INTRODUCTION TO SHIMURA VARIETIES

ANDREAS MIHATSCH

Abstract. These are lecture notes for a course on Shimura varieties I am currently teaching at Zhejiang University. Comments are highly welcome and much appreciated.

Contents

1. Introduction

In this first lecture, we will learn, very roughly, what Shimura varieties are and why they are interesting. Everything brought up today will be covered in much more detail later in the course, and it will be perfectly normal that many terms will be new during a first reading. Our goal today is only to get an overview.

1.1. **Why study Shimura varieties?** Shimura varieties combine two interesting properties:

- They are varieties defined over number fields which makes them interesting from a number theory perspective. Most importantly, their étale cohomology groups are representations of Galois groups of number fields.
- Their definition is in terms of reductive algebraic groups $G/\mathbb{Q}$. They come equipped with an action of the adelic points $G(\mathbb{A}_f)$, which implies that their étale cohomology groups are also $G(\mathbb{A}_f)$-representations.

Hence, the étale cohomology groups of Shimura varieties are both Galois and $G(\mathbb{A}_f)$-representations. Conjecturally, this two-fold structure is described by the global Langlands correspondence. Conversely, one can use the cohomology of Shimura varieties to prove important cases of this correspondence. This is the main motivation for our course, and our overall aim is to learn about several important ideas in this context.

Let us mention that Shimura varieties are also interesting for other reasons. For example, the study of heights on the Siegel variety plays an important role in Faltings's proof of the Mordell Conjecture [3]. Another example is the Gross–Zagier formula [5], which states an identity between height pairings of complex multiplication points on the modular curve and derivatives of L-functions. It plays a major role in the proof of cases of the Birch–Swinnerton-Dyer Conjecture. Its higher-dimensional generalizations, the arithmetic Gan–Gross–Prasad Conjectures [4,14], are an important topic in current arithmetic geometry research. In a related direction, the Kudla program [7] seeks to establish connections between cycles on Shimura varieties and modular forms or Eisenstein series. The proof of the averaged Colmez conjecture [1,13] has been an application of such ideas.

Date: September 17, 2025.

1.2. **This course.** The first part of our course will be an introduction to Shimura varieties. We will learn how to define them in terms of moduli spaces of abelian varieties and how to relate this definition to the group-theoretic one of Deligne. One of our goals is to obtain familiarity with the adelic formalism which will become important later.

In the second part of the course, we will study the cohomology of Shimura varieties. We will first get to know Matsushima's formula, which expresses the Betti cohomology of compact Shimura varieties in terms of automorphic representations. We will then learn about point counting in characteristic p (Langlands–Kottwitz method). The aim here is to give an orbital integral expression for the number of $\mathbb{F}_{p^n}$-points of the reduction mod p of the Shimura variety.

1.3. **References.** The following two are our main background references.

- The introductory lecture notes by Milne [10]. They focus on the group-theoretic definition of Shimura varieties and the definition of canonical models.
- The first few articles in the lecture notes volume [6]. They provide an introduction to PEL type Shimura varieties. The article of Yihang Zhu [15] is directly related to the material of the second part of the course.

1.4. **Prerequisites.** We will assume as little as possible. The only necessary background is some familiarity with varieties and algebraic number theory.

In the rest of this introduction, we sketch the definition of Shimura varieties and give an outline of the course contents.

1.5. **Shimura data.** Shimura varieties are attached to Shimura data. The formalism starts with a reductive group G over $\mathbb{Q}$. For example, G might be one of the following.

- $G = \mathrm{GL}_2$
- $G = \mathrm{GSp}_{2g}$, the general symplectic group in $2g$ variables. Let $J = \begin{pmatrix} & 1_g \\ -1_g & \end{pmatrix}$ be the matrix defining the standard symplectic form on $\mathbb{Q}^{2g}$. Then GSp is defined by
$$\mathrm{GSp}_{2g}(\mathbb{Q}) = \big\{g \in GL_{2g}(\mathbb{Q}) \mid {}^t g \cdot J \cdot g = c \cdot J \text{ for some } c \in \mathbb{Q}^\times\big\}. \tag{1.1}$$
It is related to the usual symplectic group Sp_{2g} by the exact sequence
$$1 \longrightarrow \mathrm{Sp}_{2g} \longrightarrow \mathrm{GSp}_{2g} \xrightarrow{c} \mathrm{GL}_1 \longrightarrow 1.$$
The map c is called the *similitude factor*. Note that $\mathrm{GSp}_2 = \mathrm{GL}_2$ and $\mathrm{Sp}_2 = \mathrm{SL}_2$, recovering the previous example.
- $G = \mathrm{U}(V)$, a unitary group. Let $K/\mathbb{Q}$ be an imaginary quadratic extension. (This means that $\mathbb{R} \otimes_\mathbb{Q} K \cong \mathbb{C}$.) Let V be an n-dimensional hermitian K-vector space. If V is not positive or negative definite then $\mathrm{U}(V)$ can occur as part of a Shimura datum.

