INTRODUCTION TO SHIMURA VARIETIES

RUI CHEN

1. Introduction

This is my study note for Shimura varieties based on the section 2 from the note of Kaiwen Lan [Lan17] and the chapter from the thesis of Boxer [Box15].

2. Shimura varieties

Let's assume G is a connected reductive group over $\mathbb{Q}$, we consider a manifold $\mathcal{D}$ with a smooth transitive action of $G(\mathbb{R})$.

Example 2.1. The group $G(\mathbb{R}) = SL_2(\mathbb{R})$ acting transitively on the upper half plane

$$\mathcal{H} := \{\ z \in \mathbb{C} :\ \mathrm{Im}(z) > 0\ \}$$

for each $z \in \mathcal{H}$, the stabilizer of $i \in \mathcal{H}$ is $SO_2(\mathbb{R})$, so we have

$$\mathcal{H} = SL_2(\mathbb{R}) \cdot i = SL_2(\mathbb{R})/SO_2(\mathbb{R})$$

For any $(G, \mathcal{D})$ as above, and for any open compact subgroup U of $G(\mathbb{A}^\infty)$, we can define the double coset subspace

$$X_U := G(\mathbb{Q})\backslash\mathcal{D} \times G(\mathbb{A}^\infty)/U$$

where $G(\mathbb{Q})$ acts diagonally on $\mathcal{D} \times G(\mathbb{A}^\infty)$ from the left hand side, and U acts only on $G(\mathbb{A}^\infty)$ from the right-hand side.

The group $G(\mathbb{A}^\infty)$ has a natural right action on the collection $\{X_U\}_U$ induced by

$$\mathcal{D} \times G(\mathbb{A}^\infty) \cong \mathcal{D} \times G(\mathbb{A}^\infty) :\ (x, h) \mapsto (x, hg)$$

which maps $X_{gUg^{-1}}$ to X_U as $h(gug^{-1})g = hgu$. Such an action provides a natual Hecke actions on the limit of cohomology groups $\lim H^*(X_U, \mathbb{C})$. This is crucial for relating the geometry of such double cosets to the theory of automorphic representations.

Let $\mathcal{D}^+$ be a connected component of $\mathcal{D}$, which admits a transitive action of $G(\mathbb{R})^+$, the identity component of $G(\mathbb{R})$ in the real analytic topology, let $G(\mathbb{R})_+$ denote the stabilizer of $\mathcal{D}^+$ in $G(\mathbb{R})$, let

$$G(\mathbb{Q})_+ := G(\mathbb{Q}) \cap G(\mathbb{R})_+$$

which is a subgroup of $G(\mathbb{Q})$ stabilizing $\mathcal{D}^+$, it is known that

$$\#(G(\mathbb{Q})_+\backslash G(\mathbb{A}^\infty)/U)) < \infty$$

which means there exists a subset $\{g_i\}_{i\in I}$ of $G(\mathbb{A}^\infty)$ indexed by a finite set I such that we have a disjoint union

$$G(\mathbb{A}^\infty) = \sqcup G(\mathbb{Q})_+ g_i U$$

then

$$\begin{aligned} X_U &\cong G(\mathbb{Q})_+\backslash\mathcal{D}^+ \times G(\mathbb{A}^\infty)/U \\ &= \sqcup_{i\in I} G(\mathbb{Q})_+\backslash\mathcal{D}^+ \times G(\mathbb{Q})_+ g_i U/U \\ &= \sqcup_{i\in I} \Gamma_i\backslash\mathcal{D}^+ \end{aligned}$$

where $\Gamma_i := G(\mathbb{Q})_+ \cap g_i U g_i^{-1}$, each Γ_i is an arithmetic subgroup of $G(\mathbb{Q})$.

If each Γ_i acts freely on $\mathcal{D}^+$, then X_U is a manifold because $\mathcal{D}^+$ is, this is the case when each Γ_i is a neat arithmetic subgroup of $G(\mathbb{Q})$.

Date: September 2024.

Shimura varieties are not just the double coset spaces X_U, for arithmetic applications, it is desirable that each X_U is an algebraic variety over $\mathbb{C}$ with a model defined over some canonically determined number field. This leds to the definition of a Shimura datum $(G, \mathcal{D})$. Consider the Deligne torus $S = \mathrm{Res}_{\mathbb{C}/\mathbb{R}} \mathbb{G}_m$.

