Last revised 3:40 p.m. September 28, 2024

Analysis on arithmetic quotients

Bill Casselman
University of British Columbia
cass@math.ubc.ca

Chapter II. Arithmetic groups

Bill Casselman
University of British Columbia
cass@math.ubc.ca

This chapter is a brief introduction to geometric properties of arithmetic subgroups of $\mathrm{SL}_2(\mathbb{R})$.

Let

$$\begin{aligned} G &= \mathrm{SL}_2(\mathbb{R}) \\ K &= \mathrm{SO}(2) \\ \mathcal{H} &= \text{ the upper half plane, which may be identified with } G/K \\ P &= \text{ subgroup of upper triangular matrices in } G \\ N &= \text{ subgroup of unipotent matrices in } P. \end{aligned}$$

Contents

1. Proper discrete subgroups

A subgroup Γ of G is said to be **discrete** if there exists a neighbourhood U of the identity in G containing no element of Γ other than I. The most important examples are $\mathrm{SL}_2(\mathbb{Z})$ and its congruence subgroups, which I shall examine in detail later on. But it is valuable not to restrict consideration to these alone, since the role of geometry (as opposed to number theory) becomes more apparent without such restrictions.

ELEMENTARY PROPERTIES.

II.1.1. Lemma. *Suppose Γ to be a discrete subgroup of G. Then*

(a) there exists a neighbourhood U of 1 in G such that $\gamma(U) \cap U \neq \emptyset$ only if $\gamma = I$;
(b) any subset of Γ is closed in G;
(c) if U is any compact subset of G then the set of γ in Γ with $\gamma(U) \cap U \neq \emptyset$ is finite.

Proof. For (a), let U_* be a neighbourhood of 1 not containing any other element of Γ, and let U be such that $U \cdot U \subset U_*$, $U = U^{-1}$.

For (b), let Θ be any subset of Γ, x in its complement, U as in (a). The neighbourhood xU of x contains at most one element of Γ. There then exists a neighbourhood of x contained in xU and not containing any element of Θ.

For (c), let $V = U \cdot U^{-1}$, which is compact. The intersection of Γ with V is compact, and covered by disjoint neighbourhoods of each of its points. This intersection must therefore be finite. ∎

II.1.2. Lemma. *The group Γ acts discretely on $\mathcal{H}$.*

That is to say, given z on $\mathcal{H}$ there exists a neighbourhood V of z such that whenever γV meets V (for γ in Γ) we have $\gamma(z) = z$.

The set of such γ is contained in the isotropy subgroup of z, which is a conjugate of K. Because of (c) above, this set is finite.

Proof. Identify $\mathcal{H}$ with G/K. Choose any compact neighbourhood U of z, and let U_* be its inverse image in G, which is also compact. By Lemma II.1.1, the set of γ such that $\gamma U_* \cup U_* \neq \emptyset$ is finite, and this is precisely the set of γ such that γU meets U. Let Θ be the subset of γ in this set which do not fix z, and choose the neighbourhood V to be contained in U and small enough so that V and the γV for all γ in Θ are disjoint.

CUSPS. For any q in $\mathbb{R} \cup \infty$, let P_q be the stabilizer in $\mathrm{SL}_2(\mathbb{R})$ of q, with unipotent radical N_q. Thus $P = P_\infty$. The intersection $\Gamma \cap N_q$ is discrete in N_q, hence either (a) trivial or (b) generated by a single element. In the second case the quotient $\Gamma \cap N_q \backslash N_q$ is compact and q is said to be a **cusp** of Γ.

II.1.3. Lemma. *If q is a cusp of Γ then the intersection $\Gamma \cap P_q$ is either (a) the same as $\Gamma \cap N_q$ or (b) the slightly larger group*

$$\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot (\Gamma \cap N_q) \, .$$

The second case occurs if and only if $\pm I$ is in Γ.

Proof. The image of $\Gamma \cap P_q$ modulo $\Gamma \cap N_q$ must normalize the lattice $\Gamma \cap N_q$ in N_q.

