RATIONAL POINTS ON ATKIN-LEHNER TWISTS OF MODULAR CURVES

PETE L. CLARK

These are the (more detailed) notes accompanying a talk that I am to give at the University of Pennsylvania on July 21, 2006. The topic is rational points on Atkin-Lehner twists of the modular curves $X_0(N)$. Apart from being an interesting Diophantine problem in its own right, there is an ulterior motive: $\mathbb{Q}$-rational points correspond to "elliptic $\mathbb{Q}$-curves" and thus to projective Galois representations. We will see that this leads to a realization of infinitely many new groups $PSL_2(\mathbb{F}_p)$ as Galois groups *conditional* on the Birch Swinnerton-Dyer conjecture, and to a "natural" infinite sequence of curves violating the Hasse principle.

These last two results, which are taken from a 2005 note [Cl1] and a very recent preprint [Cl2] of mine, might sound deep and/or impressive, but the proofs are easy to the point of raising the question of why they were not done before. In response, I would have to say that this circle of objects and ideas – so close to the number-theoretic mainstream (what I was taught in grad school were the three mainstays of number theory – Diophantine geometry, Galois theory, and automorphic forms – are all clearly present and up to their usual tricks) – seems profoundly underexplored. I think there are many interesting and *tractable* problems here, and I will try to justify this impression (to the extent that I am able; I have almost nothing intelligent to say about the automorphic side of things) as much as to showcase any result of mine.

1. Quadratic twists

Let $C_{/\mathbb{Q}}$ be an algebraic curve endowed with a $\mathbb{Q}$-rational involution ι. Then, for each quadratic field extension $\mathbb{Q}(\sqrt{d})/\mathbb{Q}$, we may define a new curve called $\mathcal{T}(C, \iota, \mathbb{Q}(\sqrt{d})/\mathbb{Q})$ – or, when we can get away with it, C_d. It is a curve which becomes isomorphic to C over $\mathbb{Q}(\sqrt{d})$ but has a twisted Galois action on its $\mathbb{Q}(\sqrt{d})$-rational points: if σ_d is the generator of $\mathrm{Gal}(\mathbb{Q}(\sqrt{d})/\mathbb{Q})$, then for $P \in C_d(\mathbb{Q}(\sqrt{d}))$,

$$\sigma_d(P) := \iota(\sigma_d(P)),$$

where the $:=$ is to be interpreted as in computer science: the *new* σ_d is the *old* σ_d followed by ι. (This a special case of the principle of Galois descent: $H^1(\mathbb{Q}, \langle\iota\rangle) = \mathrm{Hom}(\mathbb{Q}, \mathbb{Z}/2\mathbb{Z}) \cong \mathbb{Q}^\times/\mathbb{Q}^{\times 2}$.) In plainer terms, a $\mathbb{Q}(\sqrt{d})$-rational point P is $\mathbb{Q}$-rational on C_d iff $\sigma_d(P) = \iota(P)$.

Let us make the convention that d ranges through squarefree (positive or negative) integers, and that $C_1 = C$ itself. Then there is a close relationship between the $\mathbb{Q}$-points on C/ι, the $\mathbb{Q}$-points on the various curves C_d, and *certain* quadratic points on C. On the one hand, since $C_d/\iota = C/\iota$, every $\mathbb{Q}$-point on C_d maps to a $\mathbb{Q}$-point on C/ι. Moreover, the preimage of every *nonbranch* point $P \in (C/\iota)(\mathbb{Q})$ is a set of two points $\{Q_1,\ Q_2\}$ which is stable both under ι and under the action

of Galois, so $Q_1 = \sigma_d(Q_2)$ for a unique squarefree d, and these two points are $\mathbb{Q}$-rational on C_d (and not on $C_{d'}$ for any other squarefree d'). On the other hand a $\mathbb{Q}$-rational branch point has, set-theoretically, a unique preimage $Q \in C(\mathbb{Q})$ which satisfies, for all d, $\sigma_d(Q) = Q$ and $\iota(Q) = Q$, so rational ι-fixed points stay rational on *all* the twists C_d.

Theorem 1. *Let $(C, \iota)_{/\mathbb{Q}}$ be a curve endowed with an involution ι. Suppose:*
(i) $\{P \in C(\mathbb{Q}) \mid \iota(P) = P\} = \emptyset$.
(ii) There exists $P_0 \in C(\overline{\mathbb{Q}})$ such that $\iota(P_0) = P_0$.
(iii) $(C/\iota)(\mathbb{Q})$ is finite.
(iv) For all $\ell \leq \infty$, $C(\mathbb{Q}_\ell) \neq \emptyset$.
Then the primes $p \equiv 1 \pmod 4$ such that the twisted curve $C_p = \mathcal{T}(C, \iota, \mathbb{Q}(\sqrt{p})/\mathbb{Q})$ violates the Hasse principle over $\mathbb{Q}$ have positive density.

We will content ourselves with the following remarks: combining (i) and (iii) we get that $C_d(\mathbb{Q}) = \emptyset$ for all but finitely many squarefree d (prime or otherwise). The hypothesis (ii) ensures that $C_p(\mathbb{Q}_p) \neq \emptyset$ for any prime p splitting completely in the field of definition of P_0. By an argument using (in particular) the Weil bounds for curves over finite fields, one finds that the other places can be handled by requiring p to be a quadratic residue modulo sufficiently many small primes ℓ, and we conclude by applying the Cebotarev density theorem.

