Analysis on arithmetic quotients

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

Chapter II. Arithmetic groups

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

Let $G = \text{SL}_2(\mathbb{R})$, $K = \text{SO}(2)$ and let $\mathcal{H}$ be the upper half plane, which may be identified with $G/K$.

Contents

1. Proper discrete subgroups
2. The group $\text{SL}_2(\mathbb{Z})$
3. Congruence groups
4. References

1. Proper discrete subgroups

Let $\Gamma$ be a discrete subgroup of $G$. This means that there exists a neighbourhood $U$ of the identity in $G$ containing no element of $\Gamma$. The most important examples are $\text{SL}_2(\mathbb{Z})$ 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.

II.1.1. Lemma. Suppose $\Gamma$ to be a discrete subgroup of a locally compact group $G$. Then

(a) there exists a neighbourhood $U$ of 1 in $G$ such that $\gamma(U) \cap U \neq \emptyset$ only if $\gamma = 1$;

(b) any subset of $\Gamma$ is closed in $G$;

(c) if $U$ is any compact subset of $G$ then the set of $\gamma$ in $\Gamma$ 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 $\Gamma$, and let $U$ be such that $U \cdot U \subset U^*$, $U = U^{-1}$.

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

For (c), let $V = U \cdot U^{-1}$, which is compact. The intersection of $\Gamma$ 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 $\Gamma$ acts discretely on $\mathcal{H}$.

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

Proof. Identify $\mathcal{H}$ with $G/K$. Choose any compact neighbourhood $U$ of $z$, and let $U_\gamma$ be its inverse image in $G$, which is also compact. By Lemma II.1.1, the set of $\gamma$ such that $\gamma U_\gamma \cup U_\gamma \neq \emptyset$ is also finite. The set of $\gamma$ such that $\gamma U$ meets $U$ is finite. Let $\Theta$ be the subset of $\gamma$ 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 $\gamma V$ for all $\gamma$ in $\Theta$ are disjoint.

For any $q$ in $\mathbb{R} \cup \infty$, let $P_q$ be the stabilizer in $\text{SL}_2(\mathbb{R})$ of $q$, with unipotent radical $N_q$. 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 $\Gamma$. If $q$ is a cusp then there exists $g$ in $\text{SL}_2(\mathbb{R})$ with
Another way to put this: the domain of the number of points in the \( \Gamma \) canonical map from \( \Gamma \) to \( \mathbb{H} \) is not. Therefore if the second case were to occur, then for any \( y \) the width of \( \gamma \) grows with the height of \( \gamma(z) \). But if \( U_{\gamma(z)} \) is wide enough, its translations by elements of \( \Gamma \) must intersect eventually, a contradiction. In other words, the alternative (b) does not happen. We have thus proved:

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

\[
\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \cdot (\Gamma \cap N_q)
\]

**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 convenience, I will for a moment take the cusp \( q \) to be \( \infty \) and let \( P = P_{\infty} \). I shall assume a generator of \( \Gamma \cap N \) to be \( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \). These can always be made valid upon replacing \( \Gamma \) by a conjugate, if necessary.

Consider the images under various \( \gamma \in \Gamma \) of the parabolic domain \( \mathcal{H}_Y \). According to Proposition I.3.4 these images are either \( \mathcal{H}_Y \) itself or circles tangent to \( \mathbb{R} \), and there are two distinct possibilities for the second class: (a) these circles all lie below some fixed height or (b) the tops of these circles lie arbitrarily high. In the first case, there is a minimum to the possible values of \( c \) for elements \( \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \) of \( \Gamma \), while in the second there is not. Therefore if the second case were to occur, then for any \( z \) in \( \mathcal{H} \) the heights of elements in the orbit \( \Gamma z \) are also unbounded. Fix a non-Euclidean disk around \( I \) small enough so that such that all the \( \gamma U (\gamma \in \Gamma) \) are either disjoint or equal. According to Proposition I.3.4 the widths of \( U_{\gamma(z)} \) grows with the height of \( \gamma(z) \). But if \( U_{\gamma(z)} \) is wide enough, its translations by elements of \( \Gamma \cap N \) must intersect eventually, a contradiction.

In other words, the alternative (b) does not happen. We have thus proved:

**II.1.4. Proposition.** If \( q \) is a cusp of \( \Gamma \), then there exists a parabolic domain \( D \) in the neighbourhood of \( q \) with the property that \( \gamma D \cap D \neq \emptyset \) if and only if \( \gamma \) lies in \( P_q \).

In other words, there exists in the neighbourhood of \( q \) a parabolic domain \( D \) with the property that the canonical map from \( \Gamma \cap P_q \) to \( \Gamma \) is an injection. In this circumstance I shall call the image a parabolic domain of \( \Gamma \) in \( \mathcal{H} \) (in the neighbourhood of the cusp \( q \)).

Another way to put this:

**II.1.5. Proposition.** Assume \( \infty \) to be a cusp of \( \Gamma \). For a given \( z \), there exist modulo \( \Gamma \cap P_{\infty} \) only a finite number of points in the \( \Gamma \)-orbit of \( z \) with greater height.

And yet another:

**II.1.6. Proposition.** If \( \infty \) is a cusp of \( \Gamma \) then the non-zero matrix entries \( c \) are bounded away from 0.

I’ll call a discrete subgroup \( \Gamma \) of \( \text{SL}_2(\mathbb{R}) \) proper if it has this property:

**The quotient \( \Gamma \backslash \mathcal{H} \) is the union of a finite number of parabolic domains and a compact subset.**

From now on in this chapter I’ll assume \( \Gamma \) to be proper. If \( \Gamma \) is proper, so is any commensurable group.

