LECTURES ON SHIMURA CURVES 8: REAL POINTS

PETE L. CLARK

1. Introduction

... do not exist, unless $B = M_2(\mathbb{Q})$, is the short version.

To be more precise, let F be a totally real field of degree g, of narrow class number 1, and B/F a quaternion algebra split at exactly one infinite place ∞_1 of F. Let $\mathcal{N}$ be an integral ideal prime to the discriminant of B. Let us write Γ, $\Gamma_0(\mathcal{N})$, $\Gamma_1(\mathcal{N})$, $\Gamma(\mathcal{N})$ for the corresponding congruence subgroups of Γ, the image in $PGL_2(\mathbb{R})$ of the positive norm units of a maximal order $\mathcal{O}$ (unique up to conjugacy, by our class number assumption) of B. We write $X = \Gamma\backslash\mathcal{H}$, $X_0(\mathcal{N}) = \Gamma_0(\mathcal{N})\backslash\mathcal{H}$, $X_1(\mathcal{N}) = \Gamma_1(\mathcal{N})\backslash\mathcal{H}$, $X(\mathcal{N}) = \Gamma(\mathcal{N})\backslash\mathcal{H}$. In fact, for brevity, let us write $X_\bullet(\mathcal{N})$ for a statement which is valid for any of the above curves.

Theorem 1. *a) There is a canonical real structure on each of the curves $X_\bullet(\mathcal{N})$.*
b) In the case when the field of definition of Shimura's canonical model is F itself (i.e., for X, $X_0(\mathcal{N})$, $X_1(\mathcal{N})$), this real structure is the same as the one obtained by the base change $\infty_1 : F \to \mathbb{R}$.
c) These curves have $\mathbb{R}$-points if and only if $B = M_2(\mathbb{Q})$.

Remark: This is a theorem of Shimura, except that I have also considered real structures on the curves $X(\mathcal{N})$.[1] We will, alas, not prove part b). Note the striking part c) – in the case $F = \mathbb{Q}$, it kills all our hopes of finding $\mathbb{Q}$-rational points on $X_\bullet^D(N)$.

On the other hand, we should not expect these curves to have real points, because of the following result.

Proposition 2. *There is no QM-abelian surface (A, ι) defined over $\mathbb{R}$.*

Proof: Suppose $A_{/\mathbb{R}}$ is an abelian surface endowed with a map $\iota : \mathcal{O} \hookrightarrow \mathrm{End}(A)$. Let $G = A(\mathbb{R})^0$, i.e., the connected component of the origin in the real locus. G is a real Lie group isomorphic to $S^1 \times S^1$, and since abelian variety endomorphisms preserve the origin, we get an action of $\mathcal{O}$ on $\mathrm{End}(G)$. This induces an action of B on $H^1(G, \mathbb{Q}) \cong \mathbb{Q}^2$, i.e., a two-dimensional $\mathbb{Q}$-representation of B. But this gives an injection $B \hookrightarrow M_2(\mathbb{Q})$, so B must be split.

Remarks: (a) The same argument works to show that there are no CM elliptic curves with all endomorphisms defined over $\mathbb{R}$ (and in other situations as well).

(b) In view of this, there cannot of course be a QM surface together with any kind of level structure defined over $\mathbb{R}$, so an excellent question to ask is why this

[1] I have no doubt that Shimura knows this case as well.

simple argument does not prove the theorem in the case $F = \mathbb{Q}$. The answer is subtle but important: since X^D is a coarse moduli space, if $K \subset \mathbb{C}$ is a nonalgebraically closed field, a K-rational point $P \in X^D(K)$ does not necessarily correspond to *any* QM structure defined over K. In other words, a QM surface need not be defined over its field of moduli.

It is known that this distressing phenomenon can really happen. (On the other hand, $X_1^D(N)$ for $N \geq 4$ is a fine moduli space, so here the argument does work to show that $X_1^D(N)(\mathbb{R}) = \emptyset$.) Indeed, a precise understanding of the obstruction of defining a QM surface over its field of moduli is afforded by the following theorem of Jordan.

Theorem 3. *(Jordan) Let $P \in X^D(L)$ be an L-valued point on the Shimura curve X^D (here L is a field of characteristic zero). Then P corresponds to a QM surface $(A, \iota)_{/L}$ if and only if $B \otimes_{\mathbb{Q}} L \cong M_2(L)$ (i.e., iff L splits B).*

Exercise 1: Show that Theorem 3 plus Proposition 2 implies Theorem 1c) in the case $F = \mathbb{Q}$.

One implication is easy:

Lemma 4. *If $(A, \iota)_{/L}$ is a QM abelian surface defined over L, then L splits B.*

Exercise: Prove it. (Hint: the argument is much the same as Proposition 2 except that we look at the action of B on $H^0(A, \Omega^1_L)$, a 2-dimensional L-vector space.)