Once we assume that C has points everywhere locally (e.g. if it has an obvious $\mathbb{Q}$-point!) and admits an involution ι, the other hypotheses look rather mild: (i) holds e.g. if $g(C)$ is even (for "topological" reasons), (iii) holds whenever $g(C/\iota) \geq 2$ by Faltings' finiteness theorem, and surely (ii) holds "generically" whenever there are at least two branch points.

Remark: Note that (i) and (ii) are *necessary* for the conclusion of the theorem: as we saw, if (i) did not hold, we get $\mathbb{Q}$-points on all quadratic twists (an observation which we will be able to exploit later!). If (ii) did not hold, then it follows from a theorem of Chevalley-Weil that only finitely many of the twists can have points even everywhere locally.[1] It might be interesting to try to weaken (iii) or (iv).

Remark: To answer (again) a question I was asked at the talk, the density can be bounded below in terms of the genera of C and C/ι.

We will want to apply this theorem to Atkin-Lehner twists of modular curves, so let us now talk about them.

2. A semistable model for $X_0(N)$

Let us choose a squarefree[2] integer $N = p_1 \cdots p_r > 1$ and consider the modular curve $X_0(N)$. As a Riemann surface we recognize it as the upper half plane $\mathcal{H}$ modulo the subgroup $\Gamma_0(N)$ of $SL_2(\mathbb{Z})$ consisting of matrices which are upper triangular modulo N – or more precisely, the compactification thereof obtained by

[1] This fact – which holds for all geometrically Galois unramified coverings – is what makes possible the *descent* method for studying rational points on curves [Poo].

[2] In the talk itself I took N to be prime, for simplicity: almost nothing is lost.

adding 2^r cusps. The noncuspidal points[3] of this Riemann surface parameterize elliptic curves over the complex numbers together with a distinguished order N cyclic subgroup C.[4] Or, what is a clearly equivalent but sometimes useful alternate viewpoint, it parameterizes triples (E, E', f) where $f : E \to E'$ is a degree N isogeny (indeed all such f arise as $E \to E/C$ for C as above, and conversely). We also have a natural involution W_N, which carries f to its dual isogeny $f' : E' \to E$ (uniquely characterized by $f' \circ f = [N]$).

Although $X_0(N)$ is already interesting as a Riemann surface,[5] it is very much more interesting as a curve defined over a smaller base, like $\mathbb{Q}$. But let's go whole hog and introduce the *integral canonical model*, i.e., define a curve over $\mathbb{Z}$ (a.k.a. an "arithmetic surface") whose basechange to $\mathbb{C}$ is $Y_0(N)$. Indeed we can just cling to the moduli problem: we want the coarse moduli scheme attached to the functor which takes a scheme S to the isomorphism classes of pairs (E, C), where E/S is an elliptic curve and C is an "order N subgroup" of E. As long as N is invertible on S this works *exactly* as in the complex case, and we get a smooth curve $X_0(N)_{/\mathbb{Z}[1/N]}$.[6] When we work over a $\mathbb{Z}_p$-scheme with $p \mid N$ things become more interesting: we must speak explicitly in terms of subgroup schemes. The generic elliptic curve E in characteristic p has $E[p] \cong \mathbb{Z}/p\mathbb{Z} \times \mu_p$, and the moduli problem here instructs us to choose – along with a necessarily étale order N/p subgroup – an order p subgroup scheme of $E[p]$, and instead of $\#\mathbb{P}^1(\mathbb{F}_p) = p + 1$ choices, there are visibly just 2. Thus the fiber of $X_0(N) \to X_0(N/p)$ over each ordinary point consists of two (reduced!) points. On the other hand there are always finitely many nonsingular points, for which $E[p]$ is a nontrivial extension of $\alpha_p = \mathbb{G}_a[p]$ by itself, and here there is only *one* order p subgroup scheme.

We have given what is somewhere between an explanation and a proof of the following statement: $X_0(N)_{/\mathbb{F}_p}$ is obtained from two copies of the smooth curve $X_0(N/p)_{/\mathbb{F}_p}$ by glueing each supersingular point on the first copy to a supersingular point on the second copy. There is a slight twist: the isogeny $E \mapsto E/\alpha_p$ is the Frobenius map, so a point $(E, C_p, C_{N/p})$ gets glued not to the identical elliptic curve on the other copy, but to its twist by $\mathbb{F}_{p^2}/\mathbb{F}_p$ Frobenius. So the special fiber looks like "double helix" in which – as with real DNA! – the two strands are somehow glued together with orientations reversed. Note well that the Atkin-Lehner involution w_N interchanges the two components.

In particular, $X_0(N)_{/\mathbb{Z}}$ is a *semistable* model for $X_0(N)$ (i.e., the only singularities of the geometric fibers are ordinary double points). A key point is that this "modular" $\mathbb{Z}$-model of $X_0(N)$ is not necessarily *regular*; however, because it is a *semistable* model, it is easy to construct from it a regular model, and in fact the minimal regular model (assuming the genus is positive).[7] For this we need to know

[3]In fact, the cusps can be viewed as parameterizing semistable singular curves of arithmetic genus one, an observation of Néron that is useful in many contexts, but not for us today.