Next, the formalism requires the datum of a homomorphism of real algebraic groups
$$h : \mathbb{C}^\times \longrightarrow G(\mathbb{R}) \tag{1.2}$$
which satisfies certain axioms introduced by Deligne [2]. Such an h is called a *Deligne homomorphism.* If $g \in G(\mathbb{R})$ is a real point of G, then we may conjugate h to define a new Deligne homomorphism,
$$\big(ghg^{-1}\big)(z) := gh(z)g^{-1}.$$
Let $S_h \subset G(\mathbb{R})$ denote the centralizer of h, meaning the subgroup of elements g with $ghg^{-1} = h$. The quotient $X = G(\mathbb{R})/S_h$ is precisely the set of Deligne homomorphisms

that are conjugate to h. An important consequence of Deligne's axioms is that X is a finite union of hermitian symmetric domains for $G(\mathbb{R})$. In particular, it is a complex manifold. The pair (G, X) is called a *Shimura datum*.

Example 1.1. Consider $G = \mathrm{GL}_2$. We can embedd $\mathbb{C}$ into $\mathrm{M}_2(\mathbb{R})$ as $\mathbb{R}$-algebra by

$$h(a+bi) := \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

If we restrict this embedding to unit groups, then we obtain a Deligne homomorphism $h : \mathbb{C}^\times \to \mathrm{GL}_{2,\mathbb{R}}$. Its centralizer is precisely $h(\mathbb{C}^\times)$ and the quotient X is the set of complex structures on $\mathbb{R}^2$. Since $\mathbb{C}^\times$ is connected and since $GL_2(\mathbb{R})$ has two connected components, X has two connected components. We want to give a more explicit description of X.

Recall that $\mathbb{P}^1(\mathbb{C})$ is the space of complex lines in $\mathbb{C}^2$. Clearly, the Lie group $\mathrm{GL}_2(\mathbb{C})$ acts on it by its natural action on $\mathbb{C}^2$. The subgroup $\mathrm{GL}_2(\mathbb{R})$ preserves the real projective line $\mathbb{P}^1(\mathbb{R})$ and hence acts on the complement,

$$\mathrm{GL}_2(\mathbb{R}) \ \circlearrowright \ \mathbb{C}\backslash\mathbb{R}, \quad g\cdot\tau = \frac{a\tau+b}{c\tau+d}. \tag{1.3}$$

The complement $\mathbb{C}\backslash\mathbb{R}$ is the union of the upper and lower half plane which we often denote by $\mathbb{H}^\pm$. As an open subset of $\mathbb{C}$, it is naturally a complex manifold. Let us compute the stabilizer of i:

$$\begin{aligned} i = \frac{ai+b}{ci+d} \quad &\iff \quad -c+di = ai+b \\ &\iff \quad a = d, \quad c = -b. \end{aligned} \tag{1.4}$$

That is, the stabilizer of i is precisely $h(\mathbb{C}^\times)$. Moreover, it is clear that $\mathrm{GL}_2(\mathbb{R})$ acts transitively on $\mathbb{H}^\pm$ because

$$\begin{pmatrix} a & b \\ & 1 \end{pmatrix} \cdot i = ai + b.$$

Hence, we see that

$$X \xrightarrow{\sim} \mathbb{H}^\pm, \quad ghg^{-1} \longmapsto g\cdot i \tag{1.5}$$

as smooth manifolds in a $\mathrm{GL}_2(\mathbb{R})$-equivariant way. We have not defined the complex structure on X, but it is, in fact, given by the complex structure on $\mathbb{H}^\pm$ under (1.5).

Remark 1.2. Some groups, such as GL_n with $n \geq 3$, cannot occur as part of a Shimura datum. For example, the dimension of the symmetric space for $\mathrm{GL}_3(\mathbb{R})$ is

$$\dim \mathrm{SL}_3(\mathbb{R}) - \dim \mathrm{SO}(3) = 8 - 3$$

which is odd and hence cannot be a complex manifold.

1.6. **Shimura varieties over $\mathbb{C}$.** Given a Shimura datum (G, X), one next defines a complex variety in the following way. Let $\mathbb{A}$ denote the ring of adeles of $\mathbb{Q}$, and let $\mathbb{A} = \mathbb{A}_f \times \mathbb{R}$ be its factorization into finite and archimedean part. (We will review these definitions later in the course.) Given an open compact subgroup $K \subset G(\mathbb{A}_f)$, the quotient $G(\mathbb{A}_f)/K$ is a discrete countably infinite set with transitive $G(\mathbb{A}_f)$-action. Hence, the product $X \times G(\mathbb{A}_f)/K$ is a countable union of copies of X. We consider the diagonal action

$$G(\mathbb{Q}) \ \circlearrowright \ X \times G(\mathbb{A}_f)/K.$$

If K is small enough then the $G(\mathbb{Q})$-action is free. (The technical term is "neat" and we will get to know it later in the course.) It is also properly discontinuous, so we can form the quotient complex manifold

$$\mathrm{Sh}_K(G,X)(\mathbb{C}) := G(\mathbb{Q})\backslash\big(X \times G(\mathbb{A}_f)/K\big). \tag{1.6}$$

At this point, we have defined the complex points of the *Shimura variety for Shimura datum* (G,X) *and level* K as a complex manifold. The theorem of Bailey–Borel states that there is a unique way to endow it with an algebraic structure.