On the other hand, the other implication in Jordan's theorem is deeper than Theorem 1c).

The following (probably rather difficult, at the moment) exercise gives the "generic" example of nondefinability over the field of moduli:

Exercise 2*) For $D > 1$, let $L = \mathbb{Q}(X^D)$, i.e., the function field of the (canonical $\mathbb{Q}$-model) of X^D. Show that L does not split B.

Problem: Find some elementary to deduce Theorem 1c) from Proposition 2, or as many instances as possible.

2. Antiholomorphic involutions

Of course $\mathbb{C}/\mathbb{R}$ is a Galois extension of fields – $\mathbb{R} = \mathbb{C}^{z \mapsto \overline{z}}$ – surely the first nontrivial example of such for all of us. In general, if L/K is a Galois extension of fields with Galois group G, we have a variety $X_{/L}$ and would like to have a model of X defined over K, there is a recipe for how and when this occurs, due to Weil. For each $\sigma \in G$, put $X^\sigma = X \times_\sigma L$. A necessary condition is that $X^\sigma \cong X$ for all σ. But this is not sufficient: we need a set of isomorphisms $f_\sigma : X^\sigma \to X$ satisfying the cocycle condition: for all $\sigma, \tau \in G$, $f_{\sigma\tau} = f_\sigma \circ \sigma(f_\tau)$. We refer to a family of such isomorphisms as **descent data** on X for L/K.

Knowledge of the above fact is a way to distinguish between arithmetic geometers and complex algebraic geometers. On the other hand, the automorphism of complex conjugation on $\mathbb{C}$ plays a distinguished role – it is a *continuous* (even smooth) automorphism of $\mathbb{C}$, so that even differential geometers need to know the notion of a "real structure" on a complex manifold V. There are several equivalent ways to phrase it:

(RS1) Let $c(V) = V \times_c \mathbb{C}$ be V endowed with its conjugate complex structure. Then a (type 1) antiholomorphic involution on V is a diffeomorphism $f : V \to c(V)$ with the property that $c(f) = f^{-1} : c(V) \to V$.

Let us say a few word about the conjugate complex structure $c(V)$. We can view a complex manifold as a topological space endowed with a sheaf of $\mathbb{C}$-algebras $\mathcal{O}_V$ which upon restriction to each subset in an open covering is isomorphic to $\mathbb{C}^n$ with its standard sheaf $\mathcal{O}$ of holomorphic functions. Applying complex conjugation to $\mathcal{O}$ on $\mathbb{C}^n$ essentially amounts to taking the sheaf $c(\mathcal{O})$ whose functions are $c(f)(z) = \overline{f(z)}$, i.e., the complex conjugates of the usual holomorphic functions.

On the other hand, using f to identify V and $c(V)$, we may view c as an involution on V, leading to:

(RS2) A (type 2) antiholomorphic involution on V is an involutory diffeomorphism f of V with the property that $f^*(\mathcal{O}_V) = c(\mathcal{O}_V)$.

(RS3) In the case when V is of complex dimension one, we can characterize the complex analytic diffeomorphisms of V as those which, for every point p on V, induces an $\mathbb{R}$-linear map T_pV to $T_{f(p)}(V)$ which is conformal, i.e., orientation-preserving and angle-preserving. Then we get a third definition of f as an anti-conformal diffeomorphism, i.e., reverses orientation but preserves unoriented angles.

We leave as an exercise for the reader to verify that (RS1) and (RS2) are equivalent to each other and to (RS3) in the one-dimensional case, and that (RS1) really is the standard descent data which determines an $\mathbb{R}$-model for V. In particular, the real points are the fixed points of the involution f of (RS2) and (RS3).

Let X be a Riemann surface (not necessarily compact), and let $\mathrm{Aut}_{\mathrm{a\,/\,c}}(X)$ be the group of anti/conformal automorphisms of X. This group contains $\mathrm{Aut}_C(X)$ with index at most 2; the index is 2 exactly when X admits some real structure. Moreover the inequivalent real structures correspond to the conjugacy classes of antiholomorphic involutions in $\mathrm{Aut}_{\mathrm{a\,/\,c}}(X)$.

Example 1: Let $V = \mathbb{CP}^1$ be the complex projective line. There are two inequivalent anticonformal involutions: one of them is induced by complex conjugation on the set of lines through the origin in $\mathbb{C}^2$ and so has fixed point set $\mathbb{RP}^1$. The other is given by the antipodal map, which is fixed point free: the real structure is represented by the unique conic over $\mathbb{R}$ without real points.

Exercise 3: Prove it.

