SHIMURA CURVES LECTURE 10: QUATERNIONIC MODULI

PETE L. CLARK

Let $B/\mathbb{Q}$ be an indefinite quaternion algebra of discriminant D (we allow the case of $D = 1$, i.e., $B \cong M_2(\mathbb{Q})$) and $\mathcal{O}_N$ a level N Eichler order in B. By taking $\Gamma(\mathcal{O}_N) = \Gamma(B, \mathcal{O}_N)$ to be the elements of $\mathcal{O}^\times$ of positive reduced norm,[1] we get an arithmetic Fuchsian group and hence a Riemann surface $\mathcal{O}_N \backslash \mathcal{H}$. So as not to prejudice matters, let us temporarily denote the Shimura curve associated to any quaternion order $\mathcal{O}$ by $X(\mathcal{O})$.[2] Let N be a positive integer which is prime to D. By definition a level N Eichler order is the intersection of two maximal orders, $\mathcal{O}_N = \mathcal{O} \cap \mathcal{O}'$. Here N can be characterized in any of the following ways:

(i) The discriminant of $\mathcal{O}_N$ is $N \cdot D$.
(ii) N is the common index $[\mathcal{O} : \mathcal{O}_N] = [\mathcal{O}' : \mathcal{O}_N]$.

Let $N = \prod_p p^{n_i} = \prod_p N_p$. Then the completion of $\mathcal{O}_N$ at p is equal to the intersection of the standard maximal order $(\mathcal{O})_p = M_2(\mathbb{Z}_p)$ of $B_p = M_2(\mathbb{Q}_p)$ with the order

$$(\mathcal{O}_N)_p = \begin{bmatrix} \mathbb{Z}_p & p^n\mathbb{Z}_p \\ p^{-n}\mathbb{Z}_p & \mathbb{Z}_p \end{bmatrix},$$

the conjugate of $M_2(\mathbb{Z}_p)$ by the element

$$w_{N_p} = \begin{bmatrix} p^n & 0 \\ 0 & 1 \end{bmatrix}.$$

As we saw in our study of global orders, for two maximal orders $\mathcal{O}$ and $\mathcal{O}'$ of B, there exists an element of $B^\times$ which we shall denote w_N such that $\mathcal{O}' = w_N \mathcal{O} w_N^{-1}$. (In the split case, we can indeed take $w_N = \prod w_{N_p}$.) The reduced norm of such an element w_N is well-determined up to a square, and we may choose w_N so as to lie in $\mathcal{O}$ and have reduced norm N. When n_p is odd this is called an **Atkin-Lehner involution**.

We wish to address the following question: what is the modular intepretation of $X(\mathcal{O}_N)$ and of the two maps

$$f_1 : X(\mathcal{O}_N) \to X(\mathcal{O}),\ f_2 : X(\mathcal{O}_N) \to X(w_N \mathcal{O} w_N^{-1}) \ \ ?$$

An excellent start is the observation that when $D = 1$, the order $\mathcal{O}_N$ is precisely $\begin{bmatrix} \mathbb{Z} & \mathbb{Z} \\ N\mathbb{Z} & \mathbb{Z} \end{bmatrix}$, so that $\Gamma(\mathcal{O}_N) = \Gamma_0(N)$. This gives us the idea that the curve $XX(\mathcal{O}_N)$ should parameterize $\mathcal{O}$-QM abelian surfaces with some kind of "quaternionic $\Gamma_0(N)$ level structure," which is the right idea. (In fact, in earlier lectures

[1] I now allow myself to refer to a discrete subgroup of $GL_2(\mathbb{R})^+$ as a Fuchsian group; such a thing acts on $\mathcal{H}$ and the action is effective upon passing to the quotient by ± 1.

[2] Note that in the split case $B = M_2(\mathbb{Q})$, we are getting a noncompactified modular curve, which would be more traditionally denoted by Y "of something." We will not do this here.

we have denoted $X(\mathcal{O}_N)$ by $X_0^D(N)$ and $\Gamma(\mathcal{O}_N)$ by $\Gamma_0^D(N)$, so to a certain extent we have presupposed the explanation which we are now giving.) However, our task here is to make this precise, and also to resolve the following issue: we also have an interpretation of $X(\mathcal{O}_N) = \Gamma(\mathcal{O}_N)\backslash\mathcal{H}$ as a moduli space of abelian surfaces of the form $\mathbb{C}^2/\Phi(\mathcal{O}_N)\begin{bmatrix} \tau \\ 1 \end{bmatrix}$, where $\Phi : B \hookrightarrow B \otimes \mathbb{R} = M_2(\mathbb{R})$, in other words as $\mathcal{O}_N$-QM abelian surfaces.