[4]Since N is squarefree the cyclicity is guaranteed.

[5]E.g., which complex elliptic curves does it cover?

[6]The facts that the coarse moduli space exists, and that it is smooth, are of course not obvious, but rather follow from work of Igusa.

[7]Calling this curve $X_0(N)_{/\mathbb{Z}}$ is nonstandard (and bad): Mazur and Rapoport call it $M_0(N)_{/\mathbb{Z}}$ ("M" standing for "moduli", presumably) and reserve $X_0(N)_{/\mathbb{Z}}$ for the regularization of this model. This is more sensible because given any positive genus curve C over $\mathbb{Q}$, there is a unique

what the strict complete local ring at each of the supersingular points – when viewed as a closed point in the ambient arithmetic surface – looks like. It will be isomorphic to $W(\overline{\mathbb{F}_p})[[x,y]]/(xy-p^a)$ for some positive integer a, and to get a regular model we must blow up $a-1$ times. There is a recipe for what a is, but let me come back to it later when it shall seem much more exciting.

This was interesting enough, but why do I bring it up? Certainly it is key knowledge for much of the deep work on modular curves: for instance, Ken Ribet used this model – together with a description of the regular model for Shimura curves[8] $X_0^D(N)$ which is partly analogous but requires an extra tool (Cerednik-Drinfeld uniformization) – to show that modularity of elliptic curves implies Fermat's last theorem. It also comes up prominently in the following result:

Theorem 2. *(mostly[9] Mazur) For $N > 163$, $X_0(N)(\mathbb{Q})$ consists only of cusps.*

When studying the Diophantine geometry of any curve $C_{/\mathbb{Q}}$, it is nice to have an explicit regular $\mathbb{Z}$-model, as one can use it to determine whether or not the curve has $\mathbb{Q}_p$-rational points. Indeed, (one of many versions of) Hensel's Lemma says that $C(\mathbb{Q}_p) \neq \emptyset$ iff there exists a *nonsingular* $\mathbb{F}_p$-rational point on the special fiber. This remark may seem strange for two reasons: first, many of us are taught that it is *easy* to determine whether or not a curve – or any variety – defined over $\mathbb{Q}$ has points over any given $\mathbb{Q}_p$. This is true in a certain sense – given a particular set of defining equations, there is in principle and sometimes even in practice an algorithm to determine the complete set of places $p \leq \infty$ for which $C(\mathbb{Q}_p) \leq \infty$. But aside from the fact that most varieties are not *a priori* given to us by an explicit set of equations, it is clear that if we are interested in analyzing local points on an *infinite family* of cognate varieties, then unless we have an infinite amount of time the algorithmic approach is not the way to go.

But secondly: of course we *do* have $\mathbb{Q}$-rational points on $X_0(N)$ for all squarefree N: the cusps are all $\mathbb{Q}$-rational.[10] So of course local methods are useless for determining the $\mathbb{Q}$-rational points on $X_0(N)$. (But they will be relevant for the curves $C(N,p)$ to be defined shortly.) This is in fact what makes the problem, along with Fermat's Last Theorem, so difficult: we have at least some powerful tools – e.g., local methods and the Brauer-Manin obstruction[11] – to show that a curve has no rational points at all, but there is no general method for showing statements like "The only rational points on C are..."

---

minimal regular model, or in other words a canonical nicest way to extend C to a curve over $\mathbb{Z}$. We should reserve the notation $C_{/\mathbb{Z}}$ for this model. Why I chose the bad notation I now forget.

[8]Some of the results I will describe have Shimura curve analogues; for others it is an interesting problem whether such analogues exist. Indeed, as Prof. Chai commented at the end of the talk, one would like to work in the generality of Shimura curves over totally real fields.

[9]Mazur's paper treats the case of prime N only; to get from this case to squarefree N is not difficult. But in fact, [RI] beings with a list of all known exceptional values of N (the largest being $N = 163$), squarefree or otherwise, and this list was eventually shown to be complete by Kenku and others.

[10]When N is not squarefree all of the cusps need not be $\mathbb{Q}$-rational, but at least one of them always is.

[11]It has been conjectured that the Brauer-Manin obstruction is the only one to the existence of rational points on curves. Since hidden in this assertion is the claim that Shafarevich-Tate groups are finite, this is not likely to be proved anytime soon, but it has strongly influenced recent work in the subject.

