Shimura varieties and the work of Langlands

J.S. Milne

November 18, 2009, v1.0

Abstract

Shimura varieties are initially defined as complex manifolds (quotients of Hermitian symmetric domains by congruence groups), but is known that they are algebraic varieties defined in a natural way over number fields. The oldest examples are the elliptic modular curves (quotients of the complex upper half plane by congruence subgroups of $\mathrm{SL}(2,\mathbb{Z})$). In the talk, I'll explain these two sentences, and also why Langlands was so interested in Shimura varieties.

Contents

These are my notes for a talk in the "What is a . . . ?" seminar, University of Michigan, November 17, 2009. They are available at `www.jmilne.org/math/`.

1 Shimura varieties as complex manifolds

a. Shimura curves

Every simply connected Riemann surfaces is isomorphic to the Riemann sphere, the complex plane, or the complex upper half plane (isomorphic to the open unit disk), and so every Riemann surface is a quotient of one of these by a discrete group. The Shimura curves are the quotients of the complex upper half plane by a congruence group, and so it remains for me to explain what I mean by a congruence group.

ELLIPTIC MODULAR CURVES

The group $\mathrm{SL}_2(\mathbb{R})$ acts on $D \stackrel{\text{def}}{=} \{z \in \mathbb{C} \mid \Im(z) > 0\}$ by

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

and $\mathrm{SL}_2(\mathbb{R})/\{\pm I\} \simeq \mathrm{Hol}(D)$ (the group of holomorphic automorphisms of D). The most obvious discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ is $\mathrm{SL}_2(\mathbb{Z})$. The next most obvious is the principal congruence subgroup of level N,

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \middle| \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}.$$

A congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup containing a principal congruence subgroup. Let Γ be a congruence subgroup whose $\bar{\Gamma}$ in $\mathrm{Hol}(D)$ is torsion-free. Then $\bar{\Gamma}$ acts freely on D and $\bar{\Gamma}\backslash D$ is a complex manifold, called an *elliptic modular curve*. This is a Shimura curve, but the elliptic modular curves aren't the only Shimura curves.

QUATERNIONIC SHIMURA CURVES

A quaternion algebra over a field F is an algebra of the form

$$\begin{aligned} B(a,b) &= F + Fi + Fj + Fk, \\ & i^2 = a, \quad j^2 = b, \quad ij = k = -ji, \quad ab \neq 0. \end{aligned}$$

When $F = \mathbb{R}$, there are exactly two quaternion algebras, namely, $B(1,1) \approx M_2(\mathbb{R})$ and $B(-1,-1) =$ Hamilton's quaternion algebra, which is a division algebra. The conjugate of a quaternion $\alpha = w + xi + yj + zk$ is $\bar{\alpha} = w - xi - yj - zk$, and its norm is

$$\mathrm{Nm}(\alpha) = \alpha\bar{\alpha} = w^2 - ax^2 - by^2 + abz^2.$$

Let B be a quaterion algebra over a totally real number field[1] F of degree $[F:\mathbb{Q}] = d$. There is an algebraic group G over $\mathbb{Q}$ such that, for any $\mathbb{Q}$-algebra R,

$$G(R) = \{\alpha \in B \otimes_{\mathbb{Q}} R \mid \mathrm{Nm}(\alpha) = 1\}.$$

By assumption $F \otimes_{\mathbb{Q}} \mathbb{R} \approx \mathbb{R}^d$. Correspondingly, $B \otimes_{\mathbb{Q}} \mathbb{R} \approx \mathbb{H}^{d-r} \times M_2(\mathbb{R})^r$, and

$$G_{\mathbb{R}} \approx (\text{compact group}) \times \mathrm{SL}_2(\mathbb{R})^r.$$

[1] By this I mean a field of the form $F = \mathbb{Q}[\alpha]$ where α is a root of polynomial in $\mathbb{Q}[X]$ whose roots are all real.

Therefore, when $r = 1$, there is a surjective homomorphism

$$G(\mathbb{R}) \to \mathrm{Hol}(D)$$

with compact kernel. The Shimura curves are exactly the quotients $\bar{\Gamma}\backslash D$ where Γ is a congruence subgroup of $G(\mathbb{Q})$ whose image $\bar{\Gamma}$ in $\mathrm{Hol}(D)$ is torsion-free.[2]

When $B = M_2(\mathbb{Q})$ we get back the elliptic modular curves. Otherwise B is a division algebra, and the curves are compact — they are called the *quaternionic Shimura curves*.

b. Congruence subgroups

Let G be a algebraic group over $\mathbb{Q}$. Choose an embedding of G into GL_n, and let $G(\mathbb{Z}) = G(\mathbb{Q}) \cap \mathrm{GL}_n(\mathbb{Z})$. A subgroup of $G(\mathbb{Q})$ is *arithmetic* if it is commensurable[3] with $G(\mathbb{Z})$, and an arithmetic subgroup is a *congruence* subgroup if it contains

$$\Gamma(N) = G(\mathbb{Q}) \cap \{A \in \mathrm{GL}_n(\mathbb{Z}) \mid A \equiv 1 \mod N\}$$

for some N. These definitions are independent of the choice of the embedding of G into GL_n.