For each $Y > 0$ define the region in $\mathcal{H}$

$$\mathcal{H}_Y := \left\{ z \in \mathcal{H} \,\middle|\, \mathrm{IM}(z) > Y \right\} \, .$$

For each g in $\mathrm{SL}_2(\mathbb{R})$ let $c(g)$ be the c appearing in the matrix expression

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \, .$$

The function $|c(\gamma)|$ is a function on $(\Gamma \cap P)\backslash\Gamma/(\Gamma \cap P)$.

The following Proposition states a basic property of cusps in several slightly different ways.

II.1.4. Proposition. *Suppose ∞ to be a cusp of Γ. Then*

(a) *for any z in $\mathcal{H}$, there exist only a finite number of points in its Γ-orbit modulo $\Gamma \cap P$ with greater height;*

(b) *there exists a minimum value $c_\Gamma \neq 0$ of $|c(\gamma)|$ as γ ranges over $\Gamma - \Gamma \cap P$;*

(c) *for $Y \geq 1/c_\Gamma$*

$$\{\gamma \in \Gamma \,|\, \gamma(\mathcal{H}_Y) \cap \mathcal{H}_Y \neq \emptyset\} = \Gamma \cap P \, .$$

(d) *in these circumstances the canonical map*

$$(\Gamma \cap P)\backslash\mathcal{H}_Y \longrightarrow \Gamma\backslash\mathcal{H}$$

is an embedding.

Proof. If $\Gamma \subset P$, there is nothing to be proved, so assume that this is not the case. The possible values of $c(\gamma)$ then include non-zero numbers. Conjugating by an element of $\mathrm{SL}_2(\mathbb{R})$, I may assume that $\Gamma \cap N = N(\mathbb{Z})$.

(a) Suppose z in $\mathcal{H}$. Choose a non-Euclidean disk U centred at z such that the γU are all either disjoint or equal. If there exist γ with $\gamma(z)$ of arbitrary height, then according to Lemma I.4.3 the widths of the $\gamma(U)$ increase in proportion. But if two are high enough, some $\gamma(U)$ and $\gamma(U) + n$ will overlap, a contradiction.

(b) In a moment, we'll need:

II.1.5. Proposition. *If $c \neq 0$ then the image of $\mathcal{H}_Y$ under the element*

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

in $\mathrm{SL}_2(\mathbb{R})$ is the open disc of height $1/(c^2Y)$ just touching $\mathbb{R}$ at the rational point a/c.

Proof. Since $g(\infty) = a/c$, according to Proposition I.2.1 the image of the horizontal line $y = Y$ is a circle. Since g takes $\mathbb{P}^1(\mathbb{R})$ to itself, it must be tangent to $\mathbb{R}$ at $g(\infty) = a/c$, say with top at $a/c + iy$. According to Corollary I.3.1, the imaginary part of $g(x + iY)$ is

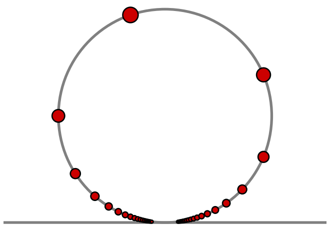
$$\frac{Y}{|c(x+iY)+d|^2} = \frac{Y}{(cx+d)^2 + c^2Y^2} \, .$$

It achieves the maximum $1/(c^2Y)$ when $cx + d = 0$.

Now suppose

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

with $c \neq 0$, and suppose $z = x + iy$. What can we say about the set of $g\nu(z)$ as ν traverses $\Gamma \cap N$? According to the Lemma, they will all lie on the circle touching the real line at a/c, and of height $1/cy^2$. They will be distributed discretely, like this:



The highest will not oftenlie at the top of this circle, but it will not be far away.

II.1.6. Lemma. *Suppose*

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} ,$$

in $\mathrm{SL}_2(\mathbb{R})$, $z = x + iy$ in $\mathcal{H}$. The point z_* among the $g(z+n)$ of maximal height satisfies

$$\frac{y}{c^2(1/4 + y^2)} \leq \mathrm{IM}(z_*) \leq \frac{1}{c^2y} \, .$$

Proof. The height of $\gamma(z + n)$ is

$$\frac{y}{(cx + cn + d)^2 + c^2y^2} \, .$$

We can choose n such that $|cx + cn + d| \leq c/2$, and then the denominator is at most $c^2/4 + c^2y^2$.

If $c(\gamma)$ is not bounded from below, this is unbounded from above. That contradicts (a).