Theorem 1.3 (Bailey–Borel, see [10, Corollary 3.16])**.** *There exists a quasi-projective complex variety* $\mathrm{Sh}_K(G,X)_{\mathbb{C}}$ *such that there exists an isomorphism of complex manifolds* $\mathrm{Sh}_K(G,X)_{\mathbb{C}}(\mathbb{C}) \xrightarrow{\sim} \mathrm{Sh}_K(G,X)(\mathbb{C})$. *This variety is unique up to isomorphism.*

Remark 1.4. Simple examples of non-unique algebraic structures on complex manifolds can be found in [11].

Example 1.5. Let us again consider the case $G = \mathrm{GL}_2$ and let us give an example of a connected component of (1.6). Let $\widehat{\mathbb{Z}} = \prod_{p<\infty} \mathbb{Z}_p$ be the subring of integral elements of $\mathbb{A}_f$. For $n \geq 1$, consider the kernel

$$K(n) = \ker\left(\mathrm{GL}_2(\widehat{\mathbb{Z}}) \longrightarrow \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})\right)$$

which is an open compact subgroup of $G(\mathbb{A}_f)$. It is small enough (in the above sense) if $n \geq 3$. The intersection

$$\Gamma(n) := \mathrm{GL}_2(\mathbb{Q}) \cap K(n)$$

is the classical congruence subgroup

$$\Gamma(n) = \left\{ \gamma \in \mathrm{GL}_2(\mathbb{Z}) \;\middle|\; \gamma \equiv \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \bmod n \right\}.$$

The quotients $\Gamma(n)\backslash\mathbb{H}^+$ and $\Gamma(n)\backslash\mathbb{H}^-$ will be two of the connected components of the complex manifold $\mathrm{Sh}_{K(n)}(\mathrm{GL}_2, \mathbb{H}^\pm)$.

1.7. **Shimura varieties over number fields.** Finally, one descends $\mathrm{Sh}_K(G,X)$ to a number field. Starting from a Shimura datum (G,X), Deligne defines a number field $E \subset \mathbb{C}$ called the *reflex field*. In a suitable sense, it is the smallest field over which the conjugacy class X is defined.

Example 1.6. Consider the three examples from §1.5.

- If $G = \mathrm{GL}_2$ or more generally $G = \mathrm{GSp}_{2g}$, then the reflex field is $\mathbb{Q}$.
- If $G = U(V)$ is a non-definite unitary group for an imaginary-quadratic field $K/\mathbb{Q}$, then the reflex field is the subfield $E \subset \mathbb{C}$ that is isomorphic to K.

Deligne [2] gave a definition of *canonical model* of $\mathrm{Sh}_K(G,X)_{\mathbb{C}}$ over E. It is a variety $\mathrm{Sh}_K(G,X)$ over $\mathrm{Spec}(E)$ together with an isomorphism

$$\mathbb{C} \otimes_E \mathrm{Sh}_K(G,X) \xrightarrow{\sim} \mathrm{Sh}_K(G,X)_{\mathbb{C}}$$

that satisfies a certain reciprocity law for complex multiplication points. Deligne proves that the canonical model $\mathrm{Sh}_K(G,X)$ is unique up to isomorphism if it exists.

Theorem 1.7 (Borovoi, Milne [8])**.** *For every Shimura datum, the canonical model exists.*

Definition 1.8. Let (G,X) be a Shimura datum with reflex field E and let $K \subset G(\mathbb{A}_f)$ be a sufficiently small evel subgroup. The Shimura variety of level K attached to (G,X) is the canonical model $\mathrm{Sh}_K(G,X)$ from Theorem 1.7.

Remark 1.9. Historically, the study of Shimura varieties started with Shimura in the 1960s. He first considered moduli spaces of abelian varieties with **P**olarization, **E**ndomorphisms, and **L**evel structure (PEL). These are the Shimura varieties defined by *PEL type* Shimura data.

Shimura also studied several non-PEL cases and defined the corresponding Shimura varieties as varieties over number fields. Deligne [2] gave a group-theoretic framework for Shimura's work. His definition in terms of a reciprocity law for complex multiplication points is extrapolated from the Shimura–Taniyama reciprocity law for abelian varieties with complex multiplication. Deligne also constructed the canonical model for abelian type Shimura varieties. The proof of existence in the general case was completed by Milne based on ideas of Borovoi. See here for a short summary of the history by Milne [9, §6].