Definition 2.2. Let G be a connected reductive group, and require $\mathcal{D}$ to be a $G(\mathbb{R})$-conjugacy class of homomorphisms

$$h : S \to G_{\mathbb{R}}$$

satisfying the following conditions:

(1) The representation defined by h and the adjoint representation of $G(\mathbb{R})$ on $\mathfrak{g}$ induces a decomposition

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}^+ \oplus \mathfrak{p}^-$$

such that

$$z \in S(\mathbb{R}) = \mathbb{C}^\times$$

acts by 1, $z/\overline{z}$ and $\overline{z}/z$ on the three summands.
(2) $h(i)$ induces a Cartan involution on $G^{ad}(\mathbb{R})$.
(3) $G^{ad}_{\mathbb{R}}$ has no non-trivial $\mathbb{Q}$-simple factor H such that $H(\mathbb{R})$ is *compact*.

Here a Cartan involution θ on a linear algebra group G over $\mathbb{R}$ is an involution θ of G such that $G^\theta(\mathbb{R})$ is compact, G has a Cartan involution if and only if G is reductive.

By definition $\mathcal{D} \cong G/K_\infty$ for K_∞ the stabilizer of $h(\mathbb{S})$, and in practice this is the stabilizer of $h(i)$, from the axiom for Shimura data this is a maximal compact subgroup of $G(\mathbb{R})$ mod center.

Example 2.3. For $G = GL_2/\mathbb{Q}$

$$h_0 : \ S(\mathbb{R}) = \mathbb{C}^\times \longrightarrow GL_2(\mathbb{R})$$
$$z = x + iy \longmapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

we have $\mathrm{Stab}_{h_0}(GL_2(\mathbb{R})) = O(2) \cdot \mathbb{R}^\times$, this is indeed compact modulo the center.

Given a homomorphism

$$h : S \longrightarrow G_{\mathbb{R}}$$

we can get $h_{\mathbb{C}} : \mathbb{G}_{m,\mathbb{C}} \times \mathbb{G}_{m,\mathbb{C}} \to G_{\mathbb{C}}$, we define

$$\mu : \mathbb{G}_{m,\mathbb{C}} \to \mathbb{G}_{m,\mathbb{C}} \times \mathbb{G}_{m,\mathbb{C}} \to G_{\mathbb{C}}$$

via $z \mapsto (z, 1)$.

Fact: There is a one one correspondence between the $G_{ad}(\mathbb{R})$-conjugacy classes of $h : S \to G_{\mathbb{R}}$ satisfying axioms (1) and (2) and the $G(\mathbb{C})$ conjugacy classes of $\mu : \ \mathbb{G}_m \to G_{\mathbb{C}}$ that are *minuscule*.

Here minuscule, means if we choose T a maximal torus of G, such that $\mu \in X_*(T)_+$, view μ as $X^*(T)^+$ corresponds to $G^\vee \to GL(V_\mu)$, then all weights in V_μ have multiplicity one.

Remark 2.4. This fact is useful for the classification of Shimura data.

Now we are going to state the beautiful results concerning the canonical models of Shimura varieties

Theorem 2.5. *Suppose $(G, \mathcal{D})$ is a Shimura datum, then the whole collection of complex manifolds $\{X_U\}$ with U varying over among neat open compact subgroups of $G(\mathbb{A}^\infty)$ is the complex analytification of a canonical collection of smooth quasi-projective varieties over $\mathbb{C}$, moreover the analytic covering map*

$$X_U \longrightarrow X_{U'}$$

where $U \subset U'$ are given by the complex analytifications of canonical finite etale algebraic morphisms between the corresponding varieties.

This theorem is proved by constructing the so called *Satake-Baily-Borel* or the *minimal* compactifications of X_U or $\Gamma \backslash \mathcal{D}^+$ which are projective varieties over $\mathbb{C}$.

Theorem 2.6. *Suppose $(G, \mathcal{D})$ is a Shimura datum, then there exists a number field F_0 given as a subfield of $\mathbb{C}$ depending only on $(G, \mathcal{D})$, called the reflex field of $(G, \mathcal{D})$ such that the whole collection of complex manifolds $\{X_U\}_U$ with U varying over among neat open compact subgroups of $G(\mathbb{A}^\infty)$, is the complex analytification of the pull back to $\mathbb{C}$ of a canonical collection of smooth quasi-projective varieties over F_0, which satisfies certain additional properties qualifying them as the canonical models of $\{X_U\}_U$, moreover the analytic covering maps*

$$X_U \to X_{U'}$$

when $U \subset U'$ are also given by the complex analytifications of canonical finite etale algebraic morphisms defined over F_0 between corresponding canonical models.