Let c_Γ be the greatest lower bound of these values of $c(\gamma)$. It remains to be seen that there exists γ with $|c(\gamma)| = c_\Gamma$. This will follow from:

II.1.7. Lemma. *For each $C > 0$ there exist only a finite number of possible values of $c(\gamma)$ less than C.*

Proof. Choose $Y \geq 1/c_\Gamma$. The sets $\gamma\mathcal{H}_Y$ are either disjoint or equal, and they lie in the region $\mathrm{IM}(z) < 1/c_\Gamma^2 Y$. If $c(\gamma) \leq C$ then the diameter of $\gamma\mathcal{H}_Y$ is at least $1/C^2 Y$ annd its (Euclidean) area is at least $4\pi/C^4 Y^2$. It is equivalent modulo translations $z \mapsto z + n$ to one contained in the region $|x| \leq 1$. There can be only a finite set of such disks, since the area of the region $|x| \leq 1$, $y \leq 1/c_\Gamma^2 Y$ is finite. ▯

(c) If γ lies in Γ and $c = c(\gamma) \neq 0$, the image of $\mathcal{H}_Y$ is a circle of height $1/c^2 Y$. If $Y \geq 1/c_\Gamma$, then $1/c^2 Y \leq 1/c_\Gamma^2 Y \leq Y$, so $\mathcal{H}_Y$ and $\gamma\mathcal{H}_Y$ are disjoint.

Item **(d)** is now immediate. ▯▯

Suppose q to be a cusp, and that $g(q) = \infty$. The element g may be chosen so that $g(\Gamma \cap N_q)g^{-1} = N \cap \mathrm{SL}_2(\mathbb{Z})$. I will call the pull-back of the coordinates x, y along this g **parabolic coordinates** x_q, y_q in the neighbourhood of q. They are well defined up to a shift of x. I define a **parabolic domain** in $\mathcal{H}$ to be the pull back of some $\mathcal{H}_Y$, or in other words a region $y_q \geq Y$. According to Proposition II.1.5 this is either $\mathcal{H}_Y$ or a circle tangent to the real numbers.

Since every cusp can be transformed to ∞ by some element of $\mathrm{SL}_2(\mathbb{R})$:

II.1.8. Proposition. *If q is a cusp of Γ, then there exists a parabolic domain D in the neighbourhood of q such that*

$$\{\gamma \in \Gamma \mid \gamma(D) \cap D \neq \emptyset\} = \Gamma \cap P_q\,.$$

In other words, there exists in the neighbourhood of q a parabolic domain D with the property that the canonical map from $\Gamma \cap P_q\backslash D$ to $\Gamma\backslash\mathcal{H}$ is an injection. In this circumstance I shall call the image a **parabolic domain of** $\Gamma\backslash\mathcal{H}$ (in the neighbourhood of the cusp q). For example, if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ then, by the remark made just after Proposition II.1.4, any $Y \geq 1$ will do.

PROPER SUBGROUPS. I'll call a discrete subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$ **proper** if the quotient $\Gamma\backslash\mathcal{H}$ is a finite union of parabolic domains and a compact set. If Γ is proper, so is any subgroup of finite index, and conversely.

- *From now on in this chapter I'll assume Γ to be proper.*

The closure of any parabolic domain in $\mathbb{P}^1(\mathbb{R})$ is a single cusp fixed by a conjugate of P_∞ in $\mathrm{SL}_2(\mathbb{R})$. Let $\mathcal{H}^*$ be the union of $\mathcal{H}$ and all the cusps of Γ. One can assign it a topology—a set is open if and only if its intersection with every parabolic domain is open. Then:

II.1.9. Proposition. *A proper discrete subgroup Γ acts discretely on $\mathcal{H}^*$, and the quotient $\Gamma\backslash\mathcal{H}^*$ is compact.*

The measure

$$\frac{dx\,dy}{y^2}$$

is invariant with respect to the action of G. The area of a parabolic domain $\Gamma\backslash\mathcal{H}_Y$ is therefore

$$\int_0^1 dx \int_Y^\infty \frac{dy}{y^2} = \frac{1}{Y} < \infty\,,$$

and we deduce:

II.1.10. Proposition. *The area of $\Gamma\backslash\mathcal{H}$ is finite.*

Remark. It was proved in [Siegel:1971] that, conversely, if Γ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ with the area of $\Gamma\backslash\mathcal{H}$ finite, then Γ is proper. [Borel:1997] includes a proof, too.