However, there is one favorable case for the enumeration of $\mathbb{Q}$-points on a curve C (of positive genus; the other case is easy): namely, if its Jacobian has a $\mathbb{Q}$-factor A of rank zero, then – assuming we have at least one point! – there is a finite-to-one map from $C(\mathbb{Q})$ to $A(\mathbb{Q})$, so we can just compute the preimages and see how many of them are $\mathbb{Q}$-rational. This *always* happens for $X_0(N)$ – here A is the quotient by the famed *Eisenstein ideal* of $\operatorname{End} J_0(N)$ – so as soon as Mazur was able to show this (quite deep) "fact" he knew immediately that all positive genus $X_0(N)$ had finitely many $\mathbb{Q}$-rational points. (Note that his work preceded Faltings' finiteness theorem.) Again this is quite different from finding the rational points on *all* the curves $X_0(N)$ at once; for the latter, many more wonderful ideas are needed.[12]

3. Atkin-Lehner quotients and Atkin-Lehner twists

For a much sterner challenge, consider the curve

$$X_0^+(N) = X_0(N)/w_N.$$

The Atkin-Lehner theory of signs of functional equations – together with the conjecture of Birch and Swinnerton-Dyer – implies that the Jacobian $J_0^+(N)$ of $X_0(N)$ has the property that every $\mathbb{Q}$-factor has rank exceeding its dimension.[13] For example, $X_0^+(37)$ is the first elliptic curve of positive rank. Thus none of the standard methods apply for finding the rational points on $X_0^+(N)$, and the only reason that we know that $X_0^+(N)(\mathbb{Q})$ is finite when the genus is at least 2 – that is, for $N > 131$ – is because of Faltings' finiteness theorem.

In fact, unlike $X_0(N)$, it is the case that for infinitely many N there are non-cuspidal rational points on $X_0^+(N)$ – namely, under certain (well understood) congruence conditions there will exist a rational CM point. Let us call a noncuspidal, non-CM point *exceptional*. Then it is a "folk conjecture" that for sufficiently large N, there are no exceptional $\mathbb{Q}$-rational points. As far as I know, the only reason to believe this is what I might call *Horatio's philosophy* on rational points on curves: except for the ones which we can see with our own eyes, $\mathbb{Q}$-rational points are few and far between.

Why do we care about rational points on $X_0^+(N)$? Well, it too has a natural and interesting moduli interpretation, closely related to that of $X_0(N)$: $\mathbb{Q}$-points on $X_0(N)$ correspond to dual pairs of N-isogenies $\iota : E \to E'$, $\iota' : E' \to E$ which are as a pair defined over $\mathbb{Q}$: in other words, either (E, ι) is itself defined over $\mathbb{Q}$ – the trivial case, which by Mazur's theorem we can rule out when $N > 163$ – or (E, ι) is defined over some quadratic field $\mathbb{Q}(\sqrt{d})$ and the nontrivial automorphism σ of this quadratic field sends (E, ι) to its dual. Such a thing is called a *quadratic* $\mathbb{Q}$-curve of degree N.[14]

[12] In the paper [EI], Barry had "only" enough wonderful ideas to handle the easier case of $X_1(N)$; and only a year or two later did he see to solve the $X_0(N)$ case [RI].

[13] In particular, one of the "relations" in the Eisenstein ideal is $w_N = -1$, so $J_0^+(N)$ is *a priori* orthogonal to the Eisenstein quotient.

[14] In fact, a result of Elkies says that all non-CM elliptic curves which are isogenous to all of their Galois conjugates come from a rational point on the full Atkin-Lehner quotient $X_0(N)/W$, i.e., are defined over a multiquadratic extension.

Such curves are interesting for many reasons that we shall not have time to explain: see [Ell] for more motivation. We shall concentrate on the following property: for p prime to N, a quadratic $\mathbb{Q}$-curve gives rise to a *projective* Galois representation, i.e., to a homomorphism $\rho : \mathrm{Gal}_{\mathbb{Q}} \to PGL_2(\mathbb{F}_p)$.

Here is a brief explanation of how this works: first, if E were defined over $\mathbb{Q}$, by letting $\mathrm{Gal}_{\mathbb{Q}}$ act on $E[p]$ we would get a homomorphism $\mathrm{Gal}_{\mathbb{Q}} \to \mathrm{Aut}(E[p]) = GL_2(\mathbb{F}_p)$. Since the data of $\mathbb{Q}$-curve is two elliptic curves defined over $K = \mathbb{Q}(\sqrt{d})$, each one of them gives a representation $\mathrm{Gal}_K \to GL_2(\mathbb{F}_p)$. Now consider the fact that the cyclic N-isogeny between them induces an isomorphism on the p-torsion subgroups (indeed on the entire p-adic Tate module, since $(p, N) = 1$), and moreover – since E, E' are not CM, they do not have any endomorphisms of order N – this isomorphism is unique up to a scalar matrix. So the generator σ of $\mathrm{Gal}_{K/\mathbb{Q}}$ induces an $\mathbb{F}_p$-linear map from $E[p]$ to $E'[p]$, and by composing with ι' we get a canonical element of $\mathrm{Aut}(\mathbb{P}E[p]) = PGL_2(\mathbb{F}_p)$.