Remarks

1.1. For any homomorphism $G \to G'$ of algebraic groups, the map $G(\mathbb{Q}) \to G'(\mathbb{Q})$ sends arithmetic groups to arithmetic groups, but the similar statement is far from true for congruence subgroups. In particular, the image in $\mathrm{PSL}_2(\mathbb{Q})$ of a congruence subgroup of $\mathrm{SL}_2(\mathbb{Q})$ is an arithmetic subgroup, but not usually a congruence subgroup.

1.2. The congruence subgroups define a topology on $G(\mathbb{Q})$. When G is a simply connected[4] semisimple algebraic group, the completion of $G(\mathbb{Q})$ with respect to this topology is $G(\mathbb{A}_f)$ where $\mathbb{A}_f$ is the ring of finite adèles,

$$\mathbb{A}_f = \mathbb{Q} \otimes \hat{\mathbb{Z}}, \quad \hat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}.$$

Using this, one can show that the congruence subgroups of $G(\mathbb{Q})$ are exactly the subgroups of the form

$$G(\mathbb{Q}) \cap (\text{compact open subgroup of } G(\mathbb{A}_f)).$$

1.3. The reason we allow only congruence subgroup is for the arithmetic — it doesn't matter over $\mathbb{C}$.

1.4. Let X be a nonsingular projective over $\mathbb{C}$. Then X has a model[5] over $\mathbb{Q}^{\mathrm{al}}$ (algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$) if and only if there exists a subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ of finite index such that $X(\mathbb{C})$ contains $\Gamma\backslash D$ as the complement of a finite set (Belyi). So if we allowed all arithmetic subgroups, we would be allowing all curves defined over $\mathbb{Q}^{\mathrm{al}}$.

1.5. Most arithmetic subgroups of $\mathrm{SL}_2(\mathbb{Z})$ are not congruence. To see this, note that $\mathrm{SL}_2(\mathbb{Z})/\Gamma(N) \simeq \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ has a filtration with uncomplicated quotients; in particular, the only simple quotients of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ are the groups $\mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$ where p is a prime dividing N. On the other hand, $\mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$ is the free group generated by an element of order 2 and an element of order 3, namely, Namely by S and ST where $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Therefore, any group generated by an element of order 2 and an element of order 3 is a quotient of $\mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$. Most finite simple groups, including the monster, satisfy this condition.

[2]For technical reasons, we also allow quotients $\Gamma'\backslash D$ where Γ' is a torsion-free subgroup of $\mathrm{Hol}(D)$ containing $\bar{\Gamma}$ as a subgroup of finite index.

[3]Two subgroups Γ and Γ' of a group are said to be commensurable if $\Gamma \cap \Gamma'$ is of finite index in both. For example, the subgroups $a\mathbb{Z}$ and $b\mathbb{Z}$ of $\mathbb{R}$ are commensurable if and only if $a/b \in \mathbb{Q}$. Commensurability is an equivalence relation.

[4]A connected algebraic group G is *simply connected* if there does not exist a connected algebraic group G' and a surjective homomorphism $G' \to G$ with finite nontrivial kernel. For example, SL_2 is simply connected as an algebraic group even though $\mathrm{SL}_2(\mathbb{R})$ is not simply connected as a Lie group.

[5]Let X be an algebraic variety over $\mathbb{C}$; by a *model* of X over a subfield L of $\mathbb{C}$, I mean an algebraic variety over L that gives back X by extension of scalars $L \to \mathbb{C}$.

c. Hermitian symmetric domains

A *bounded symmetric domain* is a bounded connected open subset of some space $\mathbb{C}^n$ such that every point is an isolated fixed point of an involution (holomorphic automorphism of order 2). A *hermitian symmetric domain* is any complex manifold isomorphic to a bounded symmetric domain. Bounded symmetric domains are simply connected.

For example, the open unit disk is a bounded symmetric domain, and the complex upper half plane is a hermitian symmetric domain (in fact, up to isomorphism, it is the only one-dimensional such domain).