Example 2: Let $V = \mathbb{C}$ be the complex plane (or, the complex affine line). It is (very) well known that the anticonformal involutions are precisely the reflections through lines in the plane, and that any two are conjugate inside the isometry group of $\mathbb{C}$, hence certainly inside the group of anti/conformal automorphisms of $\mathbb{C}$. (Recall that the former is a real Lie group of dimension 3 and the latter is a real Lie group of dimension 4.) Thus we may as well take the anticonformal involution to be complex conjugation itself. The fixed point set is, of course, the affine line over $\mathbb{R}$.

Example 3: Let $\Lambda \subset \mathbb{C}$ be a lattice. Suppose that Λ is stabilized by complex conjugation. (Concretely, if $\Lambda = \mathbb{Z}1 \oplus \mathbb{Z}\tau$, then we are requiring $\overline{\tau} = a + b\tau$ for some $a, b \in \mathbb{Z}$.) Then $c : \mathbb{C} \to \mathbb{C}$ descends to an antiholomorphic involution on $\mathbb{C}/\Lambda$. This gives the complex elliptic curve $E = \mathbb{C}/\Lambda$ the structure of a real curve of genus one; call it X_Λ.

Exercise 4: a) Let $\mathfrak{o}$ be an imaginary quadratic order. We have seen that $\mathfrak{o}$ may naturally be viewed as a lattice in $\mathfrak{o} \otimes \mathbb{R} = \mathbb{C}$. Note that $\mathfrak{o}$ is evidently stable under complex conjugation. What fact about CM elliptic curves does this prove?
b) For any c-stable lattice Λ, describe the set $X_\Lambda(\mathbb{R})$ of fixed points of complex conjugation. (Much easier:) Show in particular that it is always nonempty, so that X_Λ is an elliptic curve over $\mathbb{R}$.
c) Can every elliptic curve with real j-invariant be uniformized by a c-stable lattice?
d) Suppose Λ is a c-stable lattice, so $\mathbb{C}/\Lambda$ has an $\mathbb{R}$-structure. Is this $\mathbb{R}$-structure unique?

Example 3: Let $V = \mathcal{H}$ be the upper halfplane. Then $\mathrm{Aut}_{\mathrm{a\,/\,c}}(\mathcal{H}) = PGL_2(\mathbb{R})$, where a matrix $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ of negative determinant acts on $\mathcal{H}$ by $g(z) = c(\frac{az+b}{cz+d})$. Note that the subgroup of matrices whose determinant (class) is a square has index 2 and is isomorphic to $PSL_2(\mathbb{R})$. In other words, $\mathcal{H}$ admits real structures. For instance, we could take our anticonformal involution to be given by $f(z) = -\overline{z}$ corresponding to the diagonal matrix with entries -1 and 1. The fixed point set is the line $i\mathbb{R}^{>0}$. In fact all involutions in $PGL_2(\mathbb{R}) \setminus PGL_2(\mathbb{R})^+$ are conjugate, and correspond to reflections through geodesics in the upper halfplane. Thus this real structure is unique.

Example 4: Let $V = Y(1) = PSL_2(\mathbb{Z})\backslash\mathcal{H}$. One notices immediately that $i\mathbb{R}^{>0}$ is a symmetry line for the standard fundamental domain for $PSL_2(\mathbb{Z})$, which suggests that $PSL_2(\mathbb{Z})$ has a real structure coming from the matrix $g = \mathrm{diag}(1, -1)$. Note that indeed the normalizer of $PSL_2(\mathbb{Z})$ in $PGL_2(\mathbb{R})$ is $PGL_2(\mathbb{Z})$, which contains the image of g. In the case at hand, this is not terribly exciting (as a Riemann surface, V is the affine line, which we already saw had a unique real structure), but suggests a vast generalization.

Let $\Gamma \subset PSL_2(\mathbb{R})$ be a Fuchsian group of the first kind, and let $N_{\mathrm{a\,/\,c}}(\Gamma)$ be the normalizer of Γ in $PGL_2(\mathbb{R})$. It follows easily from our work on Fuchsian groups that $N_{\mathrm{a\,/\,c}}(\Gamma)/\Gamma$ is a finite group consisting of anti/conformal automorphisms of Γ.

Exercise 5: Suppose that Γ is of hyperbolic type (no cusps or elliptic points). Show that $N_{\mathrm{a\,/\,c}}(\Gamma)/\Gamma$ is the full group of anti/conformal automorphisms of $V = \Gamma\backslash\mathcal{H}$.

In particular, V admits a real structure if and only if $N_{ac}(\Gamma)$ contains an element of $PGL_2(\mathbb{R})^-$.

