Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer Allianz- bzw. Nationallizenz frei zugänglich. / This publication is with permission of the rights owner freely accessible due to an Alliance licence and a national licence respectively.

msp

ALGEBRA AND NUMBER THEORY 15:3 (2021)
https://doi.org/10.2140/ant.2021.15.747

Residual Galois representations of elliptic curves with image contained in the normaliser of a nonsplit Cartan

Samuel Le Fourn and Pedro Lemos

It is known that if $p > 37$ is a prime number and $E/\mathbb{Q}$ is an elliptic curve without complex multiplication, then the image of the mod p Galois representation

$$\bar{\rho}_{E,p} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(E[p])$$

of E is either the whole of $\mathrm{GL}(E[p])$, or is *contained* in the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}(E[p])$. In this paper, we show that when $p > 1.4 \times 10^7$, the image of $\bar{\rho}_{E,p}$ is either $\mathrm{GL}(E[p])$, or the *full* normaliser of a nonsplit Cartan subgroup. We use this to show the following result, partially settling a question of Najman. For $d \geq 1$, let $I(d)$ denote the set of primes p for which there exists an elliptic curve defined over $\mathbb{Q}$ and without complex multiplication admitting a degree p isogeny defined over a number field of degree $\leq d$. We show that, for $d \geq 1.4 \times 10^7$, we have

$$I(d) = \{p \text{ prime} : p \leq d - 1\}.$$

1. Introduction

Let p be a prime, and let E be an elliptic curve defined over $\mathbb{Q}$. Fix an algebraic closure $\overline{\mathbb{Q}}$ of $\mathbb{Q}$, and denote by $E[p]$ be the group of p-torsion points of $E(\overline{\mathbb{Q}})$. This is a 2-dimensional $\mathbb{F}_p$-vector space endowed with an $\mathbb{F}_p$-linear action of the Galois group $G_{\mathbb{Q}} := \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. We thus have an associated Galois representation

$$\bar{\rho}_{E,p} : G_{\mathbb{Q}} \to \mathrm{GL}(E[p]).$$

When E does not have complex multiplication, Serre [1972] shows that, for p large enough, the image of $\bar{\rho}_{E,p}$ is the whole of $\mathrm{GL}(E[p])$. In the same paper he asks whether it is possible to prove a uniform lower bound exists for his result to hold, i.e., whether there exists a positive constant B such that if $E/\mathbb{Q}$ is an elliptic curve without complex multiplication and p is a prime larger than B, then $\bar{\rho}_{E,p}$ is surjective. This problem is commonly known as Serre's uniformity question. The progress made towards finding an answer to it can be summarised in the following result, due to the work of several mathematicians,

Lemos is funded by the Royal Society Research Fellows Enhancement Award RGF\EA\181052.
MSC2010: primary 11G05; secondary 11G18.
Keywords: Galois representations of elliptic curves, Serre uniformity problem, nonsplit Cartan subgroup.

amongst whom we highlight Bilu, Mazur, Parent, Rebolledo and Serre (the terminology used in the statement of the following theorem will be explained in the next section).

Theorem 1.1 [Mazur 1978; Bilu and Parent 2011b; Bilu et al. 2013; Serre 1981]. *Let $E/\mathbb{Q}$ be an elliptic curve without complex multiplication. Suppose that p is a prime not lying in the set $\{2,3,5,7,11,13,17,37\}$. If the image of $\bar{\rho}_{E,p}$ is not $\mathrm{GL}(E[p])$, then it is **contained** in the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}(E[p])$.*

The main result of this paper is the following improvement on Theorem 1.1.

Theorem 1.2. *Let $E/\mathbb{Q}$ be an elliptic curve without complex multiplication. Let p be a prime number, and suppose that one of the following statements holds*:

(a) $p > 1.4 \times 10^7$.

(b) $p \notin \{2, 3, 5, 7, 11, 13, 17, 37\}$ *and* $j(E) \notin \mathbb{Z}$.

*If $\bar{\rho}_{E,p}$ is not surjective, then its image **is** the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}(E[p])$.*

We will at times mention Theorem 1.2(a) or Theorem 1.2(b), by which we mean the result of Theorem 1.2 assuming condition (a) or (b), respectively.

The proof of Theorem 1.2(a) (Section 6) shows, in fact, that if an elliptic curve E without complex multiplication is such that $j(E) \in \mathbb{Z}$ and admits a prime p not in the set $\{2, 3, 5, 7, 11, 13, 17, 37\}$ such that the image of $\bar{\rho}_{E,p}$ is neither $\mathrm{GL}(E[p])$ nor the normaliser of a nonsplit Cartan subgroup, then $\log|j(E)| \leq \max(12000, 7\sqrt{p}) \leq 27000$. In particular, there are only finitely many such elliptic curves up to isomorphism. One would could then hope that the remaining cases might be treated algorithmically, but the authors admit they could not find a reasonably efficient way to do so. However, we wish to point out that some work has already been done in this direction. For example, in [Bajolet et al. 2012], the integral points of $X_{\mathrm{ns}}^{+}(p)$ are determined for all primes $p \leq 97$. Unfortunately, the algorithms employed there (which are already great improvements over existing techniques) need several CPU years to compute even the single case $p = 97$. Solving the remaining cases $p \leq 1.4 \cdot 10^7$ in our case thus appears as a serious technical challenge deserving of its own project.

As an immediate application of Theorem 1.2, we are able to partially settle a question of Najman [2018]. Let $d \geq 1$ be a positive integer. Najman [2018] defines $I(d)$ to be the set of primes p for which there exists an elliptic curve E defined over $\mathbb{Q}$ without complex multiplication and an isogeny $\phi : E_{/K} \to E'$ of degree p defined over a number field K of degree $\leq d$. For instance, a celebrated result of Mazur [1978] states that

$$I(1) = \{2, 3, 5, 7, 11, 13, 17, 37\}.$$

Najman [2018] shows that

$$I(d) \subseteq I(1) \cup \big\{p \text{ prime} : p \leq d-1 \text{ when } p \equiv 1 \pmod 3\big\} \cup \big\{p \text{ prime} : p \leq 3d-1 \text{ when } p \equiv 2 \pmod 3\big\}.$$

Assuming that $\bar{\rho}_{E,p}$ is surjective whenever $E/\mathbb{Q}$ does not have complex multiplication and $p \notin I(1)$, he proves that one has

$$I(d) = I(1) \cup \{p \text{ prime} : p \leq d-1\}. \tag{1-1}$$

Theorem 1.2 allows us to remove the condition on the surjectivity of $\bar{\rho}_{E,p}$, albeit adding one on the size of d.

Theorem 1.3. *For $d \geq 1.4 \times 10^7$, we have $I(d) = \{p \text{ prime} : p \leq d-1\}$.*

The proof of this result is a simple combination of Theorem 1.2 and Najman's own arguments. We refer the reader to [Najman 2018] for details.

Plan of the proof of Theorem 1.2. From now on, we assume that a basis of $E[p]$ has been chosen and systematically identify $\mathrm{GL}(E[p])$ with $\mathrm{GL}_2(\mathbb{F}_p)$.

The following result of Zywina will be used to prove Theorem 1.2.

Proposition 1.4 [Zywina 2015, Proposition 1.13]. *Let $E/\mathbb{Q}$ be an elliptic curve without complex multiplication. Let $p \notin I(1)$ be a prime such that $\bar{\rho}_{E,p}$ is not surjective.*

(1) *If $p \equiv 1 \pmod 3$, then $\bar{\rho}_{E,p}(G_{\mathbb{Q}})$ is the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$.*

(2) *If $p \equiv 2 \pmod 3$, then $\bar{\rho}_{E,p}(G_{\mathbb{Q}})$ is either the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, or is conjugate in $\mathrm{GL}_2(\mathbb{F}_p)$ to the group*

$$G(p) := \left\{a^3 : a \in C_{\mathrm{ns}}(p)\right\} \cup \left\{\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \cdot a^3 : a \in C_{\mathrm{ns}}(p)\right\},$$

where $C_{\mathrm{ns}}(p)$ is an explicit choice of nonsplit Cartan subgroup made in the next section.

Remark 1.5. For the convenience of the reader, and following a suggestion made by an anonymous referee, we reproduce Zywina's proof of this result in Appendix B.

We are then reduced to showing that when $p \equiv 2 \pmod 3$ the image of $\bar{\rho}_{E,p}$ cannot be contained in a conjugate of $G(p)$. When the j-invariant of $E/\mathbb{Q}$ is not integral, we will rule out this possibility using Mazur's formal immersion argument (see [Mazur 1978]). More precisely, an elliptic curve defined over $\mathbb{Q}$ whose residual Galois representation $\bar{\rho}_{E,p}$ has image contained in $G(p)$ will give rise to a $\mathbb{Q}$-rational point x on a modular curve $X_{G(p)}$. If the j-invariant is not in $\mathbb{Z}$, then some prime ℓ divides the denominator. We will first point out that ℓ cannot be p (this is Proposition 5.2). It will then follow that there exists a $\mathbb{Q}$-rational point x in the residue class modulo λ of a cusp c (here, λ is a prime of the residue field of the cusp c dividing ℓ). We will show the existence of a nontrivial quotient of the jacobian of $X_{G(p)}$ with finite Mordell–Weil group (this is Section 3) and use the standard formal immersion arguments due to Mazur to prove that such a point cannot exist (Sections 4 and 5). This will give us Theorem 1.2(b).

Remark 1.6. The reader will notice that this situation contrasts with that of the modular curve $X_{\mathrm{ns}}^+(p)$ associated to the normaliser of a nonsplit Cartan. Indeed, it is well-known that the conjecture of Birch and Swinnerton-Dyer implies the inexistence of a nontrivial quotient of the jacobian of $X_{\mathrm{ns}}^+(p)$ with finite

Mordell–Weil group. This is a major obstacle to the study of the rational points of $X_{\mathrm{ns}}^{+}(p)$, and thus to giving a positive answer to Serre's uniformity question.