Example 1.10. Consider the two cases from Example 1.6. The unitary group $U(V)$ has no PEL type Shimura data. For the group GSp_{2g}, there exists a PEL type Shimura datum (GSp_{2g}, X). Consider a principal congruence level subgroup

$$K(n) = \ker\left(\mathrm{GSp}_{2g}(\widehat{\mathbb{Z}}) \longrightarrow \mathrm{GSp}_{2g}(\mathbb{Z}/n\mathbb{Z})\right)$$

with $n \geq 3$. Then the canonical model $\mathrm{Sh}_{K(n)}(\mathrm{GSp}_{2g}, X)$ can be described as a moduli space of principally polarized abelian varieties with level-n-structure. For example, if we look at $\mathbb{C}$-points and specialize to GL_2, then we obtain

$$\mathrm{Sh}_{K(n)}(\mathrm{GL}_2, X)(\mathbb{C}) \xrightarrow{\sim} \{(E,\eta)/\mathbb{C}\}/\sim \tag{1.7}$$

where the right hand side denotes the set of isomorphism classes of pairs (E,η) with

- E an elliptic curve over $\mathbb{C}$,
- $\eta : (\mathbb{Z}/n\mathbb{Z})^{\oplus 2} \xrightarrow{\sim} E[n]$ a choice of basis for the n-torsion.

The datum η is called a *level structure* for E. Proving (1.7) will be one of our first goals.

1.8. **Hasse–Weil ζ-function.** We mentioned in §1.1 that Shimura varieties are interesting for the Galois representations defined by their étale cohomology (among other reasons). A concrete way to package this information is the *Hasse–Weil ζ-function*.

Let X be a smooth projective variety over $\mathbb{Q}$. Recall that for almost all primes p, there exists a smooth projective model $\mathcal{X}_p$ of X over $\mathbb{Z}_{(p)}$.

Definition 1.11. Let S be a finite set of primes such that for each $p \notin S$, X has a smooth projective model $\mathcal{X}_p$ over $\mathbb{Z}_{(p)}$. For $p \notin S$, define the p-factor

$$\zeta_p(X,s) := \exp\left(\sum_{n=1}^{\infty} |\mathcal{X}_p(\mathbb{F}_{p^n})| \cdot \frac{p^{-ns}}{n}\right). \tag{1.8}$$

This expression converges and defines a holomorphic function on the half plane $\mathrm{Re}(s) > \dim(X)$. An equivalent definition is as follows. Let $|\mathcal{X}_p|$ denote the set of closed points of $\mathcal{X}_p$. For $x \in |\mathcal{X}_p|$, the residue field $\kappa(x)$ is a finite extension $\mathbb{F}_{p^d}$ of $\mathbb{F}_p$. Then

$$\zeta_p(X,s) = \prod_{x\in|\mathcal{X}_p|} (1 - |\kappa(x)|^{-s})^{-1}. \tag{1.9}$$

The partial *Hasse–Weil ζ-function of* X is defined as the product

$$\zeta^S(X,s) := \prod_{p\notin S} \zeta_p(X,s)$$

which converges on the right half plane $\mathrm{Re}(s) > \dim(X) + 1$.

Example 1.12. (1) For $X = \mathrm{Spec}(\mathbb{Q})$, we can take $S = \emptyset$ and $\mathcal{X}_p = \mathrm{Spec}(\mathbb{Z}_{(p)})$. Then (1.9) comes out as $(1-p^{-s})^{-1}$ and $\zeta(\mathrm{Spec}(\mathbb{Q}), s)$ is the Riemann ζ-function.

(2) The number of $\mathbb{F}_q$-points of projective space is given by

$$\mathbb{P}^m(\mathbb{F}_q) = 1 + q + q^2 + \ldots + q^m.$$

Substituting this into the definition, we find

$$\zeta(\mathbb{P}^m, s) = \zeta(s)\zeta(s-1)\cdots\zeta(s-m).$$

(3) Let $E/\mathbb{Q}$ be an elliptic curve and let S be as in Definition 1.11. Then

$$\zeta^S(E,s) = \zeta^S(s)\cdot\zeta^S(s-1)\cdot L^S(E/\mathbb{Q},s)^{-1}$$

where ζ^S is the partial Riemann ζ-function and $L^S(E/\mathbb{Q},s)$ the partial L-function of the elliptic curve.

A surprising property of $\zeta_p(X,s)$ is that it is independent of the choice of integral model $\mathcal{X}_p$ at p. This follows from the Grothendieck–Lefschetz fixed point formula and proper smooth base change for étale cohomology which provide the following expression for the factor at p:

$$\zeta_p(X,s) = \prod_{i=0}^{2\dim(X)} \det\left(1-\sigma_p\cdot t \mid H^i_{\text{ét}}(X_{\overline{\mathbb{Q}}},\mathbb{Q}_\ell)\right)^{(-1)^{i+1}}\Big|_{t=p^{-s}} \tag{1.10}$$

Here, we still assume that X has a smooth projective model over $\mathbb{Z}_{(p)}$ and denote by $\sigma_p \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ a Frobenius element at p. The right hand side of (1.10) is independent of $\mathcal{X}_p$.

The Riemann ζ-function has a meromorphic continuation to the complex plane which satisfies a functional equation.

Conjecture 1.13 (Hasse, Weil)**.** For every smooth projective variety $X/\mathbb{Q}$, the Hasse–Weil ζ-function $\zeta^S(X,s)$ admits a meromorphic continuation and satisfies a functional equation.