Remark

1.6. There is a better (equivalent) definition of hermitian symmetric domain. A *hermitian space* is a smooth manifold M endowed with a riemannian structure g and a complex structure J that are compatible in the sense that $g(Jx, Jy) = g(x, y)$ for all $m \in M$ and all $x, y \in T_m M$.[6] A hermitian space is *symmetric* if every point is an isolated fixed point of an involution. A *hermitian symmetric domain* is a hermitian symmetric space with negative (sectional) curvature. [7]

CLASSIFICATION OF HERMITIAN SYMMETRIC DOMAINS

Every hermitian symmetric domain decomposes as a product of simple (irreducible) hermitian symmetric domains D whose automorphism groups $\mathrm{Hol}(D)$ are simple Lie groups. The *type* of the domain is defined to be the type of the Lie group $\mathrm{Hol}(D)$. The number of simple hermitian symmetric domains of each type is given by the following table:

A_n	B_n	C_n	D_n	E_6	E_7	E_8	F_4	G_2
n	1	1	3	2	1	0	0	0

d. Shimura varieties

Let D be a hermitian symmetric domain, and let $\mathrm{Hol}(D)^+$ be the identity component of $\mathrm{Hol}(D)$. Then $\mathrm{Hol}(D)^+$ is a connected semisimple Lie group with trivial centre. Consider a simply connected semisimple algebraic group G over $\mathbb{Q}$ and a surjective homomorphism

$$G(\mathbb{R}) \to \mathrm{Hol}(D)^+$$

with compact kernel (such pairs always exist, and are classified). Let Γ be a congruence subgroup in $G(\mathbb{Q})$ whose image $\bar{\Gamma}$ in $\mathrm{Aut}(D)^+$ is torsion-free. Then $\bar{\Gamma}$ acts freely D, and so the quotient $\bar{\Gamma}\backslash D$ is a complex manifold. We'll see shortly that it is, in fact, an algebraic variety. The Shimura varieties are the algebraic varieties that arise in this way. In other words, an algebraic variety over $\mathbb{C}$ is a Shimura variety if its universal covering space (in the topological sense) is a hermitian symmetric domain, and its fundamental group is the image of a congruence group as above.

DEFINITION 1.7. A *Shimura datum*[8] is a triple $(G, D, G(\mathbb{R}) \to \mathrm{Aut}(D)^+)$ with G a semisimple algebraic group G over $\mathbb{Q}$, D a hermitian symmetric domain, and $G(\mathbb{R}) \to \mathrm{Aut}(D)^+$ a surjective homomorphism with compact kernel.

[6] The reason for the name is that g is the real part of a hermitian form on the complex tangent space.

[7] A bounded symmetric domain has a hermitian metric that is invariant under all holomorphic automorphisms, namely, the Bergman(n) metric. This metric has negative curvature, and so every complex manifold isomorphic to a bounded symmetric domain is a hermitian symmetric domain in the sense of (1.6). Conversely, if D is a hermitian symmetric domain in the sense of (1.6), then the Harish-Chandra embedding realizes it as a bounded symmetric domain.

[8] Strictly speaking, this is a connected Shimura datum.

In what follows Γ will always be a subgroup of $G(\mathbb{Q})$ (arithmetic or congruence) and $\bar{\Gamma}$ will be its image in $\mathrm{Hol}(D)$. Also I usually write (G, D) for a Shimura datum.

2 Shimura varieties as algebraic varieties over $\mathbb{C}$

a. Chow's theorem

For a nonsingular variety X over $\mathbb{C}$, the set $X(\mathbb{C})$ of points of X has a natural structure of complex manifold. The functor $X \rightsquigarrow X(\mathbb{C})$ is faithful, but it is far from surjective on arrows or on objects. For example, $\mathbb{A}^1(\mathbb{C}) = \mathbb{C}$, and there are many nonpolynomial holomorphic maps $\mathbb{C} \to \mathbb{C}$, for example, the exponential functions. Moreover, a Riemann surface arises from an algebraic curve if and only if it can be compactified by adding a finite number of points. In particular, if a Riemann surface is an algebraic curve, then every bounded function is constant so, for example, the upper half plane is not an algebraic curve.

THEOREM 2.1 (CHOW 1949). *The functor $X \rightsquigarrow X(\mathbb{C})$ from nonsingular projective algebraic varieties to projective complex manifolds is an equivalence of categories.*

A complex manifold is *projective* if it can be realized as a closed submanifold of $\mathbb{P}^n(\mathbb{C})$ for some n, and similarly for algebraic varieties. Chow's theorem remains true when singularities are allowed.

b. The Baily-Borel theorem

THEOREM 2.2 (BAILY-BOREL 1966). *Let $X = \bar{\Gamma}\backslash D$ be the quotient of a hermitian symmetric domain by a torsion-free arithmetic group $\bar{\Gamma}$. Then X is an algebraic variety.*