In the case where $j(E) \in \mathbb{Z}$, the assumptions on the mod p Galois representation of E give rise to an integral point on $X_{G(p)}$. We then follow the steps of Bilu and Parent [2011b] as follows. First, by applying Runge's method, we obtain an upper bound for $\log|j(E)|$ which is linear in $\sqrt{p}$. An explicit version of Serre's surjectivity theorem obtained by the first author on the basis of isogeny theorems of Gaudron and Rémond [2014] provides a lower bound linear in p, which gives rise to a contradiction for $p \geq 1.4 \times 10^7$.

2. Cusps of modular curves

We give a brief review of some basic facts about cusps of modular curves and set down some notation that will be used later in the paper. The reader should be warned that our definition of *cusps at infinity* differs slightly from the standard one.

Let p be an odd prime, and let $X(p)$ be the (compactification of the) classical modular curve which classifies pairs $(E, (P, Q))$, where E is an elliptic curve and the pair (P, Q) is an $\mathbb{F}_p$-basis of $E[p]$. This is a smooth projective curve over $\mathbb{Q}$ whose base change to $\mathbb{Q}(\zeta_p)$ has $p-1$ connected components. Given a subgroup H of $\mathrm{GL}_2(\mathbb{F}_p)$, we will denote the modular curve $H\backslash X(p)$ by X_H.

Define

$$M_p := \big((\mathbb{Z}/p\mathbb{Z})^2 - \{(0,0)\}\big)/\pm 1. \tag{2-1}$$

If we regard the elements of M_p as column vectors, we have a natural left action of $\mathrm{GL}_2(\mathbb{F}_p)$ on M_p. We can therefore define an action of $\mathrm{GL}_2(\mathbb{F}_p)$ on $M_p \times \mathbb{F}_p^\times$ by letting $\mathrm{GL}_2(\mathbb{F}_p)$ act on $\mathbb{F}_p^\times$ via multiplication by the determinant.

Lemma 2.1. *There is a bijection between the cusps of $X(p)$ and the set $M_p \times \mathbb{F}_p^\times$ which is equivariant for the action of $\mathrm{GL}_2(\mathbb{F}_p)$. Moreover, if $\sigma \in G_{\mathbb{Q}}$ and c is a cusp of $X(p)$ corresponding to the pair $\left(\binom{a}{b}, d\right)$ then ${}^{\sigma}c$ corresponds to*

$$\sigma \cdot \left(\binom{a}{b}, d\right) := \left(\chi_p(\sigma)^{-1}\binom{a}{b}, \chi_p(\sigma)^{-1}d\right),$$

where χ_p is the cyclotomic character.

Proof. Following [Deligne and Rapoport 1973, VI.5], we have a canonical Galois equivariant bijection between the cusps of $X(p)$ and the set

$$\mathrm{Isom}\big(\mu_p \times \mathbb{Z}/p\mathbb{Z}, (\mathbb{Z}/p\mathbb{Z})^2\big)/\pm U,$$

where U is the group of matrices

$$\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \quad u \in \mathrm{Hom}(\mathbb{Z}/p\mathbb{Z}, \mu_p),$$

and the (left) action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is induced by its natural action on μ_p and trivial one on $\mathbb{Z}/p\mathbb{Z}$. Furthermore, the action of $\mathrm{GL}_2(\mathbb{F}_p)$ corresponds to composition (in other words, to left matrix multiplication).

Given a class γ in $\mathrm{Isom}\big(\mu_p \times \mathbb{Z}/p\mathbb{Z}, (\mathbb{Z}/p\mathbb{Z})^2\big)/\pm U$ represented by

$$(\zeta_p, 0) \mapsto (a, b), \quad (1, 1) \mapsto (c, d),$$

we associate to it the element

$$\left(\begin{pmatrix} a \\ b \end{pmatrix}, \det \gamma\right) \in M_p \times \mathbb{F}_p^\times,$$

where the determinant of γ is defined to be $ad - bc$. It is easy to see that this function is well defined. It is also clear that this function commutes with the actions of $\mathrm{GL}_2(\mathbb{F}_p)$ and of the Galois group, so we just need to check that it is a bijection. But this is clearly surjective, as, given a pair $(a, b) \in (\mathbb{Z}/p\mathbb{Z})^2 - \{(0, 0)\}$, it is always possible to find a pair (c, d) such that $ad - bc$ is equal to a given element of $\mathbb{F}_p^\times$. As the two sets have the same number of elements, the result follows. □

Corollary 2.2. *If H is a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, then there is a bijection between the set of cusps of X_H and the set $H\backslash(M_p \times \mathbb{F}_p^\times)$. Moreover, if $\det H = \mathbb{F}_p^\times$, this bijection induces a bijection between the set of cusps of X_H and $(H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p$.*

Proof. The first assertion follows immediately from Lemma 2.1 and the definition of X_H. In order to prove the second one, start by observing that, given a class in $H\backslash(M_p \times \mathbb{F}_p^\times)$, there is always a representative of this class whose second entry is 1 (this is due to the assumption that $\det H = \mathbb{F}_p^\times$). Therefore, the map $(H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p \to H\backslash(M_p \times \mathbb{F}_p^\times)$ given by

$$\begin{pmatrix} a \\ b \end{pmatrix} \mapsto \left(\begin{pmatrix} a \\ b \end{pmatrix}, 1\right)$$

is well-defined and bijective. □

Corollary 2.3. *Let H be a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ such that $\det H = \mathbb{F}_p^\times$. Under the identification of Corollary 2.2, the Galois orbit of a cusp of X_H represented by an element $\binom{a}{b}$ of M_p is the set of cusps of X_H represented by the elements in the set*

$$\left\{\gamma \cdot \begin{pmatrix} a \\ b \end{pmatrix} \in M_p : \gamma \in H\right\}.$$

In particular, we obtain a one-to-one correspondence between the Galois orbits of cusps of X_H and the set $H\backslash M_p$.

Proof. For each $\lambda \in \mathbb{F}_p^\times$, choose $\gamma_\lambda \in H$ such that $\det \gamma_\lambda = \lambda$. The first observation we want to make is that we have the following equality of sets:

$$H = \big\{h\gamma_\lambda : \lambda \in \mathbb{F}_p^\times,\ h \in H \cap \mathrm{SL}_2(\mathbb{F}_p)\big\}. \tag{2-2}$$

Indeed, if $g \in H$, and if we set $d := \det g$, we have $g\gamma_d^{-1} \in H \cap \mathrm{SL}_2(\mathbb{F}_p)$. Thus, g is of the form $h\gamma_d$ for some $h \in H \cap \mathrm{SL}_2(\mathbb{F}_p)$. The other inclusion is obvious.

According to Lemma 2.1, the Galois orbit of a cusp represented by $\left(\binom{a}{b}, 1\right)$ is the set of cusps represented by the elements of the set

$$\Sigma := \left\{ \left(\begin{pmatrix} \lambda a \\ \lambda b \end{pmatrix}, \lambda \right) \in M_p \times \mathbb{F}_p^\times : \lambda \in \mathbb{F}_p^\times \right\}.$$

Of course, the set of cusps of X_H represented by Σ and the set of those represented by

$$\gamma_\lambda^{-1} \cdot \Sigma = \left\{ \left(\lambda \gamma_\lambda^{-1} \begin{pmatrix} a \\ b \end{pmatrix}, 1 \right) \in M_p \times \mathbb{F}_p^\times : \lambda \in \mathbb{F}_p^\times \right\}$$

is the same. As $\det H = \mathbb{F}_p^\times$, we know from Corollary 2.2 that we can also identify the set of cusps of X_H with $(H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p$. Therefore, the Galois orbit of our cusp is the set of cusps of X_H represented by the elements of the set

$$\left\{ \lambda \gamma_\lambda^{-1} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \in M_p : \lambda \in \mathbb{F}_p^\times \right\}. \tag{2-3}$$

As $\det \lambda\gamma_\lambda^{-1} = \lambda$, we see that $\{\lambda\gamma_\lambda^{-1} : \lambda \in \mathbb{F}_p^\times\}$ runs through a set of representatives of $(H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash H$. Equality (2-2) and the fact that the cusps of X_H represented by a set of elements of M_p does not change under the action of an element of $H \cap \mathrm{SL}_2(\mathbb{F}_p)$, yield that the Galois orbit of our cusp is represented by the set

$$\left\{ \gamma \cdot \begin{pmatrix} a \\ b \end{pmatrix} \in M_p : \gamma \in H \right\},$$

as we wanted. □

We define the *cusps at infinity* of a modular curve X_H to be those cusps represented by elements of $M_p \times \mathbb{F}_p^\times$ of the form

$$\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, a \right), \quad a \in \mathbb{F}_p^\times$$

under the identification of Corollary 2.2. Note that the set of cusps at infinity of X_H forms a full Galois orbit.

Before finishing this section, we wish to mention some of the modular curves that we will use throughout this article. We start by considering the case where H is the upper triangular subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. In this case, the curve X_H is usually denoted by $X_0(p)$. This modular curve has two distinct cusps: one cusp at infinity and one not at infinity, as one can easily check using the identification of Corollary 2.2. Both cusps are defined over $\mathbb{Q}$. The cusp at infinity will be denoted by ∞, while the other one will be denoted by 0.

Let $C_{\mathrm{sp}}(p)$ be the split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ consisting of diagonal matrices, i.e.,

$$C_{\mathrm{sp}}(p) := \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_p) : a, b \in \mathbb{F}_p^\times \right\}.$$

When $H = C_{\mathrm{sp}}(p)$, we will denote X_H by $X_{\mathrm{sp}}(p)$. The normaliser of $C_{\mathrm{sp}}(p)$ will be denoted by $N_{\mathrm{sp}}(p)$. This is the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ consisting of diagonal and antidiagonal matrices. When $H = N_{\mathrm{sp}}(p)$, the curve X_H will be denoted by $X_{\mathrm{sp}}^+(p)$. We have a canonical morphism $X_{\mathrm{sp}}(p) \to X_{\mathrm{sp}}^+(p)$ of degree 2, which is unramified at the cusps. The curve $X_{\mathrm{sp}}^+(p)$ has $(p+1)/2$ cusps, of which exactly one is at

infinity. The cusp at infinity is defined over $\mathbb{Q}$, while the others are defined over $\mathbb{Q}(\zeta_p)^+ := \mathbb{Q}(\zeta_p + \zeta_p^{-1})$. The preimage of the cusp at infinity of $X_{\mathrm{sp}}^+(p)$ in $X_{\mathrm{sp}}(p)$ consists of two cusps, both defined over $\mathbb{Q}$, of which one is at infinity and the other is not. The remaining $p-1$ cusps of $X_{\mathrm{sp}}(p)$ are defined over $\mathbb{Q}(\zeta_p)$.