A parabolic subgroup P is called Γ-rational if its fixed point on $\mathbb{P}^1(\mathbb{R})$ is a cusp of Γ. Under the assumption that Γ is proper, the number of Γ-conjugacy classes of Γ-rational parabolic subgroups is finite.

NORMS. On $G = \mathrm{SL}_2(\mathbb{R})$ we have the 'algebraic' norm

$$\sup_{\|v\|=1}\|g(v)\| \quad (v \in \mathbb{R}^2)\,,$$

which induces a norm on $\mathcal{H}$. Explicitly, as we have seen in Proposition I.6.3, this is equivalent to

(II.1.11)
$$|\!|\!|z|\!|\!| = \frac{x^2+(y+1)^2}{y}\,.$$

This in turn determines a norm on $\Gamma\backslash\mathcal{H}$:

$$|\!|\!|z|\!|\!|_{\Gamma\backslash\mathcal{H}} = \inf_{\gamma\in\Gamma}|\!|\!|\gamma(z)|\!|\!|\,.$$

II.1.12. Lemma. *If Ω is a compact subset of $\mathcal{H}$ then $|\!|\!|x|\!|\!|_{\Gamma\backslash\mathcal{H}} \asymp |\!|\!|x|\!|\!|$ on Ω.*

In fact, this remains valid for any discrete Γ.

Proof. Let M be the maximum value of $|\!|\!|x|\!|\!|$ on Ω, and let Ω_* be the subset of $\mathcal{H}$ on which $|\!|\!|x|\!|\!| \le M$. It is compact. By (c) of Proposition II.1.4 the set Ξ of γ such that $\gamma\Omega_* \cap \Omega_* \neq \emptyset$ is finite. Then for x in Ω_*

$$\inf_{\gamma\in\Gamma}|\!|\!|\gamma(x)|\!|\!| = \inf_{\gamma\in\Xi}|\!|\!|\gamma(x)|\!|\!|\,.$$

But then

$$\frac{\inf_{\gamma\in\Xi}|\!|\!|\gamma(x)|\!|\!|}{|\!|\!|x|\!|\!|}$$

is an invertible continuous function on Ω_*.

The following is elementary:

II.1.13. Lemma. *Say $X, Y > 0$. In the region $|x| \le X$, $y \ge Y$*

$$|\!|\!|z|\!|\!| \asymp y$$

and in the region $|x| \le X$, $y \le Y$

$$|\!|\!|z|\!|\!| \asymp 1/y\,.$$

More explicitly, we have

(II.1.14)
$$y \le |\!|\!|z|\!|\!| \le y\left(1 + \frac{X^2+Y+1}{Y^2}\right)$$

in the first case and

(II.1.15)
$$\frac{x^2+1}{y} \le |\!|\!|z|\!|\!| \le \frac{x^2+1}{y}\,(Y+1)^2$$

in the second.

A region $|x_q| \le X$, $y_q \ge Y$ is called a **Siegel domain** for the cusp q. Because Γ is proper, the quotient $\Gamma\backslash\mathcal{H}$ is covered by a finite number of these.

Consequently, in any Siegel domain for the cusp at ∞ we have $|\!|\!|z|\!|\!| \asymp \mathrm{IM}(z)$.

II.1.16. Proposition. *Suppose z to be in $\mathcal{H}$. In any Siegel domain there exist only a finite number of points in the Γ-orbit of z.*

Proof. Suppose U to be a compact neighbourhood of z such that The sets γU are either disjoint or identical. Suppose C_2 to be a bound on the height of points in the Γ-orbit of z, and $C_1 > 0$ to be a lower bound on $\mathfrak{S}$, and A a bound on the width of $\mathfrak{S}$. Hence every point of the orbit of z that lies in $\mathfrak{S}$ lies inside the region $|x| \le A$, $C_1 \le y \le C_2$. By Lemma I.4.3, the area of the sets γU is bounded from below, so there can be only a finite number of the sets γU meeting the rectangle.