Thus, what we really want to understand is why an $\mathcal{O}_N$-QM structure is equivalent to an $\mathcal{O}$-QM structure together with a quaternionic $\Gamma_0(N)$-level structure (and to define the latter). Because we can view $\mathcal{O}_N^\times$ as defining a compact open subgroup of $B^\times(A_f)$ – namely we take the adelic points whose component at p is the units of the local order $(\mathcal{O}_N)_p$ defined above, it is *a priori* clear that $X(\mathcal{O}_N)$ can be viewed as some kind of partial level N-structure in the adelic sense: namely, an orbit of $B \otimes \hat{\mathbb{Z}}$-equivariant isomorphisms of $\widehat{\mathcal{O}} = \mathcal{O} \otimes \hat{\mathbb{Z}}$ with the full Tate module $TA = \prod_\ell T_\ell(A)$. To give a full level N structure is equivalent giving an $\mathcal{O}$-equivariant isomorphism from $A[N]$ to $\widehat{\mathcal{O}} \otimes \mathbb{Z}/N\mathbb{Z} \cong \mathcal{O} \otimes \mathbb{Z}/N\mathbb{Z}$; note that the latter is isomorphic to $M_2(\mathbb{Z}/N\mathbb{Z})$ since N is divisible only by split primes of B. Thus full level N-structures are parameterized by $M_2(\mathbb{Z}/N\mathbb{Z})^\times \cong GL_2(\mathbb{Z}/N\mathbb{Z})$, just as in the elliptic modular case. Indeed, as long as N is prime to D, all the group theory is the same as in the $D = 1$ case, and $\Gamma(\mathcal{O}_N)$ is a (non-normal!) subgroup of $\Gamma(\mathcal{O})$ of (projective) index equal to the (projective) index of $\Gamma_0(N)$ in $\Gamma(1)$ in the classical modular case.

Exactly what is a $\Gamma_0(N)$ level structure? In the case that $D = 1$, we are looking at squares of elliptic curves $E \times E$, and the reasonable thing to take is a subgroup of the form $Q = C \times C$, where C is an order N cyclic subgroup of E. It is not hard to see that if we do this, then Q has the merit of being stabilized by $M_2(\mathbb{Z})$; in particular, $E \times E/Q$ carries a natural QM structure. In fact, as far as the N-torsion is concerned, the picture is identical in the QM case:

Let (A, ι) be an abelian surface equipped with an embedding $\iota : \mathcal{O} \hookrightarrow \mathrm{End}(A)$. The elements of $\iota(\mathcal{O})$, like any endomorphisms of any abelian group, preserve the N-torsion, so we have a representation $\mathcal{O} \hookrightarrow \mathrm{End}(A) \cong M_4(\mathbb{Z}/N\mathbb{Z})$, which factors through $\mathcal{O} \otimes \mathbb{Z}/N\mathbb{Z} \cong M_2(\mathbb{Z}/N\mathbb{Z})$. Thus, the essential data is a certain homomorphism $M_2(\mathbb{Z}/N\mathbb{Z}) \to \mathrm{End}(W)$, where $W = A[N]$ is a four-dimensional free $\mathbb{Z}/N\mathbb{Z}$-module. Up to equivalence, there is only one such homomorphism (an instance of *Morita equivalence* from the category of modules over a ring R to the category of modules over $M_n(R)$, but we can make this perfectly explicit here.) Let e_1, e_2 be the two standard idempotents for $M_2(\mathbb{Z}/N\mathbb{Z})$; then putting $V_i = e_i W$ we get a direct sum decomposition $W = V_1 \oplus V_2$ under which $M_2(\mathbb{Z}/N\mathbb{Z})$ acts in the obvious way (i.e., as a 4×4 matrix partitioned into four 2×2 blocks, each of which consists of a scalar matrix).

Let $\varphi : (A_1, \iota_1) \to (A_2, \iota_2)$ be an isogeny which respects the QM-structure. Equivalently, there exists a positive integer N such that φ is given by modding out by a finite subgroup $Q \subset A[N]$, where Q is $\mathcal{O}$-stable. By the above discussion, it is not possible for Q to have $\mathbb{Z}/N\mathbb{Z}$-rank 1: the proper, nontrivial $\mathcal{O}$-stable subspaces all have rank 2. Thus it makes sense to define an isogeny as being **QM-cyclic** if its kernel is $\mathcal{O}$-stable and isomorphic as an abelian group to $\mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$. More precisely, under the Morita equivalence, a QM-cyclic subgroup $Q \subset A[N]$ will be of the form $e_1Q \oplus e_2Q = C_1 \oplus C_2$, where C_i is an honestly cyclic subgroup of V_i. The data of Q and C_1 (or indeed also C_2) are in fact equivalent: $C_1 = e_1Q$ and $Q = \mathcal{O} \cdot C_1$. Thus, a QM-cyclic degree N isogeny determines, and is determined by, an order N cyclic subgroup C of V_1. (Note however that not any order N subgroup of $A[N]$ arises this way: a cardinality argument shows that it is in fact more likely that the $\mathcal{O}$-module generated by a point of order N is all of $A[n]$.) Moreover, it is easily seen that the subgroup of $\mathcal{O}^\times$ stabilizing a particular subgroup $C \subset V_1$ is precisely the group $\mathcal{O}_N^\times$. Thus we have proved the following