Finally, we make the following choice of nonsplit Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Fix a generator ϵ_p of the cyclic group $\mathbb{F}_p^\times$. Define

$$C_{\mathrm{ns}}(p) := \left\{ \begin{pmatrix} a & \epsilon_p b \\ b & a \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_p) : a, b \in \mathbb{F}_p^\times \right\}.$$

Its normaliser will be denoted by $N_{\mathrm{ns}}(p)$. Explicitly, this is given by

$$N_{\mathrm{ns}}(p) = C_{\mathrm{ns}}(p) \cup \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} C_{\mathrm{ns}}(p).$$

We have a canonical finite morphism $X_{\mathrm{sp}}(p) \to X_{\mathrm{sp}}^+(p)$ of degree 2. The modular curve $X_{\mathrm{ns}}^+(p)$ has $(p-1)/2$ cusps, all of them at infinity. Their field of definition is $\mathbb{Q}(\zeta_p)^+$. The modular curve $X_{\mathrm{ns}}(p)$, on the other hand, has $p+1$ distinct cusps. Like in the case of $X_{\mathrm{ns}}^+(p)$, they are all at infinity, but their field of definition is now $\mathbb{Q}(\zeta_p)$.

3. Quotients of modular jacobians with finite Mordell–Weil group

Let p be an odd prime, and let $H(p)$ denote the group $N_{\mathrm{ns}}(p) \cap N_{\mathrm{sp}}(p)$. Let H be a subgroup of $N_{\mathrm{ns}}(p)$ of index $d \geq 2$ containing $H(p)$. We shall write J_H for the jacobian of the modular curve X_H. The main result of this section is the following proposition.

Proposition 3.1. *Suppose that $p = 11$ or $p \geq 17$. Then the jacobian J_H of X_H admits a nontrivial optimal quotient A such that*

(1) *$A(\mathbb{Q})$ is finite;*

(2) *the kernel of the canonical projection $J_H \to A$ is stable under the action of the Hecke operators T_ℓ, where $\ell \neq p$ is a prime.*

Let $\pi_{\mathrm{ns}} : X_{H(p)} \to X_{\mathrm{ns}}^+(p)$ denote the canonical projection. As H is contained in $N_{\mathrm{ns}}(p)$ and contains $H(p)$ by assumption, the morphism π_{ns} factors through X_H:

$$\begin{array}{ccc} & & X_{H(p)} \\ & {\scriptstyle \pi_H} \swarrow & \downarrow {\scriptstyle \pi_{\mathrm{ns}}} \\ X_H & \xrightarrow[\pi'_H]{} & X_{\mathrm{ns}}^+(p) \end{array} \tag{3-1}$$

As the index of H in $N_{\mathrm{ns}}(p)$ is d, we have $\deg \pi'_H = d$ and $\deg(\pi_H) = (p+1)/(2d)$. Moreover, the morphisms π_H and π'_H are unramified at the cusps because π_{ns} is.

Lemma 3.2. *The modular curve X_H has cusps not at infinity.*

Proof. In the language of Corollary 2.2, the set of cusps of $X_{H(p)}$ is identified with the set

$$H(p)\backslash M_p \times \mathbb{F}_p^\times,$$

and the cusps at infinity are, by definition, those represented by an element of the form

$$\left(\binom{a}{0}, a\right), \quad a \in \mathbb{F}_p^\times.$$

It follows that $X_{H(p)}$ has $(p-1)/2$ cusps at infinity. Therefore, X_H has at most $(p-1)/2$ cusps at infinity. As π'_H is a morphism of degree d unramified at the cusps, the number of cusps of X_H is $d(p-1)/2$ (recall that $X_{\mathrm{ns}}^+(p)$ has $(p-1)/2$ cusps). Since $d \geq 2$ by assumption, this number is strictly larger than $(p-1)/2$. As a consequence, there exists a cusp of X_H which is not at infinity. □

Remark 3.3. Note that the proof of Lemma 3.2 relies crucially on the fact that $d > 1$. If $d = 1$, then X_H is the modular curve $X_{\mathrm{ns}}^+(p)$ and, indeed, all of its cusps are at infinity.

Let $\pi_{\mathrm{sp}} : X_{H(p)} \to X_{\mathrm{sp}}^+(p)$ be the canonical projection. By pulling back by π_H and pushing forward via π_{sp}, we obtain a morphism

$$\phi := \pi_{\mathrm{sp},*} \circ \pi_H^* : X_H \to \mathrm{Pic}(X_{\mathrm{sp}}^+(p)).$$

Lemma 3.4. *If $c \in X_{H(p)}(\mathbb{Q}(\zeta_p))$ is not a cusp at infinity, then $\pi_{\mathrm{sp}}(c)$ is not at infinity.*

Proof. Let $\left(\binom{a}{b}, d\right)$ be an element of $M_p \times \mathbb{F}_p^\times$ representing c. Suppose, for contradiction, that $\pi_{\mathrm{sp}}(c)$ is at infinity. Then there exists $\gamma \in N_{\mathrm{sp}}(p)$ and $\alpha \in \mathbb{F}_p^\times$ such that

$$\gamma \cdot \binom{a}{b} = \binom{\alpha}{0}.$$

It follows that $a = 0$ or $b = 0$. However, the cusps of $X_{H(p)}$ represented by elements of $M_p \times \mathbb{F}_p^\times$ of the form

$$\left(\binom{a}{0}, d\right) \quad \text{or} \quad \left(\binom{0}{b}, d\right)$$

are easily seen to be cusps at infinity, which is a contradiction. □

We also recall a well-known morphism

$$\eta : X_{\mathrm{sp}}^+(p) \to J_0(p)/(1+w_p)J_0(p)$$

(where w_p is the Atkin–Lehner involution of $J_0(p)$) that has been studied, for instance, by Mazur [1978] and by Momose [1984]. This morphism is defined as follows. Start by considering the two degeneration maps

$$d_1 : X_{\mathrm{sp}}(p) \to X_0(p) \quad \text{and} \quad d_p : X_{\mathrm{sp}}(p) \to X_0(p)$$

whose moduli interpretations are

$$d_1 : (E, (A, B)) \mapsto (E, A) \quad \text{and} \quad d_p : (E, (A, B)) \mapsto (E/B, E[p]/B).$$

We define an auxiliary morphism

$$\eta' : X_{\mathrm{sp}}(p) \to J_0(p)$$

by requiring that a point $x \in X_{\mathrm{sp}}(p)(\overline{\mathbb{Q}})$ be mapped to the class of the divisor $d_1(x) - d_p(x)$. Letting ω_p denote the involution of $X_{\mathrm{sp}}(p)$ whose moduli interpretation is

$$(E, (A, B)) \mapsto (E, (B, A)),$$

(where A and B are distinct subgroups of $E(\overline{\mathbb{Q}})$ of order p) it is easy to check that the following relations hold:

$$w_p \circ d_p = d_1 \circ \omega_p \quad \text{and} \quad w_p \circ d_1 = d_p \circ \omega_p. \tag{3-2}$$

It follows that

$$\eta' \circ \omega_p = -w_p \circ \eta'.$$

Therefore, the morphism

$$X_{\mathrm{sp}}(p) \xrightarrow{\eta'} J_0(p) \to J_0(p)/(1 + w_p)J_0(p)$$

factors through $X^+_{\mathrm{sp}}(p)$. The morphism $X^+_{\mathrm{sp}}(p) \to J_0(p)/(1 + w_p)J_0(p)$ through which it factors is the morphism η aforementioned.

Despite the following lemma being a well-known result, the authors were not able to find a reference offering a concise proof. Due to this, a proof of this lemma can be found in Appendix A.

Lemma 3.5. *Let $c \in X^+_{\mathrm{sp}}(p)(\mathbb{Q}(\zeta_p))$ be a cusp. We have*

$$\eta(c) = \begin{cases} 0 & \text{if } c \text{ is at infinity,} \\ \mathrm{cl}(0 - \infty) & \text{otherwise.} \end{cases}$$

By abuse of notation, we shall denote by η the map $\mathrm{Pic}(X^+_{\mathrm{sp}}(p)) \to J_0(p)/(1 + w_p)J_0(p)$ obtained from η using the universal property of jacobians. Define ν to be the composition

$$\eta \circ \phi : X_H \to J_0(p)/(1 + w_p)J_0(p),$$

which is clearly a morphism defined over $\mathbb{Q}$.

Lemma 3.6. *Let $c \in X_H(\mathbb{Q}(\zeta_p))$ be a cusp. If c is a cusp at infinity, we have*

$$\nu(c) = \left(\frac{p+1}{2d} - 1\right)\mathrm{cl}(0 - \infty).$$

If, on the other hand, c is not at infinity, we have

$$\nu(c) = \left(\frac{p+1}{2d}\right)\mathrm{cl}(0 - \infty).$$

Proof. If c is not at infinity, the pull-back of c by π_H is a sum of $(p+1)/2d$ cusps of $X_{H(p)}$ not at infinity. Then, using Lemma 3.4, we conclude that $\phi(c)$ is a sum of $(p+1)/2d$ cusps of $X_{\mathrm{sp}}^+(p)$ not at infinity. The image of this divisor under η is then

$$\left(\frac{p+1}{2d}\right)\mathrm{cl}(0-\infty)$$

by Lemma 3.5.