This has applications to the inverse Galois problem, although this is not immediately clear: e.g., it is well-known that $PGL_2(\mathbb{F}_p)$ occurs as a Galois group over $\mathbb{Q}$ for all p. Indeed $GL_2(\mathbb{F}_p)$ does, a stronger statement. In fact there exists a single elliptic curve $C_{/\mathbb{Q}}$ for which the representation on p-torsion is surjective for all odd primes p (a result of Serre) – and getting $GL_2(\mathbb{F}_2) \cong S_3$ as a Galois group, via elliptic curves or otherwise, is easy. On the other hand, since $PSL_2(\mathbb{F}_p)$ is not a quotient but a subgroup of $PGL_2(\mathbb{F}_p)$, the above arguments do not suffice. It may seem at first that this should also be easy to do with elliptic curves: the covering $X(p) \to X(1)$ has Galois group $PSL_2(\mathbb{F}_p)$ – however it is only geometrically Galois. A little analysis shows that one can realize it minimally as a Galois cover over $\mathbb{Q}(\sqrt{p^*})$ - i.e., we are a quadratic extension away.

Here is where $\mathbb{Q}$-curves can come in to help: if we have a $\mathbb{Q}$-curve defined over $\mathbb{Q}(\sqrt{p^*})$, and if $(\frac{N}{p}) = -1$, then the associated homomorphism in fact lands in $PSL_2(\mathbb{F}_p)$: in some sense we have "twisted away" the undesired quadratic extension. A more geometric way of saying this is as follows:

Theorem 3. *(Shih) Let N be squarefree, and p a prime with $(\frac{N}{p}) = -1$. Let $C(N, p)$ be the quadratic twist of $X_0(N)$ by w_N and $\mathbb{Q}(\sqrt{p^*})/\mathbb{Q}$. Then there exists a regular $PSL_2(\mathbb{F}_p)$-Galois covering $\psi : Y \to C(N, p)$ defined over $\mathbb{Q}$.*

Now we have an obvious strategy: find a $\mathbb{Q}$-rational point P on $C(N, p)$ such that the fiber of φ over P remains irreducible over $\mathbb{Q}$; this is equivalent to the Galois representation on the corresponding $\mathbb{Q}$-curve being surjective. Note that CM points will (apart from a few trivial cases) never work, since their Galois representations will have solvable image. Intuitively, we feel that "most" non-CM points should have surjective image, so if there are "many" $\mathbb{Q}$-rational points on $C(N, p)$, our strategy is a good one.

There is a well-known classical case: if $C(N, p) \cong \mathbb{P}^1$, then a theorem of Hilbert assures that a density one set of specializations will remain irreducible (more precisely, this property holds on the complement of a thin set): in this case we say that $PSL_2(\mathbb{F}_p)$ *occurs regularly* over $\mathbb{Q}$. Remarkably, Serre observed that it follows

from Faltings' theorem that there are also infinitely many irreducible specializations when $C(N,p)$ is an elliptic curve of positive rank[15] (so, whenever $C(N,p)(\mathbb{Q})$ is infinite). There is no similar general criterion for the case of finitely many rational points (in particular, when $X_0(N)$ has genus at least 2 – i.e., for $N > 21$).[16]

So, among the squarefree $N > 1$ for which $X_0(N)$ has genus 0 – namely

$$N = 2,\ 3,\ 5,\ 7,\ 10,\ 13;$$

or genus one – namely

$$N = 11,\ 14,\ 15,\ 17,\ 19,\ 21;$$

our task is to determine for which primes p such that $(\frac{N}{p}) = -1$, $C(N,p)(\mathbb{Q})$ is infinite.

Theorem 4.
a) (Shih) If $N \in \{2,\ 3,\ 7\}$, then $C(N,p) \cong \mathbb{P}^1$ for all p.
b) (many people) $C(5,p)(\mathbb{Q}_5) = C(10,p)(\mathbb{Q}_5) = C(13,p)(\mathbb{Q}_{13}) = \emptyset$.
c) (C—) Let N be 11 or 19. Then $C(N,p)$ is an elliptic curve, which has odd analytic rank iff $p \equiv 1 \pmod 4$.
d) (Gonzalez, C—) $C(15,p)(\mathbb{Q}_5) = C(17,p)(\mathbb{Q}_{17}) = \emptyset$.

To sum up: $PSL_2(\mathbb{F}_p)$ certainly occurs as a Galois group over $\mathbb{Q}$ if at least one of 2, 3, 7 is a quadratic *non*residue mod p. Recall that conjecture of Birch and Swinnerton-Dyer asserts, among other things, that the analytic rank is equal to the Mordell-Weil rank; that at least the two ranks are equal mod 2 is called the *parity conjecture*, and thanks to recent work of Nekovar, if only we knew that Shafarevich-Tate groups were finite this would follow. Anyway, assuming some portion of BSD, the theorem says that we also get $PSL_2(\mathbb{F}_p)$ as a Galois group if $p \equiv 1 \pmod 4$ and either 11 or 19 is a quadratic nonresidue mod p. If $p \equiv -1 \pmod 4$ then the analytic rank is *even*, but it can still be positive. A famous conjecture (much less universally believed than BSD; I am not quite ready to endorse it myself) says that most elliptic curves should have rank 0 or 1, so it may well be that the *density* of the set of primes $p \equiv 1 \pmod 4$ such that $C(11,p)(\mathbb{Q})$ or $C(19,p)(\mathbb{Q})$ is infinite is equal to zero. However, there are certainly such primes, e.g. $p = 47$ (Elkies). Based on some computer calculations, I am willing to conjecture that there are infinitely many.