The main strategy for proving these analytic properties is by relating $\zeta^S(X,s)$ to automorphic L-functions. This leads us back to Shimura varieties, where we have a (conjectural) description of the étale cohomology in terms of automorphic representations and Langlands reciprocity.

1.9. **The Langlands–Kottwitz method.** Langlands' strategy for proving Conjecture 1.13 for Shimura varieties was to find a group-theoretic expression for (1.8) that can be compared with the orbital integral side of the Arthur–Selberg trace formula. This should ultimately allow to bridge from the étale cohomology to automorphic representations. We end this introduction with a brief sketch of what is involved in this idea.

(1) Let us assume that the Shimura variety in question is of PEL type. Then, at almost all primes p, the integral model $\mathcal{X}_p$ can be constructed by simply extending the moduli description. The quantities $|\mathcal{X}_p(\mathbb{F}_q)|$ now have a very concrete meaning: they count the number of certain polarized abelian varieties with endomorphisms and level structure over $\mathbb{F}_q$.

(2) Next, one partitions the set $\mathcal{X}_p(\overline{\mathbb{F}_p})$ into isogeny classes. The main tool here is the Honda–Tate theorem which classifies the isogeny classes of abelian varieties over $\overline{\mathbb{F}_p}$ in terms of the characteristic polynomial of the Frobenius.

(3) Within each isogeny class, the number of $\mathbb{F}_{p^n}$-points is counted by a linear combination of expressions of the form

$$\mathrm{Orb}(\gamma,\mathbb{1}_{K^p})\cdot \mathrm{TOrb}(\delta,\phi_p) \tag{1.11}$$

where the first factor is a $G(\mathbb{A}_f^p)$-orbital integral, and the second factor a σ-twisted $G(\mathbb{Q}_{p^n})$-orbital integral.

(4) Finally, one hopes to compare the last expression with the orbital integral side of the Arthur–Selberg trace formula. This involves, among other things, a comparison of twisted orbital integrals with actual orbital integrals, which is the content of the Fundamental Lemma.

The final goal of this course is to explain some of these ideas in detail.

Part 1. **The Shimura variety of** GL_2

2. The upper half plane

In Example 1.1, we have introduced the action of $\mathrm{GL}_2(\mathbb{Q})$ on the union of upper and lower half plane $\mathbb{H}^\pm = \mathbb{C}\backslash\mathbb{R}$. Recall that it is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d}.$$

In Example 1.5, we have seen that we are especially interested in actions by subgroups such as $\mathrm{GL}_2(\mathbb{Z})$ and $\Gamma(n)$. Our aim in this section is to give a definition of such *arithmetic subgroups* and to prove properties about their action on $\mathbb{H}^\pm$.

Note that elements of $\mathrm{GL}_2(\mathbb{Z})$ have determinant 1 or -1, and that the elements of determinant -1 interchange upper and lower half plane. So we will focus on the action of $\mathrm{SL}_2(\mathbb{Q})$ on the upper half plane $\mathbb{H} \subset \mathbb{H}^\pm$.

2.1. **The fundamental domain.** Let $\mathcal{F}$ be the area defined by

$$\mathcal{F} = \left\{ \tau \in \mathbb{H} \;\middle|\; |\tau| \geq 1 \text{ and } -\frac{1}{2} \leq \mathrm{Re}(\tau) \leq \frac{1}{2} \right\}. \tag{2.1}$$

Its interior $\mathcal{F}^\circ$ is the open subset where $|\tau| > 1$ and $-1/2 \leq \mathrm{Re}(\tau) \leq 1/2$.

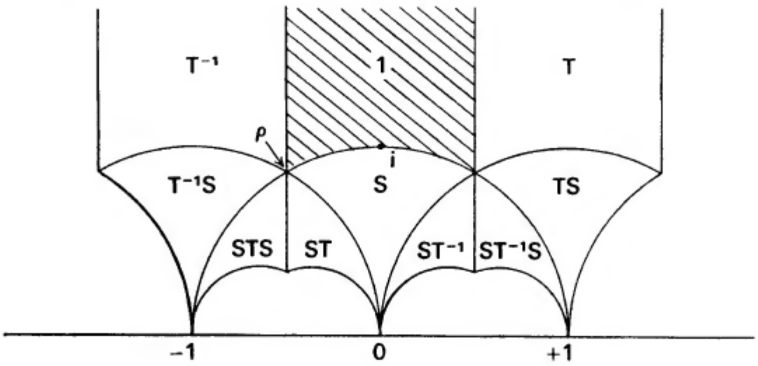

Figure 1. The area $\mathcal{F}$ is depicted in grey. The remaining areas show translates of $\mathcal{F}$ under the action of the elements S and T defined in (2.3). By Proposition 2.1 and Remark 2.2, these translates cover all of $\mathbb{H}$. The picture is taken from [12, §VII].

Proposition 2.1. *The set $\mathcal{F}$ is a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$ on $\mathbb{H}$. That is, it has the following two properties.*

(1) For every $\tau \in \mathbb{H}$, there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma\tau \in \mathcal{F}$.

(2) If $\gamma\tau \in \mathcal{F}^\circ$, then γ is uniquely determined up to multiplication by -1.