If c is a cusp at infinity, then, using the language of Corollary 2.2, it is represented by an element of $M_p \times \mathbb{F}_p^\times$ of the form

$$\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, a\right), \quad a \in \mathbb{F}_p^\times.$$

The pullback of c by π_H corresponds then to pulling back this element by the map

$$H(p)\backslash M_p \times \mathbb{F}_p^\times \to H\backslash M_p \times \mathbb{F}_p^\times.$$

There is only one cusp at infinity of $X_{H(p)}$ in the preimage of c. Indeed, the cusps in the preimage of c are all of the form γc for some $\gamma \in H \subseteq N_{\mathrm{ns}}(p)$. But the only elements of $N_{\mathrm{ns}}(p)$ fixing the line generated by $\binom{a}{0}$ are in $H(p)$, which proves the claim. It then follows from Lemma 3.4 that $\phi(c)$ is the sum of one cusp at infinity with $(p+1)/2d-1$ cusps not at infinity. Using Lemma 3.5, we conclude that

$$\nu(c) = \left(\frac{p+1}{2d} - 1\right)\mathrm{cl}(0-\infty),$$

as we wanted. □

Let $\tilde{J}_p$ be the Eisenstein quotient of $J_0(p)$ (see [Mazur 1977, II.10 Definitions 10.4]), and let ϑ be the composition

$$J_H \to J_0(p)/(1+w_p)J_0(p) \xrightarrow{\mathrm{pr}} \tilde{J}_p,$$

where the first map is the one induced by ν, and pr is the natural projection.

Proof of Proposition 3.1. Denote the image of ϑ in $\tilde{J}_p$ by B. This is an abelian variety defined over $\mathbb{Q}$, and we need to show that it is not trivial. Let c be a cusp of X_H at infinity, and let c' be one not at infinity (the existence of such a cusp is guaranteed by Lemma 3.2). Then, by Lemma 3.6, we have

$$\vartheta(\mathrm{cl}(c'-c)) = \mathrm{pr}(\mathrm{cl}(0-\infty)).$$

A theorem of Mazur [1977, III.1 Corollary 1.4] now yields that the order of $\mathrm{pr}(\mathrm{cl}(0-\infty))$ is

$$(p-1)/\gcd(p-1, 12),$$

which is not 1 because $p = 11$ or $p \geq 17$. This shows that B is not trivial. Also, $B(\mathbb{Q})$ is finite because $\tilde{J}_p(\mathbb{Q})$ is.

Let K denote the kernel of ϑ, and let K^0 denote the connected component of the identity. We define A to be J_H/K^0. As A is $\mathbb{Q}$-isogenous to B, it follows that it is not trivial and that $A(\mathbb{Q})$ is finite. This proves statement (1).

Let ℓ be a prime different from p. One easily checks that the morphism

$$J_H \to J_0(p)/(1+w_p)J_0(p)$$

commutes with the action of T_ℓ. By the work of Mazur [1978], we already know that the Hecke operator T_ℓ preserves the kernel of the projection $J_0(p)/(1+w_p)J_0(p) \to \tilde{J}_p$, from where it follows that it also preserves K. As T_ℓ is an endomorphism of abelian varieties (and is, in particular, continuous), it maps K^0 to itself. This finishes the proof of the proposition. □

4. Formal immersions

For a cusp c of X_H, let

$$\iota_c : X_{H/\mathbb{Q}(\zeta_p)} \to J_{H/\mathbb{Q}(\zeta_p)}$$

be the Abel–Jacobi map centred at c. Let A be a nontrivial quotient of J_H satisfying the conditions of Proposition 3.1. Define $f_c : X_{H/\mathbb{Q}(\zeta_p)} \to A_{/\mathbb{Q}(\zeta_p)}$ to be the composition

$$X_{H/\mathbb{Q}(\zeta_p)} \xrightarrow{\iota_c} J_{H/\mathbb{Q}(\zeta_p)} \to A_{/\mathbb{Q}(\zeta_p)},$$

where the second map is the canonical projection.

Let R be the ring $\mathbb{Z}[\zeta_p, 1/p]$. Let $X_{H/R}$ be the minimal regular model of X_H over R, and let $A_{/R}$ be the Néron model of A over R. The Néron mapping property allows us to extend the morphism f_c to a morphism

$$f_{c/R} : X_{H/R} \to A_{/R}$$

over R. If R' is an R-algebra, we will write $X_{H/R'}$, $A_{/R'}$ and $f_{c/R'}$ for the base change of $X_{H/R}$, $A_{/R}$ and $f_{c/R}$ to R'.

Recall that if X and Y are two noetherian schemes and $\gamma : X \to Y$ is a morphism, we say that γ is a *formal immersion* at the point $x \in X$ if the induced homomorphism

$$\hat{\gamma}_x^{\#} : \hat{\mathcal{O}}_{Y,f(x)} \to \hat{\mathcal{O}}_{X,x}$$

of completed local rings is surjective.

Given a prime ideal λ of R, we will write R_λ for the λ-adic completion of R at λ.

Proposition 4.1. *Let λ be a maximal ideal of R whose characteristic is $\neq 2$ (it is also different from p because p is a unit in R). The morphism f_{c/R_λ} is a formal immersion at $\tilde{c}$, where $\tilde{c}$ stands for the reduction of c modulo λ.*

As $\tilde{c}$ and $f(\tilde{c}) = 0$ are both defined over $k(\lambda)$, the residue field of R_λ, proving this proposition is equivalent to showing that the induced $k(\lambda)$-linear map of cotangent spaces of the special fibres

$$f^*_{c/k(\lambda)} : \mathrm{Cot}(A_{/k(\lambda)}) \to \mathrm{Cot}_{\tilde{c}}(X_{H/k(\lambda)}) \tag{4-1}$$

is surjective. As $X_{H/k(\lambda)}$ is 1-dimensional, it is enough to show that $f^*_{c/k(\lambda)}$ is not trivial. To prove this we will make use of a result due to Mazur [1978, Lemma 2.1], that we now recall.

Mazur's lemma. The content of this subsection is completely contained in a more general form in Mazur's paper [1978]. Given a cusp $c \in X_{G(p)}(\mathbb{Q}(\zeta_p))$, we will denote by

$$\iota_{c/R} : X_{H/R} \to J_{H/R}$$

the extension of the Abel–Jacobi map ι_c to R. We obtain a homomorphism

$$\iota^*_{c/R} : \mathrm{Cot}(J_{H/R}) \to \mathrm{Cot}_c(X_{H/R}) \tag{4-2}$$

of free R-modules.

A theorem of Raynaud asserts that the Picard variety $\mathrm{Pic}^0_{X_{H/R}}$ of $X_{H/R}$ is canonically identified with $J_{H/R}$. This identification induces an isomorphism between the respective tangent spaces:

$$i : \mathrm{H}^1(X_{H/R}, \mathcal{O}_{X_{H/R}}) \to \mathrm{Tan}(J_{H/R}).$$

Of course, $\mathrm{Cot}(J_{H/R})$ is naturally the R-dual of $\mathrm{Tan}(J_{H/R})$, while Grothendieck–Serre duality establishes an R-duality between $\mathrm{H}^0(X_{H/R}, \Omega^1_{X_H})$ and $\mathrm{H}^1(X_{H/R}, \mathcal{O}_{X_H/R})$. We thus obtain an isomorphism

$$\Theta : \mathrm{Cot}(J_{H/R}) \to \mathrm{H}^0(X_{H/R}, \Omega^1_{X_{H/R}}).$$

The natural homomorphism $v : \mathrm{H}^0(X_{H/R}, \Omega^1_{X_{H/R}}) \to \mathrm{Cot}_c(X_{H/R})$ gives then rise to a homomorphism $v \circ \Theta$ from $\mathrm{Cot}(J_{H/R})$ to $\mathrm{Cot}_c(X_{H/R})$. The following lemma, due to Mazur, says that this homomorphism is, up to a sign, $\iota^*_{c/R}$.

Lemma 4.2 [Mazur 1978]. $\iota^*_{c/R} = \pm v \circ \Theta$.

The reason why the homomorphism $v \circ \Theta$ is so useful is because we can explicitly write v in terms of the q-expansion at c of global differential forms of X_H: the point is that we can identify $\mathrm{Cot}_c(X_{H/R})$ with R in such a way that the diagram

$$\begin{array}{ccc} \mathrm{H}^0(X_{H/R}, \Omega^1_{X_H}) & \xrightarrow{q\text{-exp}} & \mathbb{Z}[\![q^{1/p}]\!] \otimes_{\mathbb{Z}} R \\ \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle \sum a_i q^{i/p} \mapsto a_1} \\ \mathrm{Cot}_c(X_{H/R}) & \xrightarrow{\cong} & R \end{array}$$

commutes.

Formal immersion at the cusps. The last results we need before we are ready to prove Proposition 4.1 are the following two lemmas.

Lemma 4.3. *Let ω be an element of $\mathrm{H}^0\big(X_{H/\mathbb{Z}[1/p]}, \Omega^1_{X_H/\mathbb{Z}[1/p]}\big)$. Let c be a cusp of X_H and let*

$$\sum_{n=1}^{\infty} a_n(c, \omega) q^{n/p} \in \mathbb{Z}[\![q^{1/p}]\!] \otimes_{\mathbb{Z}} R$$

be the q-expansion of ω at c. If $\sigma \in G_{\mathbb{Q}}$, then $a_n({}^{\sigma}c, \omega) = {}^{\sigma}a_n(c, \omega)$.