II.1.17. Theorem. *On any Siegel domain $\mathfrak{S}$*

$$|\!|\!|z|\!|\!|_{\Gamma\backslash\mathcal{H}} \asymp |\!|\!|z|\!|\!| \,.$$

Proof. We have

$$\|g\|^{-1} \cdot \|x\| \le \|v\| \le \|g\| \cdot \|x\| \,,$$

so transforming a region by g in G changes the norm by only a constant factor. Therefore we may assume that the cusp involved is ∞. I may also assume that $\Gamma \cap N = N(\mathbb{Z})$. Because of Lemma II.1.12, I may assume that Y is large enough so that $\gamma\mathcal{H}_Y \cap \mathcal{H}_Y = \emptyset$ unless γ is in $\Gamma \cap P$.

The group Γ may be partitioned into $\Gamma \cap P$ and its complement. So $|\!|\!|z|\!|\!|_{\Gamma\backslash G\mathcal{H}}$ is the minimum of

$$\inf_{\Gamma\cap P} |\!|\!|\gamma(z)|\!|\!|, \quad \inf_{\Gamma-\Gamma\cap P} |\!|\!|\gamma(z)|\!|\!| \,.$$

For z in the region $|x| \le X$, $y \ge Y$, $|\!|\!|z|\!|\!| \asymp \textsc{im}(z)$ and for γ in $\Gamma \cap P$ we have $\textsc{im}(\gamma(z)) = \textsc{im}(z)$. Therefore, up to constants

$$\inf_{\Gamma\cap P} |\!|\!|\gamma(x+iy)|\!|\!| = y \,.$$

This takes care of the first case.

If γ is not in $\Gamma \cap P$ then $z_* = \gamma(z)$ lies in the region $\textsc{im}(z_*) \le 1/c_\Gamma^2 y$, and we may shift it to lie in the region $|x| \le 1/2$, in which $|\!|\!|x_* + iy_*|\!|\!| \asymp 1/y_*$. Lemma II.1.6 tells us that

$$\frac{y}{c_\Gamma^2(1/4+y^2)} \le \textsc{im}(z_*) \le \frac{1}{c_\Gamma^2 y} \,,$$

and then

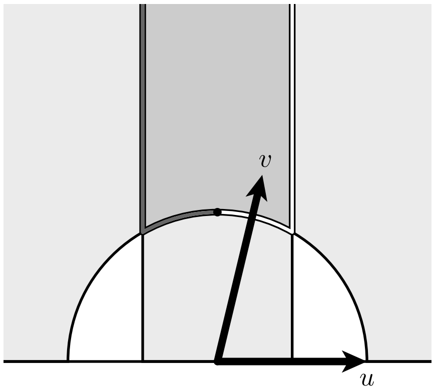
$$\inf_{\Gamma-\Gamma\cap P} |\!|\!|\gamma(z)|\!|\!| \asymp \frac{1}{\textsc{im}(z_*)} \asymp c_\Gamma^2 \textsc{im}(z) \,.$$

2. The group SL(2,Z)

In these notes a **lattice** will be a copy of $\mathbb{Z}^2$ in $\mathbb{C}$, given the Euclidean structure determined by the complex norm. The dot product in $\mathbb{C}$ is calculated as $u \bullet v = \textsc{re}(u\overline{v})$.

A **reduced basis** of a lattice is a pair u, v satisfying these conditions:

(a) the length of u is less than or equal to that of v;
(b) the perpendicular projection of v onto the real line through u lies in the closed interval $[-u/2, u/2]$;
(c) the pair u, v is positively oriented in the sense that v/u has positive imaginary component.



A strictly reduced pair

It will called **strictly reduced** if

(b′) The perpendicular projection of v onto the real line through u lies in the half open interval $[-u/2, u/2)$, and if $|u| = |v|$ it lies in $[-u/2, 0]$.

I recall that the perpendicular projection of v onto u is the complex number

$$\left(\frac{u \bullet v}{|u|^2}\right) u$$

so that condition (b) means $-1/2 \le u \bullet v/|u|^2 \le 1/2$.