Remark 2.7. As we will see later 2.1, here the canonical means to be canonical with respect to the embeddings of tori.

Example 2.8. Let $G = GL_2$, let $K = \hat{\Gamma}_0(N)$ be the subgroup

$$K = \hat{\Gamma}_0(N) = \{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G(\hat{\mathbb{Z}}) \mid c \equiv 0 \pmod{N} \}$$

the determinant map $\det : G \to \mathbb{G}_m$ induces

$$X_K(\mathbb{C}) \to \mathbb{Q}^\times \backslash \{\pm\} \times \mathbb{A}^\infty / \det(K)$$

provides a bijection from the geometrically connected components of $X_K(\mathbb{C})$ and the right hand side, since we have $\det(K) = \hat{\mathbb{Z}}$ and hence

$$X_K(\mathbb{C}) = G(\mathbb{Q}) \backslash \mathcal{H}^\pm \times G(\mathbb{A}^\infty) /\ K$$

has a single connected component, which can be identified $X_K(\mathbb{C})$ with $\Gamma_0(N)\backslash\mathcal{H}^+$ where $\Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$ is the usual congruence subgroup of matrices that reduce to upper triangular matrices modulo N, for $N \geq 3$, this is the complex points of the smooth, geometrically connected modular curve $Y_0(N)$ over $\mathbb{Q}$.

Since the maps $X_U \to X_{U'}$ are algebraic and defined over F_0, we have a canonical action of $\mathrm{Gal}(\overline{\mathbb{Q}}/F_0)$ on $\lim H^*_{et}(X_U, \mathbb{Q}_\ell)$ which is compatible with the Hecke action of $G(\mathbb{A}^\infty)$. This is an instance of the compatibility between the Hecke and Galois symmetries.

It is natural to consider the relative setting.

Definition 2.9. We define a morphism of Shimura data

$$(G_1, \mathcal{D}_1) \to (G_2, \mathcal{D}_2)$$

to be a group homomorphism $G_1 \to G_2$ and mapping $\mathcal{D}_1 \to \mathcal{D}_2$. If $\mathcal{D} = G_1(\mathbb{R}) \cdot h_0$ is the conjugacy class of some homomorphism $h_0 : \mathbb{S} \to G_{1,\mathbb{R}}$, then this means the composition of h_0 with $G_{1,\mathbb{R}} \to G_{2,\mathbb{R}}$ lies in the conjugacy class $\mathcal{D}_2$. If U_1 and U_2 are open compact subgroups of $G_1(\mathbb{A}^\infty)$ and $G_2(\mathbb{A}^\infty)$, such that U_1 is mapped into U_2, then we obtain the corresponding morphism $X_{U_1} \to X_{U_2}$, this morphism is defined over the subfield of $\mathbb{C}$ generated by the reflex fields of $(G_1, \mathcal{D}_1)$ and $(G_2, \mathcal{D}_2)$.

The most important examples are given by special points or CM points, which are *zero-dimensional* special subvarieties defined by the subgroup G_1 of G_2 that are tori, these are generalizations of the points of modular curves parametrizing CM elliptic curves with level structures. Zero dimensional Shimura varieties and special points are important because their canonical models can be defined more directly and hence they are useful for characterizing the canonical models of Shimura varieties of positive dimension.

Example 2.10. Let $E/\mathbb{Q}$ be an imaginary quadratic extension, recall that $G(\mathbb{Q})$ acts on $\mathcal{H}^\pm$ via conjugation, there is a unique point $x_\iota \in \mathcal{H}^\pm$ with $\det(x_\iota i) > 0$ and whose stabilizer in $G(\mathbb{Q})$ is $\iota(E^\times)$, for example if $E = \mathbb{Q}(\sqrt{d}i)$, then we can take $x_\iota = \sqrt{d}i \in \mathcal{H}^+$, the set of CM-points of level K can be defined as