Proof. The result follows from the analogous assertion for $X(p)$, so we show that the lemma holds in this case. Indeed, recall (see [Katz 1973, 1.2]) that the q-expansion of a modular form f at a cusp of $X(p)$ is the evaluation of f at the triple $(\mathrm{Tate}(q), \omega_{\mathrm{can}}, \alpha_p)$, where $\mathrm{Tate}(q)$ is the Tate curve over $\mathbb{Z}((q^{1/p})) \otimes R$, ω_{can} is its canonical differential, and α_p is a p-level structure. The result follows from the fact that the formation of f commutes with arbitrary base change (we may take this base change to be the conjugation by an element of the absolute Galois group of $\mathbb{Q}$). □

Lemma 4.4. *Let r be a prime number different from p, and let ω be as in the statement of Lemma 4.3. Let T_r be the r-th Hecke operator. There exists $\sigma \in G_{\mathbb{Q}}$ such that*

$$a_1(c, T_r\omega) = a_r({}^{\sigma}c, \omega).$$

Proof. By the description of the action of Hecke operators on modular forms of level p given in [Katz 1973, §1.11], we know that $a_1(c, T_r\omega) = a_r(c', \omega)$ for some cusp c' of $X_{G(p)}$ (be aware that the ω we are using here is *not* the ω used in [Katz 1973, §1.11]). In order to actually show that c' must be a conjugate of c, let us start by working on $X(p)$. So, let ω be a modular form of $X(p)$ over $\mathbb{Z}[1/p]$, and let c be a cusp of $X(p)$. The description in [Katz 1973] yields that if c is a cusp of $X(p)$ corresponding to the p-level structure on a Néron p-gon given by

$$(\zeta_p, 0) \mapsto (\alpha, \beta), \quad (1, 1) \mapsto (\gamma, \delta),$$

then $a_1(c, T_r\omega) = a_r(c'\omega)$, where c' is the cusp of $X(p)$ corresponding to the p-level structure

$$(\zeta_p, 0) \mapsto (\alpha, \beta), \quad (1, 1) \mapsto (\gamma', \delta'),$$

where $r\gamma' = \gamma$ and $r\delta' = \delta$ (in $\mathbb{Z}/p\mathbb{Z}$). Thus, using the notation of Lemma 2.1, if c is a cusp of X_H represented by $\left(\binom{\alpha}{\beta}, 1\right)$ (they all are represented by an element of this form because $\det H = \mathbb{F}_p^{\times}$), then c' will be represented by $\left(\binom{\alpha}{\beta}, r^{-1}\right)$. Identifying the set $(H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p$ with the set $H\backslash M_p \times \mathbb{F}_p^{\times}$ as was done in the proof of Corollary 2.2, the cusp c is represented by $\binom{\alpha}{\beta}$, and c' is represented by $\gamma_r\binom{\alpha}{\beta}$, where γ_r is an element of H of determinant r. Corollary 2.3 now yields that c' is in the Galois orbit of c. □

Proof of Proposition 4.1. The proof is standard. As $A_{/k(\lambda)}$ is not trivial, there exists a nonzero $\omega \in \mathrm{Cot}(A_{/k(\lambda)})$. By specialisation results due to Raynaud (as stated, for example, in [Mazur 1978, Corollary 1.1]), the requirement that the characteristic of λ is not 2 allows us to regard $\mathrm{Cot}(A_{/k(\lambda)})$ as a $k(\lambda)$-linear subspace of $\mathrm{Cot}(J_{H/k(\lambda)})$.

Let

$$\sum_{n=1}^{\infty} a_n(c, \omega) q^{n/p} \in k(\lambda)[\![q^{1/p}]\!]$$

be the q-expansion of ω at $c_{/k(\lambda)}$. Lemma 4.2 asserts that $f^*_{c/k(\lambda)}(\omega) = \pm a_1(c, \omega)$. If $a_1(c, \omega) \neq 0$, then $f^*_{c/k(\lambda)}$ is not trivial, and we are done. Suppose now that $a_1(c, \omega) = 0$. If r is a prime different from p, we know that A is stable under the action of T_r. Thus, $T_r\omega \in \mathrm{Cot}(A_{/k(\lambda)})$. By Lemma 4.4, there exists

$\sigma \in G_{\mathbb{Q}}$ such that $a_1(c, T_r\omega) = a_r({}^{\sigma}c, \omega)$. By Lemma 4.3, we then have

$$a_1(c, T_r\omega) = {}^{\sigma}a_r(c, \omega).$$

In particular, $a_1(c, T_r\omega) = 0$ if and only if $a_r(c, \omega) = 0$. If there exists a prime $r \neq p$ such that $a_r(c, \omega) \neq 0$, we see that $f^*_{c/k(\lambda)}(T_r\omega) \neq 0$ and we are done. We are going to show that such a prime always exists if $a_1(c, \omega) = 0$.

Suppose, for the sake of contradiction, that such a prime does not exist. It then follows that $a_n(c, \omega) = 0$ for every integer $n \geq 1$ such that $p \nmid n$. Using the q-expansion principle, we therefore conclude that ω is fixed by a conjugate of the group

$$U(p) := \left\{ \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} \in \mathrm{GL}_2(\mathbb{F}_p) : a \in \mathbb{F}_p \right\} \subseteq \mathrm{GL}_2(\mathbb{F}_p).$$

As ω is also fixed by the action of H, it is fixed under the action of the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ generated by $H(p)$ and the conjugate of $U(p)$ in question. An elementary argument shows that this is the whole of $\mathrm{GL}_2(\mathbb{F}_p)$, and so $\omega = 0$, which is a contradiction. Therefore, there exists a prime $r \neq p$ such that $a_r(c, \omega) \neq 0$, and the proposition follows. □

Given a point $x \in X_H(\mathbb{Q}(\zeta_p)^+)$ and a maximal ideal λ of R, define

$$B_\lambda(x) := \{y \in X_H(\mathbb{Q}) : y \equiv x \pmod{\lambda}\}.$$

In other words, $B_\lambda(x)$ is the set of $\mathbb{Q}$-rational points in the residue class of x modulo λ.

Corollary 4.5. *Let c be a cusp of X_H and let λ be a maximal ideal of R of characteristic different from* 2 (*once again, we note that the characteristic is also not p*). *We then have $B_\lambda(c) = \varnothing$.*

Proof. Say that the characteristic of λ is ℓ. Suppose, for contradiction, that $B_\lambda(c)$ is nonempty, and let $x \in B_\lambda(c)$. Consider $f_c : X_{H/\mathbb{Q}(\zeta_p)} \to A_{/\mathbb{Q}(\zeta_p)}$. The first observation one must make is that $f_c(x)$ is a torsion point in $A(\mathbb{Q}(\zeta_p))$. This is an easy argument that can be found (for the case of a different modular curve) in the paper of Darmon and Merel [1997]. It goes as follows. We first show that a multiple of $\iota_c(x)$ is a $\mathbb{Q}$-rational point in J_H. Indeed, $\iota_c(x) = \mathrm{cl}(x - c)$. Now, given $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$, we find

$${}^{\sigma}\iota_c(x) - \iota_c(x) = \mathrm{cl}(x - {}^{\sigma}c) - \mathrm{cl}(x - c) = \mathrm{cl}(c - {}^{\sigma}c),$$

because x is defined over $\mathbb{Q}$. The Drinfeld–Manin theorem yields the existence of a positive integer m_σ such that $m_\sigma \cdot \mathrm{cl}(c - {}^{\sigma}c) \in J_H(\mathbb{Q})$. Taking $m := \max_\sigma\{m_\sigma\}$, we get

$$m \cdot {}^{\sigma}\iota_c(x) = m \cdot \iota_c(x)$$

for all $\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$. It follows that $m \cdot f_c(x)$ is $\mathbb{Q}$-rational. As $A(\mathbb{Q})$ is finite, $m \cdot f_c(x)$ is torsion, and so is $f_c(x)$ in the first place.

As the characteristic of λ is not 2 nor p, reduction modulo λ gives rise to an injective group homomorphism

$$A(\mathbb{Q}(\zeta_p))_{\mathrm{tors}} \hookrightarrow A_{/k(\lambda)}(k(\lambda)).$$

As $x_{/k(\lambda)} = c_{/k(\lambda)}$ and the image of c in A is 0, knowing that $f_c(x)$ is a torsion point allows us now to conclude that $f_c(x) = f_c(c) = 0$.

To achieve a contradiction, we are now going to make use of Proposition 4.1. Consider the R_λ sections of $X_{G(p)}$ defined by x and c. Let $h_x : \hat{\mathcal{O}}_{X_H, x_{/k(\lambda)}} \to R_\lambda$ be the homomorphism of completed local rings at the special fibres induced by x, and let $h_c : \hat{\mathcal{O}}_{X_H, c_{/k(\lambda)}} \to R_\lambda$ be that induced by c. Note that as $c_{/k(\lambda)} = x_{/\lambda}$, we have $\hat{\mathcal{O}}_{X_H, x_{/k(\lambda)}} = \hat{\mathcal{O}}_{X_H, x_{/k(\lambda)}}$. The statement that $f_c(x) = f_c(c)$ means that

$$h_x \circ \hat{f}_c^{\#} = h_c \circ \hat{f}_c^{\#}.$$

But the statement of Proposition 4.1 is precisely that $\hat{f}_c^{\#}$ is surjective, which leads to $h_x = h_c$. But this is only possible if $x = c$. However, this is a contradiction, as x was assumed to be defined over $\mathbb{Q}$, but the field of definition of c is $\mathbb{Q}(\zeta_p)$. □

5. The image of the residual Galois representation in the nonintegral case

When the prime p, besides not being in the set $\{2, 3, 5, 7, 13\}$, satisfies $p \equiv 2 \pmod 3$, the group $G(p)$ (as defined in Proposition 1.4) is a proper subgroup of $N_{\mathrm{ns}}(p)$ containing $H(p)$, as the following lemma shows.

Lemma 5.1. *If $p \equiv 2 \pmod 3$, the group $H(p)$ is a subgroup of $G(p)$.*

Proof. Explicitly, the group $H(p)$ is

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & \pm a \end{pmatrix} : a \in \mathbb{F}_p^\times \right\} \cup \left\{ \begin{pmatrix} 0 & \epsilon_p b \\ \pm b & 0 \end{pmatrix} : b \in \mathbb{F}_p^\times \right\}.$$

We only have to show that these elements are contained in $G(p)$. As $p \equiv 2 \pmod 3$, the endomorphism of $\mathbb{F}_p^\times$ given by $a \mapsto a^3$ is surjective. It follows from the definition of $G(p)$ that all the elements of the form

$$\begin{pmatrix} a & 0 \\ 0 & \pm a \end{pmatrix}, \quad a \in \mathbb{F}_p^\times,$$

are contained in $G(p)$. Similarly, the function $\mathbb{F}_p^\times \to \mathbb{F}_p^\times$ given by $b \mapsto \epsilon_p^2 b^3$ is surjective, from where it follows that all the elements of the type

$$\begin{pmatrix} 0 & \epsilon_p b \\ \pm b & 0 \end{pmatrix}, \quad b \in \mathbb{F}_p^\times,$$

are in $G(p)$. □

As a consequence, we can take $H = G(p)$ and the results of the previous sections will hold.

The last ingredient we need to be ready to prove Theorem 1.2(b) is the following result [Lemos 2019]. Its usefulness resides in the fact that it implies that, in our situation, our elliptic curve has potentially good reduction at p.