The theorem is more precise. Its proof shows that the holomorphic automorphic forms on X define an embedding $X \hookrightarrow \mathbb{P}^n(\mathbb{C})$ for some n, and that the closure of the image is an algebraic variety containing the image as a Zariski open subset. The Baily-Borel theorem says that $\bar{\Gamma}\backslash D$ has a canonical structure as an algebraic variety; the next theorem implies that the algebraic structure is unique.

c. Borel's theorem

THEOREM 2.3 (BOREL 1972). *Let X be as in the preceding theorem, and let $f: V(\mathbb{C}) \to X(\mathbb{C})$ be a holomorphic map of complex manifolds where V is a nonsingular algebraic variety. Then f arises from a regular map $V \to X$ (i.e., a morphism of algebraic varieties).*

I prove Borel's theorem when V is a curve and X is the elliptic modular curve $\Gamma(N)\backslash D$ with N even. By Chow's theorem, it suffices to show that f extends to a map on the compactifications, and for this is suffices to show that f doesn't have an essential singularity at a point on the boundary of $V(\mathbb{C})$. Consider the composite

$$C(\mathbb{C}) \xrightarrow{f} \Gamma(N)\backslash D \longrightarrow \Gamma(2)\backslash D.$$

The Riemann surface $\overline{\Gamma(2)\backslash D}$ has genus zero and there are three cusps (boundary points). Therefore $\Gamma(2)\backslash D \approx \mathbb{C} \smallsetminus \{2 \text{ points}\}$, and so f can't have an essential singularity at a point of the boundary because this would violate the big Picard theorem (if f has an essential singularity at a point P, then it omits at most one value in any neighbourhood of P).

The big Picard theorem implies that any homomorphic map from a punctured disk to $\mathbb{P}^1(\mathbb{C}) \smallsetminus \{\text{three points}\}$ extends to a holomorphic map from the whole disk to $\mathbb{P}^1(\mathbb{C})$. Resolution of singularities (Hironaka) shows that V can be embedded in a nonsingular projective variety $\overline{V}$ in such a way that the boundary $\overline{V} \smallsetminus V$ is a divisor with normal crossings. This means that $\overline{V}(\mathbb{C}) \smallsetminus V(\mathbb{C})$ is locally a product of disks and punctured disks. Thus Borel's theorem generalizes the big Picard theorem in two respects: the punctured disk is replaced by a product of punctured disks and disks, and the target space is generalized from $\mathbb{P}^1 \smallsetminus \{\text{three points}\}$.

3 Abelian class field theory and complex multiplication

Let $\mathbb{Q}^{\text{al}}$ be an algebraic closure of $\mathbb{Q}$, for example, the algebraic closure of $\mathbb{Q}$ in $\mathbb{C}$, and let $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$ be the group of automorphisms of $\mathbb{Q}^{\text{al}}$ endowed with the Krull topology (that for which the open subgroups are exactly those fixing a finite extension of $\mathbb{Q}$). Then Galois theory tells us that the intermediate fields correspond to the closed subgroups of $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$. Unfortunately, we don't have a good description of $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$.

a. Abelian class field theory

However, we do have a good description of the largest Hausdorff abelian quotient $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})^{\text{ab}}$ of $\text{Gal}(\mathbb{Q}^{\text{al}}/\mathbb{Q})$. More generally, abelian class field theory provides us with a good description of $\text{Gal}(\mathbb{Q}^{\text{al}}/F)^{\text{ab}}$ for any algebraic number field F (i.e., finite extension of $\mathbb{Q}$ contained in $\mathbb{Q}^{\text{al}}$). Let F^{ab} be the composite of all finite abelian[9] extensions of F in $\mathbb{Q}^{\text{al}}$. Then $\text{Gal}(\mathbb{Q}^{\text{al}}/F)^{\text{ab}} \simeq \text{Gal}(F^{\text{ab}}/F)$.

The *ring of adèles* of F is

$$\mathbb{A}_F = F \otimes_{\mathbb{Q}} (\mathbb{R} \otimes \mathbb{A}_f).$$

It contains F as a subring (embedded diagonally). The *idèle group* of F is $\mathbb{A}_F^\times$. With the correct topology, it is a locally compact group.[10]

THEOREM 3.1. *There is a canonical surjective homomorphism*

$$\mathbb{A}_F^\times \to \text{Gal}(F^{\text{ab}}/F)$$

whose kernel is the closure of $F^\times \cdot \prod_{v \text{ complex}} F_v^\times \cdot \prod_{v \text{ real}} (F_v^\times)^+$.

The homomorphism is called *reciprocity* or *Artin map*. In particular, the finite abelian extensions of F in $\mathbb{Q}^{\text{al}}$ are in one-to-one correspondence with the open subgroups of $\mathbb{A}_F^\times$ containing the kernel of the Artin map. Unfortunately, we don't have an explicit way of constructing the field corresponding to an open subgroup, except in special cases.