The significance of this is:

II.2.1. Proposition. *If u, v is a reduced basis, then $|u|$ is minimal among the lengths of vectors in the lattice they span.*

Proof. Suppose a, b integers. Then

$$|au + bv|^2 = a^2|u|^2 + 2ab(u \bullet v) + b^2|v|^2 \ge (a^2 + b^2)|u|^2 - 2|ab|(u \bullet v) \ge (a^2 + b^2 - |a|\,|b|)|u|^2 \,.$$

But $a^2 - ab + b^2$ is the norm form of the lattice spanned by 1 and a cube root of unity, which has a minimum norm of 1.

Conversely, suppose u to be a vector of least possible length in this lattice. Since u is primitive, we may find v such that u and v form a basis of the lattice with $\textsc{im}(v/u) > 0$. Hence:

II.2.2. Proposition. *Every lattice in $\mathbb{C}$ possesses a reduced basis.*

Lattices will usually be specified by a basis, that is to say a pair of complex numbers u, v such that v/u is not a real number. By changing v to $-v$ if necessary we may assume this basis to be positive in the sense that $\textsc{im}(v/u) > 0$. There is a simple algorithm originally due, I believe, to Lagrange that finds a reduced basis explicitly, starting from a given positive basis.

(0) (Initial signed swap) If $|u| > |v|$ then replace u and v by $-v$ and u.
(1) (Translation) At this point, $|v| \ge |u|$. If necessary, replace v by $v - nu$ so as get the projection of v onto the line through u between $\pm u/2$. Explicitly, let $x = u \bullet v/|u|^2 + 1/2$ and $n = \lfloor x \rfloor$.
(2) (Signed swap) At this point the projection of v is correct. If $|u| > |v|$ then replace u and v by $-v$ and u and go to (1); otherwise stop.

Each positive basis u, v of a lattice in $\mathbb{C}$ determines the point $z = v/u$ in $\mathcal{H}$. The points of $\mathcal{H}$ in fact classify positive bases of lattice in $\mathbb{C}$ up to oriented similarity, that is up to multiplication by a non-zero complex number. Any quotient of $\mathbb{C}$ by an embedded copy of $\mathbb{Z}^2$ defines an elliptic curve, and the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathcal{H}$ parametrizes isomorphism classes of elliptic curves.

The algorithm above translates into one for points of $\mathcal{H}$:

(0) (Initial signed swap) If $|z| < 1$, replace it by $-1/z$;
(1) (Translation) Replace z by $z - n$ where $n = \lfloor \mathrm{RE}(z) + 1/2 \rfloor$, so that $n \le x + 1/2 < n + 1$.
(2) (Signed swap) If $|z| < 1$ then replace z by $-1/z$ and go to (1); otherwise stop.

II.2.3. Corollary. *The region*

$$|z| \ge 1, \quad -1/2 \le x \le 1/2$$

is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$.

The algorithm can be refined to produce a unique strictly reduced basis, and in effect to show that every z in $\mathcal{H}$ is equivalent to a unique strictly reduced pair.

Another way of describing the quotient $\Gamma\backslash\mathcal{H}$ is by saying that it parametrizes isomorphism classes of lattices up to oriented similarity, or equivalently isomorphism classes of elliptic curves, which are the quotients of $\mathbb{C}$ by embedded copies of $\mathbb{Z}^2$.

One consequence of the discussion so far is that if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ then the quotient $\Gamma\backslash\mathcal{H}$ has a single cusp. More precisely:

II.2.4. Proposition. *The cusps of* $\mathrm{SL}_2(\mathbb{Z})$ *in* $\mathbb{C}$ *are the rational points in* $\mathbb{R} \cup \infty$, *on which* $\mathrm{SL}_2(\mathbb{Z})$ *acts transitively.*

Proof. It is straightforward to see that any cusp has to be rational.

I describe an algorithm that shows how $\mathrm{SL}_2(\mathbb{Z})$ acts transitively. Given a pair of integers (c, d), the Euclidean algorithm keeps dividing c by d until there is no remainder, in which case the last divisor is the greatest common divisor. If we set $c_0 = c$, $d_0 = d$, the more precise version maintains a matrix M_n which starts out as $M_0 = I$ and at every step satisfies

$$M_n \begin{bmatrix} c_0 \\ d_0 \end{bmatrix} = \begin{bmatrix} c_n \\ d_n \end{bmatrix} .$$