Proposition 5.2 [Lemos 2019, Proposition 2.2]. *Suppose that $p \geq 5$ and the image of $\bar{\rho}_{E,p}$ is contained in the normaliser of a nonsplit Cartan subgroup of* $\mathrm{GL}(E[p])$. *Then E has potentially good reduction at every prime ℓ satisfying $\ell \not\equiv \pm 1 \pmod{p}$.*

Proof of Theorem 1.2(b). Suppose that the image of $\bar{\rho}_{E,p}$ is neither $\mathrm{GL}(E[p])$ nor the normaliser of a nonsplit Cartan subgroup of $\mathrm{GL}(E[p])$. Zywina's result (Proposition 1.4) then asserts that $p \equiv 2 \pmod 3$ and that the image G of $\bar{\rho}_{E,p}$ is conjugate to the group $G(p)$ in $\mathrm{GL}(E[p])$. By choosing an appropriate basis for $\mathrm{GL}(E[p])$, we may in fact assume that the image of $\bar{\rho}_{E,p}$ is contained in $G(p)$.

As we are assuming that $j(E) \notin \mathbb{Z}$, there is some prime ℓ such that $v_\ell(j(E)) < 0$. Proposition 5.2 shows that $\ell \equiv \pm 1 \pmod{p}$. In particular, $\ell \nmid 2p$. Let λ be any prime ideal of $\mathbb{Z}[\zeta_p]$ lying above ℓ. The elliptic curve E gives rise to a $\mathbb{Q}$-rational point x in $X_{G(p)}$. As E has potentially multiplicative reduction at ℓ, it follows that there is a cusp c of $X_{G(p)}$ such that $\tilde{x} = \tilde{c}$, where $\tilde{c}$ and $\tilde{x}$ denote the reductions of c and x modulo λ, respectively. In other words, $x \in B_\lambda(c)$. But this contradicts Corollary 4.5. □

The proof of Theorem 1.2(a) — i.e., the case where E has integral j-invariant — will be the subject of the next section.

6. Runge's method on the curve $X_{G(p)}$ and end of proof of Theorem 1.2

In this section, we deal with the case where $j(E) \in \mathbb{Z}$ and E defines a rational point of the modular curve $X_{G(p)}$, denoted by P. As we only need to treat this case, we assume that $p \equiv 2 \bmod 3$ and $p \notin I(1)$ in all this section. We will prove the following.

Proposition 6.1. *If $P \in X_{G(p)}(\mathbb{Q})$ and $j(P) \in \mathbb{Z}$,*

$$\log |j(P)| \leq 7\sqrt{p}.$$

Before proving this proposition, here are its consequences for the end of the proof of Theorem 1.2.

Corollary 6.2. *For any prime $p \notin I(1)$ congruent to $2 \bmod 3$ and any elliptic curve E over $\mathbb{Q}$ without complex multiplication, if the image of $\bar{\rho}_{E,p}$ is included in a conjugate of $G(p)$ and $j(E) \in \mathbb{Z}$, then $p \leq 1.4 \times 10^7$.*

Proof of Corollary 6.2. Assume that Proposition 6.1 holds. By the explicit surjectivity theorem of [Le Fourn 2016, (7) and Theorem 5.2] (only making use of the fact that the image is contained in the normaliser of a nonsplit Cartan), based on isogeny theorems of Gaudron and Rémond [2014], we also have (if $\log |j(E)| \geq 12 \cdot 985$)

$$p^2 \leq 4 \cdot 10^7 \left(\frac{\log |j(E)|}{12} + 3 + 4 \log(2) \right)^2,$$

which gives

$$\log |j(E)| \geq \frac{6p}{10^{3.5}} - 70.$$

This yields $p \leq 1.4 \cdot 10^7$ when combined with the bound of Proposition 6.1. Finally, if $\log |j(E)| \leq 12 \cdot 985$, we get by the same surjectivity theorem an absolute upper bound on p^2 giving a smaller upper bound on p. □

The proof of Proposition 6.1 relies on Runge's method, which starts with the following fact.

Lemma 6.3. *The set of cusps of $X_{G(p)}$ consists of two Galois orbits. One of these Galois orbits is the set of cusps at infinity, identified via Corollary 2.2 with the orbit*

$$(\mathcal{O}_{\text{cubes}})_{/\pm 1} \subset M_p, \quad \mathcal{O}_{\text{cubes}} := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{F}_p^2 \backslash \{(0,0)\} : a + \sqrt{\epsilon_p} b \text{ is a cube in } \mathbb{F}_{p^2}^\times \right\}.$$

Proof. By Corollary 2.3, we have a correspondence between the Galois orbits of the set of cusps of $X_{G(p)}$ and the set $G(p)\backslash M_p$. Fix a square root $\sqrt{\epsilon_p}$ of ϵ_p and consider the one-to-one map

$$\theta : M_p \to \mathbb{F}_{p^2}^\times / \{\pm 1\}, \quad \begin{pmatrix} a \\ b \end{pmatrix} \mapsto a + \sqrt{\epsilon_p} b.$$

Let $\gamma = \begin{pmatrix} x & \sqrt{\epsilon_p} y \\ y & x \end{pmatrix}$ be an element of $C_{\text{ns}}(p)$. An easy calculation shows that

$$\theta\left(\gamma \begin{pmatrix} a \\ b \end{pmatrix}\right) = (x + \sqrt{\epsilon_p} y)(a + \sqrt{\epsilon_p} b).$$

Moreover, the action of the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ corresponds, under θ to the action of the Frobenius automorphism on $\mathbb{F}_{p^2}^\times$. We thus see that $(\mathcal{O}_{\text{cubes}})_{/\pm 1}$ is a Galois orbit for the action of $G(p)$ on M_p. Indeed, if $\gamma \in G(p) \cap C_{\text{ns}}(p)$, then the action of γ on elements of M_p corresponds, under θ to multiplication by a cube, and so it is clear that it preserves cubes. Also, the action of Frobenius on $\mathbb{F}_{p^2}^\times$ maps cubes to cubes.

The complement of $(\mathcal{O}_{\text{cubes}})_{/\pm 1}$ in M_p is also a Galois orbit because any noncube of $\mathbb{F}_{p^2}^\times$ can be obtained from a specified noncube by multiplying by an appropriate cube and applying the Frobenius automorphism if needed. Finally, by definition of cusps at infinity, we see that $(\mathcal{O}_{\text{cubes}})_{/\pm 1}$ is precisely the set of cusps at infinity of $X_{G(p)}$. □

Proof of Proposition 6.1. As we explain right below, Lemma 6.3 announces that it is possible to apply Runge's method to integral points of $X_{G(p)}$ with respect to the j-invariant, as there are two Galois orbits of cusps. Indeed, one can then construct (following the results of [Kubert and Lang 1981]) a modular unit $U \in \mathbb{Q}(X_{G(p)})$, i.e., a function whose sets of zeros and sets of poles are the two orbits of cusps. We will then prove that $U \in \overline{\mathbb{Z}[j]}$ and $p^3/U \in \overline{\mathbb{Z}[j]}$ after looking at the q-expansions, which directly gives that $U(P)$ is an integer, nonzero and bounded when $j(P) \in \mathbb{Z}$, as $j(P)$ can be large only when P is close from the cusps. As U goes (quickly) to 0 or infinity when P comes close to cusps, studying further the q-expansions will then prove that P is far away from the cusps and allow to bound $|j(P)|$. For more general results on Runge's method on modular curves, the reader is invited to consult [Bilu and Parent 2011a] where the applicability domain (and general bound) is given for every modular curve.

To do it in practice, one needs to define properly a modular unit. We follow the results of [Kubert and Lang 1981] and ideas of Bajolet, Bilu and Matschke [Bajolet et al. 2012] here: even though we do not

exactly use their own results except at the very end, their arguments definitely inspired the construction of our modular unit. We use the following notation:

- $e(x) := e^{2i\pi x}$ for any complex number x.
- $\mathcal{H}$ is the upper half-plane and τ will denote any element of $\mathcal{H}$, for which $q_\tau := e(\tau)$ (we will generally drop the subscript when τ is obvious). For any rational number r, the convention is $q_\tau^r := e(r\tau)$.

For any nonzero pair $\underline{a} := (a_1, a_2)$ in $\mathbb{Q}^2 \cap [0, 1[^2$ (to simplify notation) with common denominator p, we can define a modular function $g_{\underline{a}}$ on $\mathcal{H}$ whose q-expansion is $\mathcal{H}$ is

$$g_{\underline{a}}(\tau) = q^{B_2(a_1)/2} e(a_2(a_1-1)) \prod_{n=0}^{+\infty} \big(1 - q^{n+a_1} e(a_2)\big)\big(1 - q^{n+1-a_1} e(-a_2)\big), \tag{6-1}$$

where $B_2(X) = X^2 - X + \frac{1}{6}$ is the second Bernoulli polynomial. There is a modular transformation formula for these units, but we only need the following fact: for $\mathcal{O}$ a subset of $\mathbb{F}_p^2 \backslash \{(0,0)\}$ (stable by $-\mathrm{Id}$ and the action of $G(p)$), choosing for every $(a,b) \in \mathcal{O}$ their lift $(\widetilde{a}, \widetilde{b}) \in (\mathbb{Z} \cap [0, p[)^2$ and then $(a_1, a_2) := (\widetilde{a}, \widetilde{b})/p$, and for $m \in \mathbb{N}^*$, the product

$$U_{\mathcal{O},m} := \prod_{(a,b)\in\mathcal{O}} g^m_{(\tilde{a}/p, \tilde{b}/p)} \tag{6-2}$$

is automorphic of degree 0 for the congruence subgroup associated to $G(p)$ and defines up to multiplication by a root of unity a function on $\mathbb{Q}(X_{G(p)})$ if

$$m \sum_{(a,b)\in\mathcal{O}} a^2 = m \sum_{(a,b)\in\mathcal{O}} b^2 = m \sum_{(a,b)\in\mathcal{O}} ab = 0 \in \mathbb{Z}/p\mathbb{Z} \tag{6-3}$$

and 6 divides $m|\mathcal{O}|$ [Kubert and Lang 1981, Theorem 5.2 in Chapter 3].

This is what we use to define our key function (with $\mathcal{O} = \mathcal{O}_{\text{cubes}}$), whose properties are summed up below.