Proof. Fix $\tau \in \mathbb{H}$ and let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ be any element. By direct computation, we see that

$$\mathrm{Im}(\gamma\tau) = \mathrm{Im}\left(\frac{(a\tau + b)(c\bar{\tau} - d)}{|c\tau + d|^2}\right) = \frac{(ad - bc)\mathrm{Im}(\tau)}{|c\tau + d|^2} = \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2}. \tag{2.2}$$

The denominator $|c\tau + d|^2$ defines a positive definite quadratic form in $(c, d) \in \mathbb{Z}^2$. It hence takes a minimum on the set of (c, d) that occur as the bottom row of an element of $\mathrm{SL}_2(\mathbb{Z})$. (These are precisely the (c, d) with $\gcd(c, d) = 1$.) So we see that $\{\mathrm{Im}(\gamma\tau) \mid \gamma \in \mathrm{SL}_2(\mathbb{Z})\}$ has a maximum.

Let γ be such that $\mathrm{Im}(\gamma\tau)$ is maximal. Consider the two matrices

$$S = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \tag{2.3}$$

and observe that they act as the very simple transformations

$$S\tau = -\frac{1}{\tau}, \qquad T\tau = \tau + 1. \tag{2.4}$$

In particular, acting with a suitable power T^m, $m \in \mathbb{Z}$, we can translate $\gamma\tau$ to assume it lies in the strip $-1/2 \leq \mathrm{Re}(z) \leq 1/2$. Then also $|\gamma\tau| \geq 1$ because otherwise $\mathrm{Im}(S\gamma\tau) > \mathrm{Im}(\gamma\tau)$ would contradict the maximality of $\mathrm{Im}(\gamma\tau)$. This proves statement (1) of the proposition.

We now prove statement (2). Assume that τ and $\gamma\tau$ both lie in $\mathcal{F}^\circ$, our aim being to show that $\gamma \in \{\pm 1\}$. After possibly replacing the pair (γ, τ) by $(\gamma^{-1}, \gamma\tau)$, we can assume that $\mathrm{Im}(\gamma\tau) \geq \mathrm{Im}(\tau)$.

It is not possible that $d = 0$. Namely, then we would have $\gamma = \pm S$. Since $|\tau| > 1$ by assumption, we would find $\mathrm{Im}(\gamma\tau) < \mathrm{Im}(\tau)$ in contradiction with our choice of γ and τ.

So we know $d \neq 0$. Consider again (2.2), which now yields $|c\tau + d|^2 \leq 1$. Clearly, if $c \neq 0$, then there is no way that $c\tau + d$ lies in the unit disk $\{|z| \leq 1\}$ (again use that $\tau \in \mathcal{F}^\circ$). So we find $c = 0$ and hence

$$\gamma = \pm \begin{pmatrix} 1 & m \\ & 1 \end{pmatrix}$$

for some $m \in \mathbb{Z}$. Since both τ and $\gamma\tau$ have real part in $(-1/2, 1/2)$, the only possibility is $m = 0$. This finishes the proof. □

Remark 2.2. One can show that the matrices S and T from (2.3) generate $\mathrm{SL}_2(\mathbb{Z})$. That is, every element of $\mathrm{SL}_2(\mathbb{Z})$ can be written as a product of the three elements S, T and T^{-1}. The proof is elementary and can be found in [12, §VII.1, Theorem 2].

2.2. **Arithmetic subgroups of $\mathrm{SL}_2(\mathbb{Q})$.** We now define arithmetic subgroups of $\mathrm{SL}_2(\mathbb{Q})$.

Definition 2.3. (1) For $n \geq 1$, we define the *principal congruence subgroup* $\Gamma(n)$ by

$$\Gamma(n) = \{\gamma \in \mathrm{SL}_2(\mathbb{Z}) \mid \gamma \equiv 1 \bmod n\}.$$

(2) We call a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Q})$ *arithmetic* if it contains a principal congruence group $\Gamma(n)$ with finite index.

The group SL_2 has a very interesting property which will come up again later. Namely, for each $n \geq 1$, the projection map

$$\mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$$

is surjective. This is not hard to show, but is also not obvious.

Example 2.4. By the surjectivity we just stated for SL_2, the image of the projection map $\mathrm{GL}_2(\mathbb{Z}) \to \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ is the set of matrices with determinant ± 1. In particular, this projection is not surjective when $n = 5$ or $n \geq 7$.

In the context of Definition 2.3, we see that $\Gamma(n) \trianglelefteq \mathrm{SL}_2(\mathbb{Z})$ is a normal subgroup of index equal to $|\,\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})|$. In particular, if a group Γ contains $\Gamma(n)$ with finite index, then it also contains all $\Gamma(mn)$ with finite index.

Proposition 2.5. *Let Γ be an arithmetic subgroup.*

(1) There exists a lattice $\Lambda \subset \mathbb{Q}^2$ such that $\Gamma \subseteq \mathrm{SL}(\Lambda)$.