For example, when $F = \mathbb{Q}$, then $\mathbb{A}_{\mathbb{Q}}^\times/(\text{kernel}) \simeq \widehat{\mathbb{Z}}^\times$, and the field corresponding to the subgroup $N\widehat{\mathbb{Z}}$ of $\widehat{\mathbb{Z}}$ is $\mathbb{Q}[\zeta_N]$, $\zeta_N = e^{2\pi i/N}$. Hilbert in the twelfth of his famous problems (ICM 1900) asked whether, for every number field F, there exist functions whose special values generate the finite abelian extensions of F.

This is still very open, but the theory of complex multiplication for elliptic curves provides a solution for quadratic imaginary fields (i.e., for fields of the form $\mathbb{Q}[\sqrt{-d}]$, $d \in \mathbb{Z}$, $d > 0$, d a nonsquare).

[9] An extension E of a field F is said to be *abelian* if it is Galois with abelian Galois group.

[10] Equivalently, let F_v be the completion of F at a prime v, and let $\mathcal{O}_v$ be the ring of integers in F_v when v is finite. Then

$$\mathbb{A}_F^\times = \textstyle\prod_v (F_v^\times : \mathcal{O}_v^\times) \stackrel{\text{def}}{=} \{(a_v) \in \prod F_v^\times \mid a_v \in \mathcal{O}_v^\times \text{ for almost all } v\}.$$

endowed with the topology for which $\prod_v \mathcal{O}_v^\times$ is an open subgroup with the product topology.

b. Complex multiplication for elliptic curves

Recall that the functor $C \rightsquigarrow C(\mathbb{C})$ defines an equivalence from the category of complete nonsingular curves over $\mathbb{C}$ onto the category of compact Riemann surfaces. The elliptic curves over $\mathbb{C}$ (curves of genus 1 with a distinguished point) are exactly the curves E such that $E(\mathbb{C}) \approx \mathbb{C}/\Lambda$ for some lattice Λ in $\mathbb{C}$. Let $E(\Lambda) = \mathbb{C}/\Lambda$. The Weierstrass $\wp$-function for Λ defines an embedding

$$z \mapsto (\wp(z):\wp'(z):1): E(\Lambda) \hookrightarrow \mathbb{P}^2(\mathbb{C})$$

whose image is the curve

$$Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3.$$

Let

$$j = \frac{1728g_2^3}{g_2^3 - 27g_3^2}.$$

Then $E(\Lambda) \approx E(\Lambda')$ if and only if $j(\Lambda) = j(\Lambda')$.

One shows that

$$\mathrm{Hom}(E(\Lambda), E(\Lambda')) = \{z \in \mathbb{C} \mid z\Lambda \subset \Lambda'\}.$$

From this it follows that $\mathrm{End}(E(\Lambda)) = \mathbb{Z}$ unless $\Lambda = \mathbb{Z}1 + \mathbb{Z}z$ where z generates a quadratic imaginary extension F of $\mathbb{Q}$. Then $\mathrm{End}(E(\Lambda))$ is a subring R of $\mathcal{O}_F$ of rank 2 over $\mathbb{Z}$, and E is said to have *complex multiplication* by the elements of R.

Let F be a quadratic imaginary extension of $\mathbb{Q}$, and let E have complex multiplication by $\mathcal{O}_F$. Then $j(E)$ generates the Hilbert class field of F (largest abelian extension of F in which all primes are unramified). Moreover, E is defined over $F(j)$, and, when F has no roots of 1 other than ± 1, F^{ab} is generated over $F(j)$ by the x-coordinates of the points of finite order on E.

We can restate this analytically. Let $j(z) = j(\mathbb{Z}1 + \mathbb{Z}z)$ for $\Im(z) > 0$. Then j is a holomorphic function on D invariant under $\Gamma(1)$, and it defines an isomorphism $\overline{\Gamma(1)\backslash D} \to \mathbb{P}^1(\mathbb{C})$. The Hilbert class field of F is generated over F by the value $j(\tau)$ of j at a generator τ of $\mathcal{O}_F$ as a $\mathbb{Z}$-algebra, and, when F has no roots of 1 other than ± 1, F^{ab} is generated over $F[j(\tau)]$ by the values $\wp(z/N)$ with $z \in \Lambda$.