Proposition 6.4. *The function on $\mathcal{H}$ defined by*

$$U := \zeta \cdot U_{\mathcal{O}_{\text{cubes}},3} = \zeta \cdot \prod_{\substack{(a,b)\in\mathbb{F}_p^2 \\ a+\sqrt{\epsilon_p} b \text{ cube in } \mathbb{F}_{p^2}^*}} g^3_{(\widetilde{a},\widetilde{b})/p}$$

with ζ the root of unity chosen such that the constant term of the q-expansion of U is 1, induces a rational function on $X(p)$ (also denoted by U) which satisfies the following properties:

- *It belongs to $\mathbb{Q}(X_{G(p)})$.*
- *Its zeroes are the cusps at infinity of $X_{G(p)}$, and its poles make up the second Galois orbit of cusps of $X_{G(p)}$.*
- *It is integral over $\mathbb{Z}[j]$ as well as $p^3 U^{-1}$.*

Proof. First, U does define a function on $\mathbb{Q}(X_{G(p)})$ by [Kubert and Lang 1981, Theorem 5.2 in Chapter 3]. Indeed, $m = 3$ is enough, and for the orbit $\mathcal{O}_{\text{cubes}}$, the vanishing conditions (6-3) hold because the set of cubes of $\mathbb{F}_{p^2}^*$ is stable by multiplication by scalars of $\mathbb{F}_p^*$, so each of the three sums of (6-3) has to be equal to itself times any scalar in $\mathbb{F}_p^*$, hence it is 0.

Regarding the divisors, each of the modular function $g_{\underline{a}}$ is nonvanishing on $\mathcal{H}$ so the divisor of U is supported on the cusps. The analysis of the q-expansion at infinity later proves that its image in $X_{G(p)}$ is a zero of U, so all its Galois conjugates are as U is $\mathbb{Q}$-rational, and the only other Galois orbit (by Lemma 6.3) must be made up with the poles of U.

For integrality, each g_{a_1,a_2} is integral over $\mathbb{Z}[j]$ [Kubert and Lang 1981, Theorem 2.2 of Chapter 2] so U is and it is easily seen from the q-expansions that

$$\prod_{(a,b)\in M_p} g^3_{\widetilde{a}/p,\,\widetilde{b}/p} = \pm p^3,$$

so p^3/U is also integral over $\mathbb{Z}[j]$. □

Consequently, for every $P \in X_{G(p)}(\mathbb{Q})$ with $j(P) \in \mathbb{Z}$,

$$U(P) \in \mathbb{Z} \quad \text{and} \quad 0 \le \log|U(P)| \le 3\log(p), \tag{6-4}$$

which is the whole point of considering this modular unit. We now use the expansion at infinity to bound q_τ.

For every $\tau \in \mathcal{H}$, gathering the q-expansions (6-1),

$$\log|U(\tau)| = \operatorname{Ord}_q(U)\log|q| + \log|\rho_U| + \log|R(\tau)|$$

with

$$\operatorname{Ord}_q(U) = 3\sum_{(a,b)\in\mathcal{O}_{\text{cubes}}} B_2(\widetilde{a}/p)/2, \quad \rho_U = \prod_{(a,b)\in\mathcal{O}_{\text{cubes}}} \rho^3_{\widetilde{a}/p,\,\widetilde{b}/p}$$

and

$$\rho_{(a_1,a_2)} = \begin{cases} -e((a_1-1)a_2/2) & \text{if } a_1 \neq 0, \\ -2i\sin(\pi a_2/2) & \text{if } a_1 = 0. \end{cases}$$

Finally,

$$\log|R(\tau)| = 3\sum_{(a,b)\in\mathcal{O}_{\text{cubes}}} \log|R_{a_1,a_2}(\tau)|$$

$$\text{where} \quad \log|R_{a_1,a_2}(\tau)| = \sum_{n\ge 0} \log\left|1 - q^{n+a_1}e(a_2)\right| + \log\left|1 - q^{n+a_1}e(a_2)\right|.$$

We will obtain the following estimates and equalities.

Proposition 6.5. *We have*

$$\operatorname{Ord}_q(U) = \frac{p^2-1}{4p}, \quad |\rho_U| = (p-1)^3,$$

and

$$|\log R(\tau)| \le 2(p^2-1)\frac{|q|}{1-|q|} + \frac{\pi^2 p(p-2)}{3|\log|q||}.$$

Proof. First, for $a = 0$, all nonzero b satisfy that $(a, b) \in \mathcal{O}_{\text{cubes}}$ because ϵ_p is a cube in $\mathbb{F}_{p^2}^*$ (check its order). This gives $(p-1)$ elements in the orbit. Moreover, $\mathcal{O}_{\text{cubes}}$ is stable by scalar multiplication by $\mathbb{F}_p^*$, hence all fibres of $(a, b) \mapsto a$ have the same cardinality except above 0. They are thus of cardinality $(p-2)/3$. This allows us to compute

$$\operatorname{Ord}_q(U) = \tfrac{3}{12}(p-1) + \frac{(p-2)}{2}\sum_{a=1}^{p-1}\big((a/p)^2 - (a/p) + \tfrac{1}{6}\big) = \frac{p^2-1}{4p}.$$

Similarly, for ρ, as all terms except for $a_1 = 0$ have modulus 1,

$$|\rho_U| = \prod_{b=1}^{p-1} |1 - e(b/p)|^3 = (p-1)^3.$$

Finally, for $R(\tau)$, we use that $\big|\log|1-z|\big| \le -\log\big|1-|z|\big|$ for $|z| \le 1$ for $n = 0$, and $\big|\log|1-z|\big| \le |z|/(1-|z|)$ for the other terms (if $a_1 = 0$, the first $n = 0$ term is put into $\rho_{(0,a_2)}$. We thus get for $a_1 \neq 0$

$$\big|\log|R_{a_1,a_2}(\tau)|\big| \le \big|\log(1-|q|^{a_1})\big| + \big|\log(1-|q|^{1-a_1})\big| + \frac{2|q|}{1-|q|},$$

$$\text{and} \quad \big|\log|R_{0,a_2}(\tau)|\big| \le \frac{2|q|}{1-|q|}.$$

Gathering the previous inequalities for the product expansion,

$$\begin{aligned}
|\log R(\tau)| &\le 2(p^2-1)\frac{|q|}{1-|q|} + 2(p-2)\sum_{a=1}^{p-1}|\log(1-x^a)|, \quad x = |q|^{1/p} \\
&\le 2(p^2-1)\frac{|q|}{1-|q|} + \frac{\pi^2(p-2)}{3|\log(x)|} \\
&\le 2(p^2-1)\frac{|q|}{1-|q|} + \frac{\pi^2 p(p-2)}{3\big|\log|q|\big|}
\end{aligned}$$

by[Bilu and Parent 2011b, Lemma 3.5]. □

Now, assume $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ is such that its reduction modulo p is of the shape

$$\begin{pmatrix} a & \epsilon_p b \\ b & a \end{pmatrix},$$

where $a + \epsilon_p b$ is *not* a cube in $\mathbb{F}_{p^2}^*$. The composition $U \circ \gamma$ is a modular unit on $X_{G(p)}$ (not necessarily defined over $\mathbb{Q}$ anymore), but by arguments similar to the previous ones, we have the following:

$$\log|U(\gamma\tau)| = \operatorname{Ord}_\gamma U \cdot \log|q_\tau| + \log|\rho_{U,\gamma}| + \log|R_\gamma(\tau)|,$$

where

$$\operatorname{Ord}_\gamma U = -\frac{p^2-1}{8p}, \quad \log|\rho_{U,\gamma}| = 0, \quad \text{and} \quad |\log R_\gamma(\tau)| \le \frac{\pi^2 p(p+1)}{3\big|\log|q|\big|}.$$

The argument behind each of those computations is that by our hypothesis on γ, the function $(a, b) \mapsto a_1((a, b)\cdot\gamma))$ on $\mathcal{O}_{\text{cubes}}$ does not have 0 in its image, and each other element of $\mathbb{F}_p^*$ has $(p+1)/3$ elements in its fibre (again by stability by multiplication by $\mathbb{F}_p^*$).

Putting this together, we obtain

$$\left|\log|U(\tau)| - \frac{p^2-1}{4p}\log|q| - 3\log(p-1)\right| \le 2(p^2-1)\frac{|q|}{1-|q|} + \frac{\pi^2 p(p-2)}{3\left|\log|q|\right|}$$

and for the choice of γ above,

$$\left|\log|U(\gamma\tau)| + \frac{p^2-1}{8p}\log|q|\right| \le \frac{\pi^2 p(p+1)}{3\left|\log|q_\tau|\right|}.$$

Now, let us assume that there is a noncuspidal point $P \in X_{G(p)}(\mathbb{Q})$ with $j(P) \in \mathbb{Z}$. There is a lift $\tau \in \mathcal{H}$ such that $|q_\tau|$ is small and a $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma\cdot\tau$ is above P in the complex uniformisation of $X_{G(p)}$. This means that P is close to the cusp $\gamma^{-1}(\infty)$. Up to Galois conjugation (which fixes P but changes the cusps), we can reduce to two situations: either $\gamma = \mathrm{Id}$ (which means that τ belongs to the usual fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$), or γ is chosen as above such that its reduction modulo p corresponds to a matrix of $C_{\text{ns}}(p)$ not in $G(p)$. In these two cases, we respectively have $U(\tau) = U(P)$ and $U(\gamma\tau) = U(P)$, and this is where we use (6-4) to bound the corresponding term in one of the two previous inequalities. The first case gives

$$\frac{p^2-1}{4p}\left|\log|q|\right| \le 3\log(p-1) + 2(p^2-1)\frac{|q|}{1-|q|} + \frac{\pi^2 p(p-2)}{3\left|\log|q|\right|}.$$