The other parts of the theorem give values of N for which the method fails. Note that the two values of N not addressed in the theorem – namely 14 and 21 – would not help to realize any new values of p, since $(\frac{\ell_1\ell_2}{p}) = -1$ implies $(\frac{\ell_1}{p}) = -1$ or $(\frac{\ell_2}{p}) = -1$.[17]

It is a remarkable fact that we have just described *all* known instances of $PSL_2(\mathbb{F}_p)$ as a Galois group over $\mathbb{Q}$, except one: in 1991, G. Malle proved that if $(\frac{7}{p}) = 1$,

[15]This is not literally an extension of Hilbert's irreducibility theorem to all regular Galois extensions of the function field of a positive rank elliptic curve; there are some extra hypotheses to check.

[16]I am not aware of a single example of a surjective homomorphism $\mathrm{Gal}_{\mathbb{Q}} \to PSL_2(\mathbb{F}_p)$ coming from a degree N $\mathbb{Q}$-curve when $g(X_0(N)) \geq 2$. Exceptional rational points on $X_0^+(N)$ are rare enough, and it is relatively unlikely that their field of definition will be $\mathbb{Q}(\sqrt{-p})$ for a prime p with $(\frac{N}{p}) = -1$.

[17]Neither does removing the requirement that N be squarefree.

$(\frac{5}{p}) = -1$, then $PSL_2(\mathbb{F}_p)$ is a (regular) Galois group over $\mathbb{Q}$, using purely group-theoretical methods. Or to put matters in chronological order: Shih showed that $PSL_2(\mathbb{F}_p)$ occurs over $\mathbb{Q}$ for $\frac{7}{8}$ of all primes; Malle showed it for $\frac{1}{4}$ of all primes, and together their results give $\frac{15}{16}$ of all primes. My results give, conditionally on BSD, $\frac{3}{8}$ of all primes, leaving $\frac{5}{128}$ of the primes unaccounted for. Before writing up my results, I did a couple of hours of computer calculation and found 614 primes for which $PSL_2(\mathbb{F}_p)$ certainly occurs over $\mathbb{Q}$ by my results but not by Shih's or Malle's, including two primes $p \equiv 1 \pmod 4$ for which the curves have analytic rank 2.

Probably you noticed among the prime values of N, we are claiming infinitely many rational points iff $N \not\equiv 1 \pmod 4$ iff $\mathbb{Q}(\sqrt{-N})$ has class number 1. (The second observation implies the first, since it is well known that if $N \equiv 1 \pmod 4$, the class number of $\mathbb{Q}(\sqrt{-N})$ is even.)

Recall from §1 that the points P which are simultaneously $\mathbb{Q}$-rational on $C(N,p)$ and on $X_0(N)$ are those satisfying the equations

$$\sigma(P) = P,\ \iota(\sigma(P)) = P,$$

or equivalently:

$$\iota(P) = \sigma(P) = P.$$

In other words, $\mathbb{Q}$-rational ι-fixed points will stay rational on all twists. Now w_N on $X_0(N)$ (for squarefree $N > 3$; $N = 2$ and $N = 3$ work a little bit differently) *always* has fixed points: namely, there is one complete Galois orbit of points corresponding to the ideal classes in the order $\mathbb{Z}[\sqrt{-N}]$, and there is a complete Galois orbit of points corresponding to the ideal classes in the maximal order of $\mathbb{Q}(\sqrt{-N})$. If N is $1 \pmod 4$, we have said the same thing twice, and there is indeed just one orbit. If N is $-1 \pmod 4$ these are distinct orders and we get two different Galois orbits. Either way, we get a Galois orbit consisting of a single element iff $\mathbb{Q}(\sqrt{-N})$ has class number 1. This already proves the result in the genus 0 case, since a conic with a rational point is $\mathbb{P}^1$. In the genus 1 case, we have a $\mathbb{Q}$-rational point, so we can take it as the origin and give $C(N,p)$ (for all p) the structure of an elliptic curve. The involution w_N has three other fixed points (you can check this either geometrically by Riemann-Hurwitz or algebraically by Dedekind's formula for the class number of a nonmaximal order), so modding out by w_N gives a degree 2 map to $\mathbb{P}^1$ branched over 4 points, at least one of which is rational. We may choose our coordinate function on $\mathbb{P}^1$ so that the image of our distinguished point P on $C(N,p)$ maps to ∞, and we then have precisely the data for a Weierstrass equation $y^2 = P_3(x)$, and w_N takes $(x,y) \mapsto (x,-y)$. This means that the quadratic twist by $\mathbb{Q}(\sqrt{p^*})$ is the usual twist by p^* in the sense of elliptic curves. In particular we are twisting the L-function by the quadratic character of conductor p, and the theory of signs of functional equations gives the result on the parity of the analytic rank!