Proposition 1. *The curve $X(\mathcal{O}_N) = \Gamma(\mathcal{O}_N)\backslash\mathcal{H}$ can be viewed as the moduli space for either of the following structures:*
(M1) Triples (A, ι, Q), where A is an abelian surface, $\iota : \mathcal{O} \hookrightarrow \mathrm{End}(A)$ is an $\mathcal{O}$-QM structure, and $Q \subset A[N]$ is an $\mathcal{O}$-stable subgroup of order N^2.
(M2) Equivariant isogenies $\varphi : (A_1, \iota_1) \to (A_2, \iota_2)$ with QM-cyclic kernel.

Because of this proposition, we feel justified in reverting to the old notation $\Gamma_0^D(N)$ for the units in $\mathcal{O}_N$ and $X_0^D(N) = \Gamma_0^D(N)\backslash\mathcal{H}$.

Okay, but what does this have to do with the $\mathcal{O}_N$-QM abelian surfaces constructed analytically above? We have a map

$$A = \mathbb{C}^2/\Phi(\mathcal{O}_N)\begin{bmatrix} \tau \\ 1 \end{bmatrix} \to \mathbb{C}^2/\Phi(\mathcal{O})\begin{bmatrix} \tau \\ 1 \end{bmatrix} = A'$$

whose kernel is isomorphic to $\mathcal{O}/\mathcal{O}_N$, i.e., cyclic of order N. It should now be clear what's going on: $\mathcal{O}_N$ is the subring of $\mathrm{End}^0(A) \cong \mathrm{End}^0(A')$ which stabilize a cyclic subgroup C_1 and hence become well-defined on the quotient A'. On the other hand, there is *another* degree N isogeny from A to the abelian surface $A'' = \mathbb{C}^2/\Phi(w_N\mathcal{O}w_N^{-1})\begin{bmatrix} \tau \\ 1 \end{bmatrix}$, with corresponding kernel C_2. Again $\mathcal{O}_N$ is the subring of endomorphisms stabilizing this cyclic subgroup. These two degeneracy maps $f_1, f_2 : X_0^D(N) \to X^D$ differ by the automorphism $q : X_0^D(N) \to X_0^D(N)$ given by conjugating elements of $\mathcal{O}_N$ by w_N. What we have then, is that the picture

$$A' \xleftarrow{f_1} A \xrightarrow{f_2} A''$$

is a symmetric version of the previous interpretation of $X_0^D(N)$. Indeed, both of the composite isogenies $f_2 \circ f_1^\vee : A' \to A''$, $f_1^\vee \circ f_2 : A'' \to A'$ are QM-cyclic and mutually dual (here we are using the fact that the canonical polarization on an $\mathcal{O}$-QM surface is principal, so certain details are being elided), which gives the modular interpretation of the automorphism w_N of $X_0^D(N)$.

Remark: In the (few!) places where this is explained in the literature, one usually finds primacy given to the group C_1 instead of Q. From the perspective of my thesis work – which studies principally polarized abelian surfaces with quaternionic multiplication defined over an extension field (analogous to studying CM elliptic curves over $\mathbb{Q}$) – the group Q is more important, since one needs to have the $\mathcal{O}$-action rationally defined in order to define the projection $C_1 = e_1Q$, but there are curves – namely, quotients of $X_0^D(N)$ by Atkin-Lehner involutions at primes dividing D, which we have not yet discussed – whose natural moduli interpretation involves Q. However:

Exercise: Define the subgroup $\Gamma_1^D(N)$ and the curve $X_1^D(N) = \Gamma_1^D(N)\backslash\mathcal{H}$. Show that its moduli interpretation is a tuple (A, ι, Q, P) with A, ι, Q as above and P a generator of the cyclic subgroup $C_1 = e_1 Q$.

But it turns out that (when $D > 1$!), the curves $X_0^D(N)$ are much more interesting than the curves $X_1^D(N)$ insofar as their rational points are concerned. Indeed, by the above exercise $X_1^D(N)$ parameterizes (certain) points of order N on abelian varieties with potentially good reduction, and it turns out that this information alone is enough (in fact, more than enough) to deduce strong restrictions on the rational points. For instance, I showed in my thesis that for any fixed p-adic field K, there exists an integer N_0 such that $N > N_0$ implies that **for all D**, $X_1^D(N)(K) = \emptyset$. This result (and in fact much stronger results) can be found in a joint paper with Xavier Xarles.