The more general case will be treated later. [7]

[6.0.1] Claim: $\theta_{8k}(z)$ is a modular form of weight $4k$ for Γ_θ. That is, for every $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_\theta$

$$\theta_{8k}(z+2) \;=\; (cz+d)^{4k} \cdot \theta_{2k}(z)$$

Proof: Let

$$j(\gamma, z) \;=\; cz + d \qquad\qquad (\text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix})$$

The *cocycle property*

$$j(\gamma\delta, z) \;=\; j(\gamma, \delta z) \cdot j(\delta, z)$$

is directly verifiable:

$$j(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} A & B \\ C & D \end{pmatrix} z) \;\cdot\; j(\begin{pmatrix} A & B \\ C & D \end{pmatrix}, z) \;=\; (c\frac{Az+B}{Cz+d} + d) \cdot (Cz+D)$$

$$=\; c(Az+B) + d(Cz+D) \;=\; (cA+dC)z + (cB+dD)$$

while

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \;=\; \begin{pmatrix} * & * \\ cA+dC & cB+dD \end{pmatrix}$$

Thus, an induction on length of expression in terms of $z \to -1/z$ and $z \to z+2$ would suffice: for $\gamma, \delta \in \Gamma_\theta$, by induction

$$j(\gamma\delta, z)^{-8k} \theta_{8k}(\gamma\delta z) \;=\; j(\delta, z)^{-8k} j(\gamma, \delta z)^{-8k} \theta_{8k}(\gamma(\delta z)) \;=\; j(\delta, z)^{-8k} \theta_{8k}(\delta z) \;=\; \theta_{8k}(z)$$

Thus, it suffices to prove $\theta_{8k}(z+2) = (cz+d)^{4k} \cdot \theta_{2k}(z)$ merely for the *generators*. For $z \to z+2$, each summand in θ_{8k} is invariant. The inversion $z \to -1/z$ is treated via *Poisson summation*: for $z = iy$ with $y > 0$, the Gaussian $v \to e^{-\pi|v|^2 y}$ on $\mathbb{R}^{8k}$ has Fourier transform $v \to y^{-4k} e^{-\pi|v|^2/y}$. By Poisson summation,

$$\theta_{8k}(iy) \;=\; \sum_{m_1,\dots,m_{8k}} e^{-\pi(m_1^2+\dots+m_{8k}^2)y} \;=\; \frac{1}{y^{4k}} \sum_{m_1,\dots,m_{8k}} e^{-\pi(m_1^2+\dots+m_{8k}^2)/y} \;=\; \frac{1}{y^{4k}} \theta_{8k}(\frac{-1}{iy})$$

This gives equality of two holomorphic functions, $\theta_{8k}(z)$ and $z^{-4k}\theta_{8k}(\frac{-1}{z})$, on the positive imaginary axis, so by the identity principle the equality holds throughout $\mathfrak{H}$. This proves the modular form condition for $z \to -1/z$. ///

---

[7] Further, *harmonic theta series*

$$\theta_{2k,P}(z) \;=\; \sum_{m_1,\dots,m_{2k}} P(m_1,\dots,m_{2k})\, e^{\pi i(m_1^2+\dots+m_{2k}^2)z}$$

with degree d homogeneous, *harmonic* polynomials P in $2k$ variables are modular forms for Γ_θ of weight $k+d$. Among other applications, these harmonic theta series are used in study of *equidistribution* of points on spheres. As usual, a polynomial P in n variables is *harmonic* when it is annihilated by the Euclidean Laplacian

$$\Delta \;=\; \frac{\partial^2}{\partial x_1^2} + \dots + \frac{\partial^2}{\partial x_n^2}$$