The negative result is a consequence of the following

Theorem 5. *Suppose N is prime. Then $C(N,p)(\mathbb{Q}_N) = \emptyset \iff N \equiv 1 \pmod 4$.*

Proof: Well, as it happens we know what the reduction of $X_0(N)$ mod N looks like: two copies of $\mathbb{P}^1$ joined along the supersingular points. The running hypothesis $(\frac{N}{p}) = -1$ means that the special fiber at N of $C(N,p)$ is the quadratic twist by

w_N and the unique quadratic extension $\mathbb{F}_{N^2}/\mathbb{F}_N$. This means that the only $\mathbb{F}_N$-rational points are the $\mathbb{F}_N$-rational supersingular points, which are all singular. It would seem then that Hensel's Lemma is telling us that $C(N,p)(\mathbb{Q}_N) = \emptyset$, but this is obviously wrong so we've been too hasty. Recall that $X_0(N)_{/\mathbb{Z}_N}$ was not necessarily *regular*. Indeed at a supersingular point, the exponent a in the strict complete local ring is equal to half the number of automorphisms of the corresponding elliptic curve E (see the appendix of [EI]). We always have ± 1 as a group of automorphisms; assuming that $N > 3$ for simplicity (we already know what happens in the other cases), then even in characteristic N the only elliptic curves with further automorphisms are the one with j-invariant 0 (which has 6 automorphisms) and the one with j-invariant 1728 (which has 4). Thus if 0 is a supersingular j-invariant, we must blow up $6/2 - 1 = 2$ times to get the regular model, replacing the singular point by a chain of two rational curves. But this doesn't help: the Galois action on this chain is still twisted, so it interchanges the two components of the chain. However, if 1728 is a supersingular j-invariant then we must blow up $4/2 - 1 = 1$ time to get a single rational component, evidently stabilized by the Galois action, giving a genus zero curve over $\mathbb{F}_N$ which necessarily has at least 3 rational points, so at least one nonsingular rational point. By a well-known result of Deuring, 1728 is a supersingular j-invariant iff -1 is not square mod N iff N is not 1 (mod 4). Done!

4. Further study of $C(N,p)$: 3 questions and an answer

It is interesting to try to study the curves $C(N,p)$ more systematically:

Problem.
a) Determine, as explicitly as possible, the set of places $\ell \leq \infty$ for which the curve $C(N,p)(\mathbb{Q}_\ell) = \emptyset$.
b) What can be said about $C(N,p)(\mathbb{Q})$ when the genus is at least two?

As both Jordan Ellenberg and I think that this would make a great thesis problem, maybe I should elaborate a bit (in the notes, if not actually in the talk). The case of genus zero is completely understood: i.e., $C(N,p)$ corresponds to some quaternion algebra over $\mathbb{Q}$: to see which one, see [Ser]. So assume the genus is positive. The Deligne-Rapoport model approach will give an answer at all primes dividing N. One can see that (for all N) $C(N,p)$ has smooth reduction at the primes ℓ not dividing Np; since smooth curves of genus one over finite fields always have rational points, this means that there are $\mathbb{Q}_\ell$-rational points in the genus one case. The most mysterious case is that of $\mathbb{Q}_p$, where I don't know what the special fiber of the minimal model looks like (one can see it is not in general semistable). But a nice result of Gonzalez, described in more detail below, addresses this. So the case of genus one is essentially solved.

At the end of [Cl1] I asked three specific questions about local and global points on $C(N,p)$:

1) Is there a necessary and sufficient condition for $C(N,p)(\mathbb{Q}_p) = \emptyset$ similarly simple to the one for $C(N,p)(\mathbb{Q}_N)$ (i.e., a congruence condition)?

I suspect the answer is yes, and that the explanation will come from combining

Theorem 5 with Gonzalez' theorem, which is intriguing because his proof could not be more different: using η function identities (!!), Gonzalez constructs a finite map to an explicit conic curve, whose corresponding quaternion algebra is $\langle N^a, p^* \rangle$, where the integer a depends on $N \pmod{24}$. At least when N is prime, his result gives precisely the implication $\Longleftarrow$ of Theorem 5, so a natural first guess is that when N is prime, $C(N,p)(\mathbb{Q}_p) = \emptyset \iff C(N,p)(\mathbb{Q}_N) = \emptyset$.

2) Is there ever a deficient prime ℓ not dividing Np?

One should be able to answer this question – at least in any given case – by applying a (suitably w_N-twisted) version of the Eichler-Selberg trace formula. Reasoning by analogy to very similar issues on Shimura curves, I very strongly suspect that the answer is yes. It would be nice to apply the sort of averaging arguments at which the analytic number theorists are so proficient in order to get some idea "how often" this occurs.

3) Is it possible for $C(N,p)$ to have points everywhere locally but not globally?

Note that such behavior cannot occur in genus zero, and it did not turn up in any of our genus one examples (but the cases $N = 14$ and $N = 21$ still deserve atttention). I had expected this to be the hardest of the three questions to answer, but last week I was able to show:

Theorem 6. *For each squarefree $N > 131$ – except $N = 163$, the set of primes $p \equiv 1 \pmod 4$ for which $C(N,p)$ has points everywhere locally but not globally has positive density.*

Indeed the proof I found for this result, suitably abstracted, gave Theorem 1. To deduce Theorem 6 from Theorem 1 – with $\iota = w_N$ – we need only remark that $X_0^+(N)$ has genus at least 2 when $N > 131$, so (iii) follows from Faltings' finiteness theorem. (i) and (ii) have already been checked; note that (iv) holds because $X_0(N)$ always has $\mathbb{Q}$-rational cusps.