Let $\mathcal{E}(N)$ be the set of isomorphism classes of pairs (E, e) with E an elliptic curve over $\mathbb{Q}^{\mathrm{al}}$ with complex multiplication by $\mathcal{O}_E$ and $e = (e_1, e_2)$ a basis for the group of points of order N on E.[11] Then $\mathcal{E}(N)$ is a finite set, and the action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{F})$ on it factors through a finite abelian quotient. From the Artin map we get an action of $\mathbb{A}_F^\times$ on $\mathcal{E}(N)$, and the main theorem of complex multiplication for elliptic curves describes this action explicitly. From this theorem, the statements in the preceding paragraphs follow.

c. Complex multiplication for abelian varieties

The theory of complex multiplication was extended to abelian varieties by Shimura, Taniyama, and Weil in the 1950s.[12]

A CM (complex multiplication) field F is a quadratic totally imaginary extension of a totally real field. Let $[F:\mathbb{Q}] = 2g$. An abelian variety A of dimension g is said to have complex multiplication by $\mathcal{O}_F$ if there is an injective homomorphism of rings $\mathcal{O}_F \to \mathrm{End}(A)$.

[11] The points of order N on E form a free $\mathbb{Z}/N\mathbb{Z}$-module of rank 2. The isomorphisms are required to respect the action of $\mathcal{O}_E$.

[12] In his collected works, Weil describes how, when he arrived at the famous Tokyo-Nikko conference in 1955 planning to speak on complex multiplication, he was disconcerted to find that two young Japanese mathematicians, Shimura and Taniyama, were planning to speak on the same topic.

The theory of complex muliplication attaches to F and A, a second CM-field F', called the reflex field. Let $\mathcal{A}(N)$ be the set of isomorphism classes of pairs (A,e) where A is an abelian variety over $\mathbb{Q}^{\mathrm{al}}$ and e is a basis for group of points of order N on A. Then $\mathcal{A}(N)$ is a finite set and the action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/F')$ on it factors through a finite abelian quotient. From the Artin map we get an action of $\mathbb{A}_{F'}^{\times}$ on $\mathcal{A}(N)$, and the main theorem of complex multiplication describes this action explicitly.

(Add on one paragraph on Hilbert 12th problem).

4 Shimura varieties as algebraic varieties over number fields

Fix a Shimura datum (G,D).

a. The base field

There is a base field $E=E(G,D)$ that always has an obvious definition. For example, suppose that G is defined by a quaternion algebra B over a totally real field F. The set of embeddings of F into $\mathbb{R}$ is partitioned into two sets depending on whether $B\otimes_F\mathbb{R}$ is $M_2(\mathbb{R})$ or $\mathbb{H}$. Let L be a Galois extension of $\mathbb{Q}$ in $\mathbb{C}$ containing all conjugates of F. Then $\mathrm{Gal}(L/\mathbb{Q})$ acts on $\mathrm{Hom}(F,\mathbb{R})=\mathrm{Hom}(F,L)$, and $E(G,D)$ is the fixed field of the subgroup of $\mathrm{Gal}(\bar{F}/\mathbb{Q})$ preserving the partition.

Each congruence subgroup of Γ defines an open subgroup of $\mathbb{A}_E^{\times}$, and $X(\Gamma)=\Gamma\backslash D$ is to be defined over the corresponding abelian extension of E.

b. Uniqueness

An algebraic variety over $\mathbb{C}$ may have no model over a subfield, or it may have many. For example, the distinct real curves

$$X^2+Y^2=1,\quad X^2-Y^2=1,\quad X^2+Y^2=1$$

become isomorphic over $\mathbb{C}$. This is only a problem with nonalgebraically closed fields, so that $X=X(\Gamma)$ will have at most one model over $\mathbb{Q}^{\mathrm{al}}$.

A model Y of $X(\Gamma)$ over $E(\Gamma)$ will be uniquely determined by the action of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E(\Gamma))$ on $Y(\mathbb{Q}^{\mathrm{al}})=X(\Gamma)(\mathbb{Q}^{\mathrm{al}})$. However, we can't specify such an action because we don't know how to name the elements of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/E(\Gamma))$. Instead, we define certain *special points* of $X(\Gamma)(\mathbb{Q}^{\mathrm{al}})$ and require that the model satisfies a condition of the following form:

> each special point $P\in Y(\mathbb{Q}^{\mathrm{al}})$ lies in an abelian extension of a specific field $E(P)$ and $\mathrm{Gal}(E(P)^{\mathrm{ab}}/E(P)$ acts on it according to a specific rule.

The special points are Zariski dense, and it is known that the condition determines the canonical model uniquely.

For Shimura varieties that parametrize, in a natural way, abelian varieties with additional structure, the theory of complex multiplication shows that the moduli variety satisfies the condition.

c. Existence

There are three cases.