(2) More precisely, there exist an integer n and an element $g \in \mathrm{GL}_2(\mathbb{Q})$ such that

$$\Gamma(m) \;\subseteq\; g\Gamma g^{-1} \;\subseteq\; \mathrm{SL}_2(\mathbb{Z}).$$

Proof. The two statements are proved by very simple and effective arguments. First, by assumption on Γ, there exists an integer n such that $\Gamma(n) \subseteq \Gamma$ with finite index. Let $\gamma_1, \ldots, \gamma_r$ be representatives for the cosets $\Gamma/\Gamma(n)$. Then Γ stabilizes the lattice

$$\Lambda := \sum_{i=1}^{r} \gamma_i \cdot \mathbb{Z}^2.$$

Indeed, since $\gamma\mathbb{Z}^2 = \mathbb{Z}^2$ for every $\gamma \in \Gamma(n)$, we can also write Λ as

$$\Lambda = \sum_{\gamma \in \Gamma} \gamma \cdot \mathbb{Z}^2,$$

and from this second expression the Γ-stability is clear. This means that $\Gamma \subseteq \mathrm{SL}(\Lambda)$ which proves statement (1).

Let $\lambda_1, \lambda_2 \in \Lambda$ be a basis as $\mathbb{Z}$-module. Viewing λ_1 and λ_2 as column vectors, the base change matrix $g = (\lambda_1\ \lambda_2)$ lies in $\mathrm{GL}_2(\mathbb{Q})$ and has the property $g\mathbb{Z}^2 = \Lambda$. Then $\mathrm{SL}_2(\mathbb{Z}) = g^{-1}\,\mathrm{SL}(\Lambda)g$ and hence $g\Gamma g^{-1} \subseteq \mathrm{SL}_2(\mathbb{Z})$.

We still need to show that $g\Gamma g^{-1}$ contains a principal congruence subgroup. This is the content of the next lemma which completes the proof. □

Lemma 2.6. *Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Q})$ be an arithmetic subgroup and $g \in \mathrm{GL}_2(\mathbb{Q})$. Then $g\Gamma g^{-1}$ is again an arithmetic subgroup.*

Proof. Let d be the gcd of all the denominators of all the entries of g and g^{-1}. Then, if $A \in d^2m\mathrm{M}_2(\mathbb{Z})$ is an integer matrix divisible by d^2m, we find $g^{-1}Ag \in m\mathrm{M}_2(\mathbb{Z})$. This shows that $g^{-1}\Gamma(d^2m)g \subseteq \Gamma(m)$ which is equivalent to

$$\Gamma(d^2m) \ \subseteq\ g\Gamma(m)g^{-1}. \tag{2.5}$$

Now, for the given Γ, choose n with $\Gamma(n) \subseteq \Gamma$. Conjugating this relation by g and using (2.5), we find $\Gamma(d^2n) \subseteq g\Gamma g^{-1}$ which proves that $g\Gamma g^{-1}$ is again arithmetic. □

In other words, Proposition 2.5 shows that the arithmetic subgroups in $\mathrm{SL}_2(\mathbb{Q})$ are precisely the $\mathrm{GL}_2(\mathbb{Q})$-conjugates of groups between $\mathrm{SL}_2(\mathbb{Z})$ and some $\Gamma(n)$.

2.3. **Stabilizers.**

Definition 2.7. We say that an arithmetic subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Q})$ is *neat* if it is torsion free.

Proposition 2.8. *Let Γ be a neat arithmetic subgroup of $\mathrm{SL}_2(\mathbb{Q})$. Then Γ acts with trivial stabilizers on $\mathbb{H}$. That is, if $\gamma\tau = \tau$ for some $\gamma \in \Gamma$ and $\tau \in \mathbb{H}$, then $\gamma = 1$.*

Proof. We have seen in (1.4) that the stabilizer of $i \in \mathbb{H}$ in $\mathrm{GL}_2(\mathbb{R})$ is a copy of $\mathbb{C}^\times$. The unit circle $\mathbb{C}^1 \subset \mathbb{C}^\times$ is compact and equals the intersection $\mathbb{C}^\times \cap \mathrm{SL}_2(\mathbb{R})$. For a general point $\tau \in \mathbb{H}$, we can write $\tau = g \cdot i$ for some $g \in \mathrm{GL}_2(\mathbb{R})$. Then the stabilizers S_i and S_τ of τ and i in $\mathrm{SL}_2(\mathbb{R})$ are then related by $S_\tau = gS_ig^{-1}$. In this way, we see that for every $\tau \in \mathbb{H}$, the stabilizer $S_\tau \subset \mathrm{SL}_2(\mathbb{R})$ is isomorphic to $\mathbb{C}^1$, in particular compact.

Assume that $\gamma\tau = \tau$, where $\gamma \in \Gamma$ and $\tau \in \mathbb{H}$. This is equivalent to $\gamma \in \Gamma \cap S_\tau$. Since $\Gamma \subset \mathrm{GL}_2(\mathbb{R})$ is a discrete subgroup, the intersection $\Gamma \cap S_\tau$ is a discrete subgroup of S_τ. Since the discrete subgroups of $\mathbb{C}^1$ are all finite cyclic (generated by a root of unity), and since Γ is torsion-free by assumption, we see that $\Gamma \cap S_\tau = \{1\}$. Hence $\gamma = 1$, and the proof is complete. □

The next proposition provides a simple criterion for detecting neatness.