Maybe I should end by saying that whereas we were able to give almost the complete story of known realizations of $PSL_2(\mathbb{F}_p)$ over $\mathbb{Q}$, the past few years have seen an explosion in the realization of $PSL_2(\mathbb{F}_{p^r})$ over $\mathbb{Q}$, for congruence conditions on p depending upon r, by methods related to modular forms. See Wiese's recent preprint [Wie] and the references therein to work of Dieulefait and others. I wish I had a better understanding of the relation of Serre's conjecture to the inverse Galois problem for $PSL_2(\mathbb{F}_q)$ (and $PGL_2(\mathbb{F}_q)$): e.g., might it be possible to deduce the *non*existence of certain Galois groups from the nonexistence of certain kinds of modular forms?

5. Recommended reading

The construction of the semistable model $X_0(N)_{\mathbb{Z}}$ is one of many important topics to be found in

[DR] P. Deligne and M. Rapoport, *Schémas de modules des courbes elliptiques*, Springer Lecture Notes in Mathematics 349, 1973.

If one is primarily interested in a *description* of the model and a discussion of the finer points about regularity at the supersingular points needed to construct the regular model, I strongly recommend the appendix, by Mazur and Rapoport, to

[EI] B. Mazur, *Modular curves and the Eisenstein ideal,* Publ. Math. I.H.E.S. 47 (1977).

The main part of this paper develops the theory needed to classify the $\mathbb{Q}$-rational points on $X_1(N)$. (These results were later extended by Kamienny, Abramovich and Merel.) The harder case of $X_0(N)$ is handled in

[RI] B. Mazur, *Rational isogenies of prime degree*, Inventiones math. 44 (1978), 129-162.

Classifying rational isogenies on number fields of higher degree, or even quadratic fields – has proven to be very much harder. Indeed, the problem of rational points on $X_0^+(N)$ is a special case of this. So the results of mine described here can be viewed as a contribution to this problem, representing about ϵ^2 percent of a solution, where ϵ is a positive infinitesimal.

Here is an excellent survey of known and conjectural methods for finding rational points on curves:

[Poo] B. Poonen, *Computing rational points on curves*, 149-172, Number Theory for the Millennium III, A.K. Peters, 2002.

It seems that (not necessarily quadratic) $\mathbb{Q}$-curves were considered explicitly by Ribet – he showed that they can be characterized among complex elliptic curves by being covered by some Riemann surface $X_1(N)_{/\mathbb{C}}$.[18] There are therefore important connections to modular forms that I have not had enough time (or really, expertise) to discuss. To learn more about $\mathbb{Q}$-curves, start with the survey article

[Ell] J. Ellenberg, *Q-curves and Galois representations*, in Modular Curves and Abelian Varieties, 2004.

and follow up on the references therein. (You will find that the material presented here partially answers some of Jordan's questions.) Or take a trip to Barcelona – $\mathbb{Q}$-curves are especially popular among Spanish mathematicians.

On the other hand, most of the fundamental ideas involved in quadratic $\mathbb{Q}$-curves are to be found in two much earlier papers of Kuang-yen Shih:

[Sh1] K.-y. Shih, *On the construction of Galois extensions of function fields and number fields*, Math. Ann. 207 (1974), 99-120.

[18]The horrible terminology is due to B. Gross, who apparently first used it in the context of CM elliptic curves. Professor (and Dean) Gross' contributions to number theory in general and the arithmetic of elliptic curves in particular are such that he may be forgiven for this.

[Sh2] K.-y. Shih, *p-torsion points on certain elliptic curves*, Comp. Math. 36 (1978), 113-129.

Shih's work is distinctly underemphasized in most of the treatises on the inverse Galois problem, the notable exception being

[Ser] *Topics in Galois Theory*, Research Notes in Mathematics 1, Jones and Bartlett, 1992,

which has a very nice treatment of this material (indeed, with the benefit of years of hindsight and his well-known expository gifts, Serre manages to give a significantly simpler presentation of the portion of Shih's work with direct Galois-theoretic applications) and proposes the extension to elliptic curves described above.

Theorems 4 and 5 appear in my short paper

[Cl1] P.L. Clark, *Galois groups via Atkin-Lehner twists*, to appear in Proc. AMS.

The proof was discussed in full here, but the paper also contains some numerical examples and enunciates a conjecture about 3-ranks of imaginary quadratic fields that would serve to make the result unconditional on BSD. The conjecture is close enough to a theorem of Belabas and Fouvry that it might be within reach, although not by me: it is a job for a serious analytic number theorist.

Theorems 1 and 6 appear in my preprint

[Cl2] P.L. Clark, *An Anti-Hasse principle for prime twists*,

of which a written version exists but has not yet been made publicly available (I worry that the appearance of this paper on the arxiv at this time would annoy some people to whom I have some outstanding mathematical commitments). Certainly you may email me for a copy.

Finally, the following preprint (to be found on the arxiv) and its references should be consulted for connections to modular forms and Serre's conjecture.

[Wie] G. Wiese, *On projective lienar groups over finite fields as Galois groups over the rational numbers*.

E-mail address: plclark@gmail.com