WHAT IS...

a Shimura Variety?

James S. Milne

Most mathematicians have encountered modular functions. For example, when the group theorists discovered the monster group, they were surprised to find that the degrees of its irreducible representations were already encoded in the q-coefficients of the j-function. The theory of Shimura varieties grew out of the applications of modular functions and modular forms to number theory. Roughly speaking, Shimura varieties are the varieties on which modular functions live.

Shimura Curves

According to the uniformization theorem, every simply connected Riemann surface is isomorphic to the Riemann sphere, the complex plane, or the open unit disk (equivalently the complex upper half plane D_1). The Shimura curves are the quotients of D_1 by the actions of certain discrete groups, which I now describe.

The action $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) z = \frac{az+b}{cz+d}$ of $\mathrm{SL}_2(\mathbb{R})$ on D_1 realizes $\mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ as the group of holomorphic automorphisms $\mathrm{Hol}(D_1)$ of D_1. Let B be a quaternion algebra over a totally real number field F such that $\mathbb{R} \otimes_F B$ is isomorphic to $M_2(\mathbb{R})$ for exactly one embedding of F into $\mathbb{R}$, and let G be the algebraic group over $\mathbb{Q}$ whose R-points for any $\mathbb{Q}$-algebra R are the elements of $B \otimes_{\mathbb{Q}} R$ of norm 1. Then $G(\mathbb{R})$ is a product of $\mathrm{SL}_2(\mathbb{R})$ with a compact group, and so there is a surjective homomorphism $\varphi\colon G(\mathbb{R}) \to \mathrm{Hol}(D_1)$ with compact kernel. A Shimura curve is the quotient of D_1 by the image in $\mathrm{Hol}(D_1)$ of a congruence subgroup of $G(\mathbb{Q})$.

For example, when $B = M_2(\mathbb{Q})$, the group G is SL_2, and we get the familiar elliptic modular curves, namely, the quotients of D_1 by a discrete subgroup Γ of $\mathrm{Hol}(D_1)$ containing the image of a principal congruence subgroup

$$\Gamma(N) \overset{\text{def}}{=} \{A \in \mathrm{SL}_2(\mathbb{Z}) \mid A \equiv I \bmod N\}.$$

In this case, the Riemann surface $\Gamma\backslash D_1$ can be compactified in a natural way by adding a finite number of points (called the cusps), and so $\Gamma\backslash D_1$ has a unique structure of an algebraic curve compatible with its analytic structure. In all other cases, $\Gamma\backslash D_1$ is compact and so is automatically an algebraic curve.

Each Shimura curve has a natural embedding in projective space. Consider, for example, the elliptic modular curve $Y(N) \overset{\text{def}}{=} \Gamma(N)\backslash D_1$. A cusp form of weight $2k$ for $\Gamma(N)$ is a holomorphic function f on D_1 vanishing at the cusps and such that

(1)
$$f(Az) = (cz+d)^{2k} \cdot f(z) \text{ for all } A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma(N).$$

For some fixed k, a basis $f_0, \ldots, f_n$ for the cusp forms of weight $2k$ defines an embedding

$$P \mapsto (f_0(P) \colon \ldots \colon f_n(P)) \colon Y(N) \to \mathbb{P}^n(\mathbb{C})$$

of $Y(N)$ as an algebraic subvariety of $\mathbb{P}^n(\mathbb{C})$. In fact, we can do better. Each of the cusps is fixed by a unipotent matrix $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$ with h a positive integer. For such a matrix, (1) becomes $f(z) = f(z+h)$, and so a cusp form f has an expansion

$$(2)\quad f(z) = a_1 q + a_2 q^2 + a_3 q^3 + \cdots, \quad q = e^{2\pi i/h}.$$

Let $\mathbb{Q}[\zeta_N]$ be the field generated over $\mathbb{Q}$ by a primitive Nth root of 1. It is possible to choose the basis $f_0, \ldots, f_n$ so that the coefficients in (2) lie in the field $\mathbb{Q}[\zeta_N]$ and, when this is done, the homogeneous polynomials defining $Y(N)$ have coefficients in $\mathbb{Q}[\zeta_N]$. Thus $Y(N)$ has a canonical model over $\mathbb{Q}[\zeta_N]$. This property of $Y(N)$ is very unusual. Typically, an algebraic variety over $\mathbb{C}$ will not have a model over an algebraic number field, and when it does, it will have many distinct models, none of which is to be preferred.

The above explanation for why $Y(N)$ has a canonical model over $\mathbb{Q}[\zeta_N]$ is that of the analysts.

James S. Milne is emeritus professor of mathematics at the University of Michigan. His email address is jmilne@umich.edu.