Example 5: Put $g = \operatorname{diag}(1, -1)$. Then, for any matrix $m = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{R})$, $gmg^{-1} = \begin{bmatrix} a & -b \\ -c & d \end{bmatrix}$. Thus g normalizes $\Gamma_0(N)$, $\Gamma_1(N)$, $\Gamma(N)$ for all N and gives an antiholomorphic involution g on $Y_0(N)$, $Y_1(N)$, $Y(N)$.

Exercise 6: a) Show that g extends to an antiholomorphic involution on the compactifications $X_0(N)$, $X_1(N)$, $X(N)$.
b) Analyze the action of g on the cusps for these curves. (Again, much easier:) Show in particular that there is always an $\mathbb{R}$-rational cusp. Why does it follow that the loci $Y_0(N)$, $Y_1(N)(\mathbb{R})$, $Y(N)(\mathbb{R})$ $Y(N)(\mathbb{R})$ are nonempty?[2]
c)* Suppose $N = p_1 \cdots p_r$ is squarefree. Show that there are at least 2^r inequivalent real forms of $X_0(N)$, and that all of them have $\mathbb{R}$-points.
d) Prove or disprove: Fix a prime $p \geq 7$. The anti/conformal automorphism group of $X(p)$ is $G = PGL_2(\mathbb{F}_p)$, and any two anticonformal involutions are conjugate in G (i.e., the real structure is unique).

Remark: We have caught a glimpse of a rather rich theory of real structures on Riemann surfaces. Especially, one can describe the complex- and real-analytic unformization theories in such a way so as to make a parallel with the theory of *p-adic* uniformization of certain algebraic curves (like Shimura curves!). We will not, alas, have time for this here; we have already waded in deeper than is strictly necessary to prove Shimura's theorem.

3. The proof of Theorem 1

Now let B/F be a quaternion algebra of type $(1, g-1)$ over the totally real field F, let $\mathcal{O}$ be a maximal order in F, and let $\Gamma = \Gamma(B, \mathcal{O}) \subset PGL_2(\mathbb{R})^+$ be, as usual, the group of units of $\mathcal{O}$ of positive norm, so that $X(\mathcal{O}) = \Gamma\backslash\mathcal{H}$ is a Shimura curve. From the preceding discussion, it should be clear what to do to get a real structure: let $G = \mathcal{O}^\times / \pm 1$. Clearly G is a subgroup of $PGL_2(\mathbb{R})$, i.e., of anti/conformal automorphisms of $\mathcal{H}$. If we assume that F has narrow class number one, then $\mathcal{O}$ has units of reduced norm -1.[3] Thus there exists an element $g \in G$, of order two and negative determinant class, and it is not hard to check that any two such elements are conjugate. Thus we have endowed $X(\mathcal{O})$ with a real structure.

Exercise 7: For any integral ideal $\mathcal{N}$ coprime to the discriminant of B, show that the matrix g normalizes every congruence subgroup $\Gamma_0(\mathcal{N})$, $\Gamma_1(\mathcal{N}), \Gamma(\mathcal{N})$ of $\Gamma(B, \mathcal{O})$. So all of these Shimura curves have canonical $\mathbb{R}$-models. (Hint: work locally.)

Finally, we must see that a Shimura curve only has real points if $B = M_2(\mathbb{Q})$.

[2]Note that the Shimura-Delignecanonical model for $Y(N)$ is defined over $\mathbb{Q}(\zeta_N)$, which, for $N > 2$, does not admit a real embedding.

[3]In general, we know there are integral elements of B whose norm is an ∞_1-negative unit, which must lie in *some* maximal order $\mathcal{O}$. Our assumption implies that all maximal orders are conjugate. It is not clear to me that this assumption is actually required.

This is now just a simple calculation: let $\Gamma \subset \Gamma(B, \mathcal{O})$ be any finite index subgroup normalized by a matrix $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathcal{O}^\times$ with $\det(g) = -1$, $g^2 = \mathrm{diag}(\alpha, \alpha)$ (a scalar matrix). Then the $\mathbb{R}$-points on $\Gamma \backslash \mathcal{H}$ correspond to those points $z \in \mathcal{H}$ such that $g(\overline{z}) = \overline{g(z)} = \gamma z$ for some $\gamma \in \Gamma$. Put $h := \gamma^{-1} g \in \mathcal{O}^\times$. Suppose $h = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then we have the equation $h(\overline{z}) = z$, or

$$c|z|^2 + b = a\overline{z} - dz.$$

Since the left handside is real, so is the right hand side, which implies that $a = -d$, so that b has reduced trace zero. On the other hand it has determinant -1, so its reduced characteristic polynomial is $T^2 - 1$. But this is problematic: since B is a division algebra, the minimal polynomial of any element is irreducible, so we conclude that $b = \pm 1$, in which case its reduced norm is 1, not -1. This contradiction completes the proof.