$$CM_K(\mathbb{C}) := \{[x_\iota, g]_K \mid g \in G(\mathbb{A}^\infty)\}$$

The $Gal(E^{ab}/E)$ action on the set CM_K is described by the reciprocity law:

$$\mathrm{Art}_E :\ E^\times \backslash \mathbb{A}_E^\times \to \mathrm{Gal}(E^{ab}/E)$$

if $\sigma \in Gal(E^{ab}/E)$ and $a = (a_\infty, a_f)$ is such that $Art(a) = \sigma$ then

$$\sigma[x_\iota, g]_K = [x_\iota, \iota(a_f)g]$$

Let G be a connected reductive group over $\mathbb{Q}$ with a Shimura variety $Sh_K(G,X)$, E its reflex field, if $G=T$ is a torus, then $Sh_K(T)(\mathbb{C}) = T(\mathbb{Q})\backslash T(\mathbb{A}_f)/K$ is a finite set, we can associate to $Sh_K(T)$ a cocharacter $\mu : \mathbb{G}_m \to T_{\mathbb{C}}$, and μ is defined over $E \subseteq \mathbb{C}$ for E the reflex field, the canonical model of $Sh_K(T)$ is described by Shimura's reciprocity map

$$\text{Rec}_\mu : \ \text{Gal}(\overline{Q}/E) \longrightarrow \text{Gal}(\overline{E}/E) \cong E^\times\backslash \mathbb{A}_E^\times / E_{\mathbb{R}}^{\times,0} \longrightarrow \mathbb{G}_{m,E}(\mathbb{Q})\backslash \mathbb{G}_{m,E}(\mathbb{A}_f) \longrightarrow T(\mathbb{Q})\backslash T(\mathbb{A}_f) \tag{2.1}$$

it is the E-scheme over $\text{Spec}(E)$ such that the induced action of $\tau \in \text{Gal}(\overline{\mathbb{Q}}/E)$ on $Sh_K(T)(\mathbb{C})$ is given by $Res_\mu(T)$.

For general (G,X), a canonical model is an E-scheme such that for every morphism $(T,\{h\}) \to (G,X)$ of Shimura data, the natural morphism

$$T(\mathbb{Q})\backslash\{h\} \times T(\mathbb{A}_f)/K \cap T(\mathbb{A}_f) \to G(\mathbb{Q})\backslash X \times G(\mathbb{A}_f)/K$$

is induced from $Sh_{K\cap T(\mathbb{A}_f)}(T,\{h\}) \to Sh_K(G,X) \times_{\text{Spec}(E)} \text{Spec } E(T,\{h\})$.

3. PEL datum

Definition 3.1. A *rational PEL datum* is a tuple $(B,*,V,\langle\ ,\ \rangle,h)$ where

- B is a finite dimensional semisimple $\mathbb{Q}$-algebra.
- $*$ is a positive involution on B, i.e. $\text{tr}_{B/\mathbb{Q}}(xx^*) > 0$ for all nonzero $x \in B$.
- V is a finitely generated left B-module.
- $\langle\ ,\ \rangle : \ V \times V \to \mathbb{Q}(1)$ is an alternating form such that
$$\langle bv, \omega\rangle = \langle v, b^*\omega\rangle$$
for all $v, w \in V$ and $b \in B$.
- $h : \ \mathbb{C} \to \text{End}_{B_{\mathbb{R}}}(V_{\mathbb{R}})$ is a homomorphism of $\mathbb{R}$-algebras such that
$$\langle h(z)v, w\rangle = \langle v, h(\overline{z})w\rangle$$
for all $z \in \mathbb{C}$ and $v, w \in V$, and such that the symmetric form $\frac{1}{2\pi i}\langle v, h(i)w\rangle$ is positive definite.

The homomorphism h defines a decomposition

$$V \otimes \mathbb{C} = V_0 \oplus V_0^c$$

as $\mathbb{C}$ vector spaces where $h(z)$ acts as z on V_0 and $\overline{z}$ on V_0^c, this decomposition is stable under the action of B and each factor is isotropic for $\langle\ ,\ \rangle$.