In each step, we set

$$\begin{aligned} c_n &= q d_n + r \quad (0 \le r < |d_n|) \\ c_{n+1} &= d_n \\ d_{n+1} &= r \\ \begin{bmatrix} c_{n+1} \\ d_{n+1} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} \begin{bmatrix} c_n \\ d_n \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} M_n \begin{bmatrix} c_0 \\ d_0 \end{bmatrix} \\ M_{n+1} &= \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} M_n \end{aligned}$$

getting in the end, assuming the greatest common divisor to be 1, a matrix M with

$$M \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} .$$

The matrix M might not have determinant 1, but that can be easily corrected. ▯

3. Congruence groups

The **principal congruence subgroup of level** N is

$$\Gamma(N) := \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) \,\middle|\, \gamma \equiv I \pmod{N} \right\} .$$

Thus $\Gamma(N)$ is the kernel of the canonical homomorphism from $\mathrm{SL}_2(\mathbb{Z})$ to $\mathrm{SL}_2(\mathbb{Z}/N)$.

II.3.1. Proposition. *The canonical homomorphism from* $\mathrm{SL}_2(\mathbb{Z})$ *to* $\mathrm{SL}_2(\mathbb{Z}/N)$ *is surjective.*

Proof. One possible proof uses the Bruhat decomposition and is valid for SL_2 over any algebraic number field, but I offer here another. It comes down to showing that given a matrix η in $M_2(\mathbb{Z})$ with

$$\det(\eta) \equiv 1 \pmod{N}$$

there exists γ in $\mathrm{SL}_2(\mathbb{Z})$ with $\gamma \equiv \eta$. Transforming η on left and right by elements of $\mathrm{SL}_2(\mathbb{Z})$, we may assume η to be diagonal (elementary divisor theorem). Suppose it is

$$\eta = \begin{bmatrix} a_* & 0 \\ 0 & d_* \end{bmatrix}$$

with $a_* d_* \equiv 1 \pmod{N}$. Choose a and d in $\mathbb{Z}$ which are inverse modulo N^2, congruent modulo N to a_*, d_* respectively. Thus

$$ad - bN^2 = 1$$

for some integer b, which implies that

$$\begin{bmatrix} a & N \\ bN & d \end{bmatrix}$$

is the matrix we are looking for. ∎

II.3.2. Corollary. *The sequence*

$$1 \longrightarrow \Gamma(N) \longrightarrow \mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/N) \longrightarrow 1$$

is exact.

By the Chinese Remainder Theorem, the ring $\mathbb{Z}/N$ is the direct product of rings $\mathbb{Z}/p^{n_p}$ if $N = \prod p^{n_p}$. The sequence

$$1 \longrightarrow \Gamma(p)/\Gamma(p^n) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/p^n) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/p) \longrightarrow 1$$

is exact, so:

II.3.3. Corollary. *If* $N = \prod p^{n_p}$ *with* $n_p > 0$ *then the index of* $\Gamma(N)$ *in* $\mathrm{SL}_2(\mathbb{Z})$ *is* $N^3 \prod_p \left(1 - \dfrac{1}{p^2}\right)$.

The group $\mathrm{SL}_2(\mathbb{Z})$ acts transitively on the relatively prime pairs (c, d) in $\mathbb{Z}^2$. Two such points are equivalent under $\Gamma(N)$ if and only if they are congruent modulo N. If a/c and a_*/c_* are two rational numbers expressed in reduced form then they are equivalent under $\Gamma(N)$ if and only if $a \equiv a_*$, $c \equiv c_* \bmod N$. Hence:

II.3.4. Proposition. *The correspondence* $a/c \mapsto \begin{bmatrix} a \\ c \end{bmatrix}$ *is a bijective* $\mathrm{SL}_2(\mathbb{Z})$*-covariant correspondence between the cusps of* $\Gamma(N)$ *and the points of* $\mathbb{P}^1(\mathbb{Z}/N)$.

The quotient $\Gamma(N)\backslash\mathcal{H}$ parametrizes isomorphism classes of elliptic curves E together with an isomorphism of the N-torsion of E with $(\mathbb{Z}/N)^2$ (often called a **level structure**).

4. References

1. Armand Borel, **Automorphic forms on SL(2,R)**, Cambridge University Press, 1997.
2. Carl Ludwig Siegel, **Topics in complex function theory**, Volume II, Wiley, 1971.