Proposition 2.9. *For all $n \geq 3$, the principal congruence subgroup $\Gamma(n)$ is neat. In particular, if $\Gamma \subseteq \Gamma(n)$ is an arithmetic subgroup, then Γ is neat.*

Proof. The minimal polynomial $\Phi_d(T)$ of a primitive d-th root of unity has degree $\varphi(d)$ (Euler φ-function). Recall that $\Phi_d(T)$ is called the d-th cyclotomic polynomial and that

$$T^m - 1 = \prod_{d|m} \Phi_d(T).$$

The only values for d such that $\varphi(d) \leq 2$ are 1, 2, 3, 4, and 6. These are precisely the values for d such that $\mathbb{Q}(\zeta_d)$ has degree ≤ 2 over $\mathbb{Q}$.

Let $n \geq 1$ and let $\gamma \in \mathrm{SL}_2(\mathbb{Q})$ be a torsion element, say $\gamma^m = 1$. Then the minimal polynomial of γ divides $T^m - 1$, and hence the characteristic polynomial $P(T)$ of γ is a product of $\Phi_d(T)$ with $d \mid m$. The only possibilities for $P(T)$ are hence

$$(T-1)^2, \quad (T+1)^2, \quad T^2+1, \quad T^2+T+1, \quad T^2-T+1.$$

If $n \geq 3$ and if γ is integral with $\gamma \equiv 1 \bmod n$, then also $P(T) \equiv (T-1)^2 \bmod n$, leaving $\gamma = 1$ as the only possibility. $\square$

Conclusion 2.10. In this lecture, we saw the definition of neat arithmetic subgroups of $\mathrm{SL}_2(\mathbb{Q})$. We have seen in Proposition 2.8 that such groups act freely on $\mathbb{H}$. So the quotient $\Gamma\backslash\mathbb{H}$ will be a Riemann surface and the quotient map

$$\mathbb{H} \longrightarrow \Gamma\backslash\mathbb{H}$$

a holomorphic covering map in the sense of topology. We can think of $\Gamma\backslash\mathbb{H}$ as being glued from finitely many translates of the fundamental domain $\mathcal{F}$ as in Figure 2.1 along their edges.

References

[1] Andreatta, Fabrizio; Goren, Eyal Z.; Howard, Benjamin; Madapusi, Keerthi; *Faltings heights of abelian varieties with complex multiplication*, Ann. of Math. (2) **187** (2018), no. 2, 391–531.

[2] Deligne, Pierre; *Travaux de Shimura*, Séminaire Bourbaki, 23ème année (1970/1971), Exp. No. 389, pp. 123–165. Lecture Notes in Math. **244**, Springer-Verlag, Berlin-New York, 1971.

[3] Faltings, Gerd; *Finiteness theorems for abelian varieties over number fields*, Invent. Math. **73** (1983), no. 3, 349–366.

[4] Gan, Wee Teck; Gross, Benedict H.; Prasad, Dipendra; *Symplectic local root numbers, central critical L-values, and restriction problems in the representation theory of classical groups* in Sur les conjectures de Gross et Prasad. I, Astérisque **346** (2012), 1–109.

[5] Gross, Benedict H.; Zagier, Don B.; *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320.

[6] Haines, Thomas (ed.); Harris, Michael (ed.); *Shimura varieties*, London Math. Soc. Lecture Note Ser., No. 457, Cambridge University Press, Cambridge, 2020.

[7] Kudla, Stephen; Rapoport, Michael; *Special cycles on unitary Shimura varieties II: Global theory*, J. Reine Angew. Math. **697** (2014), 91–157.

[8] Milne, James S.; *The action of an automorphism of $\mathbb{C}$ on a Shimura variety and its special points* in *Arithmetic and geometry*, Vol. I, 239–265. Progr. Math. **35**, Birkhäuser Boston, Inc., Boston, MA, 1983.

[9] ———; *Canonical models of Shimura curves*, lecture notes available at https://www.jmilne.org/math/articles/2003a.pdf, 2003.

[10] ———; *Introduction to Shimura varieties*, lecture notes available at https://www.jmilne.org/math/xnotes/svi.html, 2017.

[11] Jelonek, Zbigniew; *Simple examples of affine manifolds with infinitely many exotic models*, Adv. Math. **284** (2015), 112–121.

[12] Serre, Jean-Pierre; *A course in arithmetic*, Grad. Texts in Math., No. 7, Springer-Verlag, New York-Heidelberg, 1973.

[13] Yuan, Xinyi; Zhang, Shou-Wu; *On the averaged Colmez conjecture*, Ann. of Math. (2) **187** (2018), no. 2, 533–638.

[14] Zhang, Wei; *On arithmetic fundamental lemmas*, Invent. Math. **188** (2012), no. 1, 197–252.

[15] Zhu, Yihang; *Introduction to the Langlands–Kottwitz method* in [6], 115–150.