Let $(B,*)$ be a finite dimensional semisimple $\mathbb{Q}$ algebra with positive involution as above and let F be its center. We let $\mathcal{T}$ denote the set of embeddings $\tau : F \to \mathbb{C}$, via the fixed isomorphism $i : \overline{\mathbb{Q}}_p \to \mathbb{C}$, we may view it as the set of embeddings $\tau : F \to \overline{\mathbb{Q}}_p$, we have a decomposition

$$F = \prod_{[\tau]} F_{[\tau]}$$

of F into a product of number fileds, where the product is over the $\text{Aut}(\mathbb{C})$ orbits of $\mathcal{T}$, we have a corresponding decomposition

$$F = \prod_{[\tau]} F_{[\tau]}$$

of B where each $B_{[\tau]}$ is simple with center $F_{[\tau]}$. The positivity of $*$ forces it to preserve this decomposition, and hence $(B,*)$ is a product of finite dimensional simple $\mathbb{Q}$-algebras with positive involution.

We now recall that a simple $\mathbb{Q}$-algebra with positive involution $(B,*)$ falls into one of the three classes, let F denote the center of B and $F^+ \subset F$ the subfield fixed by $*$.

- (type A) F/F^+ is a totally imaginary quadratic extension of a totally real field F^+.
- (type C) $F = F^+$ is totally real and for every embedding $\tau : F \to \mathbb{R}$, $B \otimes_{F,\tau} \mathbb{R} \cong M_n(\mathbb{R})$ for some integer n.
- (type D) $F = F^+$ is totally real and for every embedding $\tau : F \to \mathbb{R}$, $B \otimes_{F,\tau} \mathbb{R} \cong M_n(\mathbb{H})$ for some integer n, where $\mathbb{H}$ denotes the real quaternion algebra.

Definition 3.2. The *reflex field* of the PEL datum $(B, *, V, \langle\ ,\ \rangle, h)$ is the subfield F_0 of $\mathbb{C}$ over which the $B \otimes \mathbb{C}$ module V_0 is defined, i.e. it is the subfield of $\mathbb{C}$ fixed by all those $\sigma \in \mathrm{Aut}(\mathbb{C})$ such that

$$V_0^\sigma := V_0 \otimes_{\mathbb{C},\sigma} \mathbb{C} \cong V_0$$

as $B \otimes \mathbb{C}$ modules. Equivalently, it is the subfield of $\mathbb{C}$ generated by all the traces $\mathrm{tr}(b|V_0)$ for all $b \in B$, where b is thought as an endomorphism of the $\mathbb{C}$ vector space V_0.

Definition 3.3. An *integral structure* on a rational PEL datum $(B, *, V, \langle\ ,\ \rangle, h)$ is the additional choice of $\mathcal{O}$ and L where

- $\mathcal{O}$ an order in B which is stable under $*$.
- L is a lattice in V which is stable by $\mathcal{O}$ and such that for all $x, y \in L$

$$\langle x, y\rangle \in \mathbb{Z}(1)$$

an *integral PEL datum* is a tuple $(\mathcal{O}, *, L, \langle\ ,\ \rangle, h)$ consisting of a rational PEL datum with an integral structure.

What does this mean and why do we want a rational PEL datum?

Given any integral PEL datum, we can define a tuple

$$(A_0, \lambda_0, i_0, (\alpha_{0,n}, \nu_{0,n}))$$

where $A = L \otimes_{\mathbb{Z}} \mathbb{R}/L$ is an abelian variety, $\lambda_0 : A_0 \to A_0^\vee$ is a polarization of the abelian variety A_0, and $A_0^\vee$ is the dual abelian variety of A_0, $i_0 : \mathcal{O}_B \to \mathrm{End}_{\mathbb{C}}(A_0)$ is an endomorphism structure and $\alpha_{0,n} :\ L/nL \cong A_0[n]$ is a principal level n-structure. These structures on the abelian varieties A_0 forms the so called PEL structures, writing down an integral PEL strucutre is the same as writing down a tuple.

The moduli problem associated to an integral PEL datum is naturally defined over the reflex field F_0, then if p is a good prime, one can even work over $\mathcal{O}_{F_0,(p)}$, for studying rationality properties of automorphic forms, it is important to work over a number field like F_0 or a more integral variant. However, for the purposes of studying congruences, as we seen in the Katz's p-adic modular forms example, it is more convient to work over a p-adic base.

References

[Box15] George Boxer. Torsion in the coherent cohomology of shimura varieties and galois representations. *arXiv preprint arXiv:1507.05922*, 2015.

[Lan17] Kai-Wen Lan. An example-based introduction to Shimura varieties. *Preprint*, 2017.