The closure of any parabolic domain in \( \mathbb{P}^1(\mathbb{R}) \) is a single cusp fixed by a conjugate of \( P_{\infty} \) in \( \text{SL}_2(\mathbb{R}) \). Since the area of a parabolic domain \( \Gamma \backslash \mathcal{H}_Y \) is

\[
\int_0^1 dx \int_0^\infty dy \frac{dy}{y^2} \leq \frac{1}{Y} < \infty
\]
we can deduce:

**II.1.7. Proposition.** The area of $\Gamma \backslash \mathcal{H}$ is finite.

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

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

Recall that on $\mathcal{H}$ we have the norm

$$\|z\| = \frac{x^2 + (y + 1)^2}{y}$$

which, according to the discussion of norms in the essay on geometry, is strictly equivalent to the norm on $\mathcal{H}$ induced by the norm

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

on $G = \text{SL}_2(\mathbb{R})$.

The following is straightforward:

**II.1.8. Lemma.** In any region $|x| \leq C$ in $\mathcal{H}$ we have $\|z\| = \sup(y, 1/y)$.

The natural norm on the quotient $\Gamma \backslash \mathcal{H}$ is

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

**II.1.9. Proposition.** On any parabolic domain $\|z\|_{\Gamma \backslash \mathcal{H}} \equiv \|z\|$.

**Proof.** We have $\|g(x)\| \leq \|g\| \cdot \|x\|$ 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 $\infty$. The parabolic domain is horizontally bounded, so the norm on it is equivalent to $\sup(y, 1/y)$. On any compact region in $\mathcal{H}$ the Proposition is trivially true, so we may assume the parabolic domain to be $\left(\Gamma \cap \mathcal{H}\right) \backslash \mathcal{H}$ with $\mathcal{H} \gg 0$. But then Proposition II.1.6 and the fact (Proposition I.3.4) that the transform of $\mathcal{H}$ by

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

is a circle with top at $1/(c^2 Y)$. \hfill \square

2. The group $\text{SL}(2,\mathbb{Z})$

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 $u \cdot v = \Re(u \overline{v})$.

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.
Chapter II. Arithmetic groups

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

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

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

In these circumstances, $u$ has the least possible length of any complex number in the lattice spanned by $u$ and $v$. 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 $\text{Im}(v/u) > 0$. Hence:

II.2.1. 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 $\text{Im}(v/u) > 0$. 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, $|v| \geq |u|$. 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 $x = u \cdot v/|u|^2 + 1/2$ 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 $\text{SL}_2(\mathbb{Z}) \backslash \mathcal{H}$ 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 $n = \lfloor \text{Re}(z) + 1/2 \rfloor$, so that $n \leq x + 1/2 < n + 1$.
(2) (Signed swap) If $|z| < 1$ then replace $z$ by $-1/z$ and go to (1); otherwise stop.

II.2.2. Corollary. The region

$$|z| \geq 1, \quad -1/2 \leq x \leq 1/2$$

is a fundamental domain for $\text{SL}_2(\mathbb{Z})$. 
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 \( \Gamma = \text{SL}_2(\mathbb{Z}) \) then the quotient \( \Gamma \backslash \mathcal{H} \) has a single cusp. More precisely:

**II.2.3. Proposition.** The cusps of \( \text{SL}_2(\mathbb{Z}) \) in \( \mathbb{C} \) are the rational points in \( \mathbb{R} \cup \{\infty\} \), on which \( \text{SL}_2(\mathbb{Z}) \) acts transitively.

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

I describe an algorithm that shows how \( \text{SL}_2(\mathbb{Z}) \) 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

\[
c_n = qd_n + r \quad (0 \leq r < |d_n|)
\]

\[
c_{n+1} = d_n
\]

\[
d_{n+1} = r
\]

\[
\begin{bmatrix} c_{n+1} \\ d_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} \begin{bmatrix} c_n \\ d_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} M_n \begin{bmatrix} c_0 \\ d_0 \end{bmatrix}
\]

\[
M_{n+1} = \begin{bmatrix} 0 & 1 \\ 1 & -q \end{bmatrix} M_n
\]

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

\[
\Gamma(N) := \{ \gamma \in \text{SL}_2(\mathbb{Z}) \mid \gamma \equiv I \pmod{N} \} .
\]

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

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

**Proof.** One possible proof uses the Bruhat decomposition and is valid for \( \text{SL}_2 \) over any algebraic number field, but I offer here another. It comes down to showing that given a matrix \( \eta \) in \( M_2(\mathbb{Z}) \) with

\[
\det(\eta) \equiv 1 \pmod{N}
\]
there exists $\gamma$ in $\text{SL}_2(\mathbb{Z})$ with $\gamma \equiv \eta$. Transforming $\eta$ on left and right by elements of $\text{SL}_2(\mathbb{Z})$, we may assume $\eta$ to be diagonal (elementary divisor theorem). Suppose it is

$$\eta = \begin{bmatrix} a_2 & 0 \\ 0 & d_2 \end{bmatrix}$$

with $a_2, d_2 \equiv 1 \pmod{N}$. Choose $a$ and $d$ in $\mathbb{Z}$ which are inverse modulo $N^2$, congruent modulo $N$ to $a_2, d_2$ 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

$$1 \rightarrow \Gamma(N) \rightarrow \text{SL}_2(\mathbb{Z}) \rightarrow \text{SL}_2(\mathbb{Z}/N) \rightarrow 1$$

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

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

is exact, so:

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

The group $\text{SL}_2(\mathbb{Z})$ 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 $a \equiv a_*, c \equiv c_*$ mod $N$. Hence:

**II.3.4. Proposition.** The correspondence $a/c \mapsto \begin{bmatrix} a \\ c \end{bmatrix}$ is a bijective $\text{SL}_2(\mathbb{Z})$-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