(a) The Shimura variety is, in a natural way, a moduli variety for abelian varieties with additional structure.
(b) The Shimura variety is a moduli variety for abelian varieties with additional structure, but not in any natural or obvious way.
(c) The Shimura variety is not a moduli variety for abelian varieties (and is not known to be a moduli variety at all).

Shimura varieties of types A, B, C and some of type D lie in case (a,b); the remainder lie in case (c).

In case (a,b), $X(\Gamma)$ represents a functor on schemes over $\mathbb{C}$. The functor is defined over $E(\Gamma)$, and so is represented by a model of $X(\Gamma)$ over $E(\Gamma)$ by descent theory (Shimura, Deligne).

Case (c) is more difficult, but the canonical models are known to exist in this case also (Borovoi, Milne).

SHIMURA CURVES

The general strategy is to choose a representation $G \hookrightarrow \mathrm{GL}(V)$ over $\mathbb{Q}$. Then each point x of D defines a Hodge structure on $V \otimes \mathbb{R}$, and one hopes that the V together with this Hodge structure is the H_1 of an abelian variety. For curves defined by quaternion algebras over $F \neq \mathbb{Q}$, this is not true but, miraculously, when you tensor the Hodge structure with another Hodge structure (also not the H_1 of an abelian variety), it becomes the H_1 of an abelian variety (Shimura).

d. Conclusion

For each Shimura datum (G, D) and each congruence subgroup of $G(\mathbb{Q})$, there is a well-defined algebraic variety $X(\Gamma)$ over well-defined number field $E(\Gamma)$ such that $X(\Gamma)(\mathbb{C}) \simeq \overline{\Gamma}\backslash D$. Usually one now calls $X(\Gamma)$ the Shimura variety attached to (G, D, Γ).

(Expand this section by about a page.)

5 Shimura varieties in the work of Langlands

Langlands was interested Shimura varieties for a number of reasons, of which I will mention only two.

a. Nonabelian class field theory (the global Langlands conjecture)

Recall that abelian class field theory for $\mathbb{Q}$ provides us with a surjective homomorphism $\mathrm{GL}_1(\mathbb{A}) \to \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q})$. A homomorphism $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}) \to \mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^\times$ factors through $\mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q})$, and so defines a homomorphism $\mathrm{GL}_1(\mathbb{A}) \to \mathbb{C}^\times$. In other words, the homomorphisms $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}) \to \mathrm{GL}_1(\mathbb{C})$ parametrize the characters of $\mathrm{GL}_1(\mathbb{A})$. Roughly speaking,[13] the global Langlands conjecture says that the homomorphisms $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{C})$ should parametrize the irreducible automorphic representations of $\mathrm{GL}_n(\mathbb{A})$. More generally, for a reductive group G over $\mathbb{Q}$, it says that the homomorphisms $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q}) \to G^\vee(\mathbb{C})$ should parametrize the irreducible automorphic representations of $G(\mathbb{A})$ ($G^\vee$ is the "Langlands dual" of G). This can be regarded as the long-sought nonabelian class field theory.

How could one prove such a correspondence? Let $\mathcal{G} = \mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$. How can we show that the representations of $\mathcal{G}$ parametrize certain of the representations of $G(\mathbb{A})$? An easy case would be if

[13] This subsection is an over simplification; nothing in the Langlands program is as simple as one would like.

there is a homomorphism $G(\mathbb{A}) \to \mathcal{G}$. That can't happen our case (except when $G = \mathrm{GL}_1$) because, for example, the representations of $\mathrm{Gal}(\mathbb{Q}^{\mathrm{al}}/\mathbb{Q})$ are finite-dimensional whereas the automorphic representations are infinite-dimensional. Another idea is to find a very large vector space V on which $\mathcal{G} \times G(\mathbb{A})$ acts. Then given a representation π of $\mathcal{G}$, we can define $V(\pi) = \mathrm{Hom}(\pi, V)$ — this will be stable under $G(\mathbb{A})$, and so will be a representation of $G(\mathbb{A})$.

Where can we find such V? Fix a Shimura datum (G, D). For each congruence group Γ, we get a Shimura variety $X(\Gamma)$ over a finite extension of $\mathbb{Q}$. We can regard $X(\Gamma)$ as a variety over $\mathbb{Q}$, and take the étale cohomology of $X(\Gamma)_{\mathbb{Q}^{\mathrm{al}}}$. This a vector space on which $\mathcal{G}$ acts. The group $G(\mathbb{Q})$ doesn't act on the individual varieties $X(\Gamma)$, but it does act on the whole family (if $g \in G(\mathbb{Q})$, then $g\Gamma g^{-1}$ is also a congruence group, and $x \mapsto gx : D \to D$ defines a holomorphic (hence algebraic) map $\Gamma\backslash D \to g\Gamma g^{-1}\backslash D$). This action on the family $(X(\Gamma))$ is defined over $\mathbb{Q}$, and so gives an action on $V \stackrel{\text{def}}{=} \varinjlim_{\Gamma} H^*_{\text{et}}(X(\Gamma))$ which commutes with $\mathcal{G}$. In fact, this action is continuous for the topology defined by the congruence subgroups, and so extends to the completion, which is $G(\mathbb{A}_f)$. So we have a large vector space on which $\mathcal{G} \times G(\mathbb{A}_f)$ acts.