DOI: http://dx.doi.org/10.1090/noti923

The geometers have an entirely different explanation. For an elliptic curve E over $\mathbb{C}$, the group $E(\mathbb{C})_N$ of elements of order N is a free $\mathbb{Z}/N\mathbb{Z}$-module of rank 2 equipped with a skew-symmetric pairing e_N taking values in the group of Nth roots of 1 in $\mathbb{C}$. A level-N structure on E is a basis (t_1, t_2) for $E(\mathbb{C})_N$ such that $e_N(t_1, t_2) = \zeta_N$. For $N \geq 3$, the canonical model of $Y(N)$ represents the functor sending a $\mathbb{Q}[\zeta_N]$-algebra R to the set of isomorphism classes of elliptic curves over R equipped with a level-N structure.

Both explanations fail when $B \neq M_2(\mathbb{Q})$: then the curves are compact, so there are no cusps and no q-expansions, and they are not moduli varieties in any natural way. Thus both the analysts and the geometers were surprised when Shimura (in 1967) proved that *all* these curves do have canonical models over specific number fields. The theory of Shimura varieties, as distinct from the theory of moduli varieties, can be said to have been born with Shimura's paper. Ihara (in 1968) attached Shimura's name to these curves.

Shimura Varieties, according to Shimura

A complex manifold is symmetric if each point is an isolated fixed point of an involution. For example, D_1 is symmetric because it is homogeneous and i is an isolated fixed point of the involution $z \mapsto \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) z = -1/z$. A connected symmetric complex manifold is called a hermitian symmetric domain if it is isomorphic to a bounded open subset of $\mathbb{C}^n$ for some n. Every hermitian symmetric domain is simply connected, and so the Riemann mapping theorem shows that D_1 is the only hermitian symmetric domain of dimension one. The connected Shimura varieties are the quotients of hermitian symmetric domains by the actions of certain discrete groups, which I now describe.

The group of holomorphic automorphisms of a hermitian symmetric domain D is a semisimple Lie group whose identity component we denote by $\mathrm{Hol}(D)^+$. To define a family of connected Shimura varieties covered by D, we need a semisimple algebraic group G over $\mathbb{Q}$ and a surjective homomorphism $G(\mathbb{R}) \to \mathrm{Hol}(D)^+$ with compact kernel. The connected Shimura varieties are then the quotients $\Gamma\backslash D$ of D by a torsion-free subgroup Γ of $\mathrm{Hol}(D)^+$ containing the image of a congruence subgroup of $G(\mathbb{Q})$ as a subgroup of finite index.

Baily and Borel proved that, as in the curve case, the modular forms on D relative to Γ embed $\Gamma\backslash D$ as an algebraic subvariety of some projective space. Thus, each manifold $\Gamma\backslash D$ has a canonical structure as an algebraic variety over $\mathbb{C}$, and a later theorem of Borel shows that the algebraic structure is in fact unique.

Shimura introduced the notion of a canonical model for these algebraic varieties. This is a model of the variety over a specific number field that is uniquely determined by specifying the fields generated by the coordinates of certain special points. Shimura and his students Miyake and Shih proved the existence of canonical models for several fundamental families of connected Shimura varieties.

Shimura Varieties, according to Deligne

When Deligne was asked to report on Shimura's work in a 1971 Bourbaki seminar, he rewrote the foundations. For Deligne, a Shimura variety is defined by a reductive group G over $\mathbb{Q}$ and a $G(\mathbb{R})$-conjugacy class of homomorphisms $h\colon \mathbb{C}^\times \to G(\mathbb{R})$ satisfying certain axioms. The Shimura variety itself is a certain double coset space. The axioms ensure that, on the one hand, this double coset space is a finite disjoint union of the varieties considered in the preceding section and on the other hand, that it is the base space for a variation of Hodge structures. Sometimes the variation of Hodge structures arises from a family of abelian varieties, in which case the existence of a canonical model follows from the theory of moduli varieties. In other cases, Deligne was able to prove the existence of a canonical model by relating the Shimura variety to one that is a moduli variety. In the remaining cases, the existence of a canonical model was proved by the author and Borovoi (somewhat independently).

Shimura varieties interested Langlands as a source of Galois representations and as a test for his idea that all zeta functions are automorphic. In a 1974 lecture he introduced the term "Shimura variety" for the varieties defined by Deligne. Once the existence (and uniqueness) of their canonical models had been demonstrated, it became customary to refer to the canonical model as the Shimura variety (rather than the variety over $\mathbb{C}$). The connected components of these varieties are the canonical models of the preceding section.

Further Reading

For Shimura's approach, I suggest looking first at his notes *Automorphic Functions and Number Theory*[1] and his ICM talks. For Deligne's approach there are the difficult original articles of Deligne[2] and the author's *Introduction to Shimura Varieties.*[3]

[1] *Lecture Notes in Mathematics, No. 54, Springer, 1968.*

[2] *Travaux de Shimura. Séminaire Bourbaki (1970/71), Exp. No. 389; Variétés de Shimura, Proc. Sympos. Pure Math., 33, AMS, 1979, pp. 247–289.*

[3] *Clay Math. Proc., 4, AMS, 2005, pp. 265–378.*