(To be continued.)

When F is replaced a global field of nonzero characteristic, Drinfeld defined analogues of Shimura varieties (Drinfeld modular varieties) and proved the Langlands correspondence for GL_2 in this way, and Lafforgue extended his results to GL_n (both received the Fields medal for their work). When F is replaced by a local field, Harris and Taylor proved in 1998 that the Langlands local correspondence for GL_n can indeed be realized in the étale cohomology of Shimura varieties.

b. Zeta functions

Recall that Riemann's zeta function is

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}, \quad s \in \mathbb{C}, \quad \Re(s) > 1.$$

The product converges for $\Re(s) > 1$, and it is known that $\zeta(s)$ extends to a meromorphic function on the whole complex plane with a simple pole at $s = 1$. Moreover, it satisfies a functional equation relating $\zeta(s)$ and $\zeta(1-s)$.

To any algebraic variety X over $\mathbb{Q}$, one can attach a zeta function $\zeta(X, s)$. Loosely speaking, it can be thought of as a generating function for the number of points of X in any finite field. It is known that

$$\zeta(X, s) = \prod_{p \text{ prime}} \zeta(X_p, s) \tag{1}$$

where each $\zeta(X_p, s)$ is *finite* product of factors $(1 - ap^{-s})^{\pm 1}$ with $|a| \leq p^{\dim X}$. For example, when X is a point, $\zeta(X, s)$ is exactly Riemann's zeta function. By comparing $\zeta(X, s)$ with $\zeta(s)$, one sees that (1) converges for $\Re(s) > \dim X + 1$. The Hasse-Weil conjecture asserts that $\zeta(X, s)$ can be continued analytically to the whole complex plane and satisfies a functional equation relating $\zeta(X, s)$ to $\zeta(X, d + 1 - s)$.

This is known in the following main cases:

- ◇ Abelian varieties with complex multiplication (Shimura and Taniyama in the 1950s). In this case, one shows that the zeta function is an alternating product Hecke L-functions.
- ◇ Elliptic modular curves (Eichler and Shimura). In this case, the zeta function is the Mellin transform of a modular form.
- ◇ Elliptic curves (Wiles et al). In this case, the zeta function was shown to be a factor of the zeta function of an elliptic modular curve (hence again the Mellin transform of a modular form).

In each case, the Hasse-Weil conjecture is proved by identifying the zeta function with another function about which one knows something.

Langlands attaches to each reductive group over $\mathbb{Q}$ an array of "automorphic L-functions". The above results say that the zeta functions of abelian varieties with complex multiplication are alternating products of automorphic L-functions for GL_1 and that those of elliptic curves and elliptic modular curves are automorphic L-functions for GL_2.

It is part of Langlands's philosophy that the zeta function of *every* algebraic variety over $\mathbb{Q}$ should be an alternating product of automorphic L-functions. However, for an arbitrary variety one has no idea how to approach this problem. Since a Shimura variety is defined by an algebraic group, one at least has a candidate for the group to which the L-functions are attached. Langlands has a conjecture that expresses the zeta function of the Shimura variety defined by G in terms of the automorphic L-functions of G and of certain auxiliary groups (the endoscopic groups of G).

With Ngo's proof of the Fundamental Lemma, there appears to some hope of proving this conjecture for compact Shimura varieties (at least, if one ignores the factors of the zeta function corresponding to bad primes), and also the Hasse-Weil conjecture for the same varieties.

(Not to be continued.)

References

Much of the material originates in Shimura's books and papers. My approach follows that in:

Deligne, Pierre. Travaux de Shimura. (French) Séminaire Bourbaki, 23ème année (1970/71), Exp. No. 389, pp. 123–165. Lecture Notes in Math., Vol. 244, Springer, Berlin, 1971. MR0498581

Deligne, Pierre. Variétés de Shimura: interprétation modulaire, et techniques de construction de modèles canoniques. (French) Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, pp. 247–289, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979. MR0546620

A more extended account can be found in:

Milne, J. S. Introduction to Shimura varieties. Harmonic analysis, the trace formula, and Shimura varieties, 265–378, Clay Math. Proc., 4, Amer. Math. Soc., Providence, RI, 2005. MR2192012

Available on the Clay website and on mine (under articles 2005).