Assuming $p \ge 100$ and $\left|\log|q|\right| \ge \sqrt{p}$, we can bound roughly the coefficients and the nondominant terms to obtain

$$\left|\log|q|\right| \le 1.2 + \frac{13p}{\left|\log|q|\right|}.$$

Proceeding similarly in the second case (with the same assumptions on p and $|q|$), we obtain

$$\left|\log|q|\right| \le 1.2 + \frac{27p}{\left|\log|q|\right|}.$$

Both cases give rise to second-degree polynomial inequalities which we can readily solve, and using then the estimates of [Bilu and Parent 2011a, Corollary 2.2], after simplification,

$$\log|j(P)| \le 7\sqrt{p}.$$

We can retrieve the remaining cases $p < 100$ by refining the estimates above (or by using the main theorem of [Bajolet et al. 2012]), and the case $\log|q| \le \sqrt{p}$ by [Bilu and Parent 2011a, Corollary 2.2] again, which concludes the proof. □

Appendix A: The proof of Lemma 3.5

Using the identification of Corollary 2.2, the action of the map d_1 on the cusps corresponds to the canonical projection

$$C_{\mathrm{sp}}(p)\backslash M_p \times \mathbb{F}_p^\times \to T(p)\backslash M_p \times \mathbb{F}_p^\times,$$

where $T(p)$ is the upper triangular subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Using (3-2), we see that $d_p = w_p \circ d_1 \circ \omega_p$. One easily checks that the action of ω_p on the cusps of $X_{\mathrm{sp}}(p)$ is given by

$$\omega_p : \left(\binom{a}{b}, d\right) \mapsto \left(\binom{b}{a}, -d\right).$$

Let c be the cusp at infinity of $X_{\mathrm{sp}}^+(p)$, and let c' be the cusp at infinity of $X_{\mathrm{sp}}(p)$ (which obviously lies over c). We will show that $\eta'(c') = 0$, from where it will immediately follow that $\eta(c) = 0$.

The image of c' under d_1 is the cusp at infinity of $X_0(p)$. To compute $d_p(c')$, we start by calculating $\omega_p(c')$. Being a cusp at infinity, c' is represented by an element of $M_p \times \mathbb{F}_p^\times$ of the form

$$\left(\binom{a}{0}, a\right) \quad a \in \mathbb{F}_p^\times.$$

The involution ω_p then maps c' to the cusp represented by $\left(\binom{0}{a}, -a\right)$. The image of this cusp under d_1 is the cusp 0 of $X_0(p)$. As w_p swaps the cusp 0 with ∞, we conclude that $d_p(c')$ is also the cusp at infinity of $X_0(p)$. It follows from the definition of η' that $\eta'(c') = 0$, as we wanted.

Now let c be a cusp of $X_{\mathrm{sp}}^+(p)$ not at infinity. Let c' be a cusp of $X_{\mathrm{sp}}(p)$ lying over it (in this situation, c' is necessarily not at infinity). We will now show that $\eta'(c') = \mathrm{cl}(0 - \infty)$, which will conclude the proof of the lemma.

Let

$$\left(\binom{a}{b}, d\right), \quad a, b, d \in \mathbb{F}_p$$

be an element of $M_p \times \mathbb{F}_p^\times$ representing the cusp c'. Note that $a, b \in \mathbb{F}_p^\times$ because otherwise c' would lie over the cusp at infinity of $X_{\mathrm{sp}}^+(p)$. As $b \in \mathbb{F}_p^\times$, the image under d_1 of c' is the cusp 0. The cusp $\omega_p(c')$ is represented by $\left(\binom{b}{a}, -d\right)$. As $a \in \mathbb{F}_p^\times$, the image of $\omega_p(c')$ under d_1 is the cusp 0. Using again the fact that w_p swaps the cusps of $X_0(p)$, we conclude that $d_p(c') = \infty$. Therefore, $\eta'(c') = \mathrm{cl}(0 - \infty)$, as we wanted. □

Appendix B: The proof of Proposition 1.4

Following a suggestion made by an anonymous referee, we add here, as an appendix, Zywina's proof of Proposition 1.4 (which is Proposition 1.13 in his paper [Zywina 2015]). We must emphasise that none of the results and ideas in this appendix are due to the authors of this paper, and that the original version of this proof can be found in [Zywina 2015]. The main reason for the existence of this appendix is the fact that Zywina's paper remains unpublished.

Fix a decomposition subgroup D_p of $G_{\mathbb{Q}}$ over p, and let I be the corresponding inertia subgroup.

Proof of Proposition 1.4. As p does not lie in $I(1)$, we know (Theorem 1.1) that the image of $\bar{\rho}_{E,p}$ is contained in the normaliser of a nonsplit Cartan subgroup. By choosing a basis of $E[p]$ appropriately, we may assume that this normaliser of nonsplit Cartan is $N_{\mathrm{ns}}(p)$. Let $j(E)$ be the j-invariant of E. We start by showing that we must have $v_p(j(E)) \geq 0$.

Suppose that $v_p(j(E)) < 0$. Our elliptic curve $E_{/\mathbb{Q}_p}$ is either isomorphic to a Tate curve over $\mathbb{Q}_p$, or is a quadratic twist of one. Therefore, there exists a character $\psi : D_p \to \mathbb{F}_p^{\times}$, trivial or quadratic, such that

$$\bar{\rho}_{E,p}|_{D_p} \sim \begin{pmatrix} \psi\chi_p & * \\ 0 & \psi \end{pmatrix}.$$

As $\chi_p : D_p \to \mathbb{F}_p^{\times}$ is surjective, it follows that the image of $\bar{\rho}_{E,p}(D_p)$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ has order divisible by $p-1$. Note however that the image of $N_{\mathrm{ns}}(p)$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ has order $2(p+1)$. If the image of $\bar{\rho}_{E,p}$ were contained in $N_{\mathrm{ns}}(p)$, then $p-1$ would divide $2(p+1)$, which is not possible because $p \geq 19$. This leads us to conclude that if the image of $\bar{\rho}_{E,p}$ is contained in the normaliser of a nonsplit Cartan subgroup, then $v_p(j(E)) \geq 0$, as we wanted.

Now that we know that $v_p(j(E)) \geq 0$, we will show that $\bar{\rho}_{E,p}(I)$ is a subgroup of index 1 or 3 of $C_{\mathrm{ns}}(p)$. Before proving this, we point out that $\bar{\rho}_{E,p}(I)$ is cyclic. Indeed, the representation $\bar{\rho}_{E,p}|_I$ factors through the tame inertia subgroup of I because the order of $N_{\mathrm{ns}}(p)$ is not divisible by p. The cyclicity of $\bar{\rho}_{E,p}(I)$ now follows from the fact that the tame inertia subgroup is pro-cyclic.

There is a finite extension K of $\mathbb{Q}_p$ of ramification degree $e \in \{1, 2, 3, 4, 6\}$ over which E acquires good reduction (see Section 5.6 of [Serre 1972]). We will denote by v the valuation of K normalised so that $v(p) = e$. Let I_K denote the inertia subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/K)$. Let

$$[p](X) = \sum_{i=1}^{\infty} a_i X^i, \quad a_i \in \mathbb{Z}_p$$

be the multiplication by p in the formal group of E. As every a_i lies in $\mathbb{Z}_p$, the integers $v(a_i)$ are nonnegative multiples of e. As either $v(a_p) = 0$ (the ordinary case), or $v(a_p) \neq 0$ and $v(a_{p^2}) = 0$ (the supersingular case), the Newton polygon of $[p](X)$ can then either start with a line segment connecting $(1, e)$ to $(p, 0)$, or with a line segment connecting $(1, e)$ to $(p^2, 0)$. Using [Serre 1972, Proposition 10, Section 1.10] and the fact that the representation

$$\bar{\rho}_{E,p}|_{I_K} : I_K \to \mathrm{GL}_2(\mathbb{F}_p)$$

is semisimple (because the order of $N_{\mathrm{ns}}(p)$ is coprime to p), we conclude that in the ordinary case we have [Serre 1972, Proposition 11, Section 1.11]

$$\bar{\rho}_{E,p}|_{I_K} \sim \begin{pmatrix} \chi_p & 0 \\ 0 & 1 \end{pmatrix}, \tag{B-1}$$

while in the supersingular case we have

$$\bar{\rho}_{E,p}|_{I_K} \otimes_{\mathbb{F}_p} \overline{\mathbb{F}}_p \sim \begin{pmatrix} \theta_2^e & 0 \\ 0 & \theta_2^{pe} \end{pmatrix}, \tag{B-2}$$

where θ_2 is a fundamental character of level 2. We show that (B-1) cannot occur.

If $\bar{\rho}_{E,p}|_{I_K}$ were as in (B-1), then the image of $\bar{\rho}_{E,p}(I_K)$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ would be a cyclic group of order $p-1$. Since the square of any element in $N_{\mathrm{ns}}(p) - C_{\mathrm{ns}}(p)$ is a scalar matrix, the order of every element in the image of $N_{\mathrm{ns}}(p)$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ divides $p+1$. In particular, we would have $p-1 \mid p+1$. However, this is not possible, as $p \geq 19$.

We are thus in situation (B-2). The group $\bar{\rho}_{E,p}(I_K)$ is therefore a cyclic group of order

$$\frac{p^2-1}{\gcd(p^2-1, e)},$$

and so the order of $\bar{\rho}_{E,p}(I)$ is a multiple of this number. Also, it follows that $\bar{\rho}_{E,p}(I)$ is contained in $C_{\mathrm{ns}}(p)$. Indeed, a generator of $\bar{\rho}_{E,p}(I)$ must have order $(p^2-1)/\gcd(p^2-1, e) \geq (p^2-1)/6$, but every element of $N_{\mathrm{ns}}(p) - C_{\mathrm{ns}}(p)$ has order dividing $2(p-1)$. As $p \geq 19$, it follows that a generator of $\bar{\rho}_{E,p}(I)$ must be an element of $C_{\mathrm{ns}}(p)$, and so $\bar{\rho}_{E,p}(I)$ is contained in $C_{\mathrm{ns}}(p)$.

As the order of $C_{\mathrm{ns}}(p)$ is p^2-1, we can therefore conclude that $\bar{\rho}_{E,p}(I)$ is a subgroup of index 1, 2, 3, 4 or 6 of $C_{\mathrm{ns}}(p)$. Note however that if this index were even, then $\bar{\rho}_{E,p}(I)$ would be contained in the subgroup of squares of $C_{\mathrm{ns}}(p)$ (as $C_{\mathrm{ns}}(p)$ is cyclic, this is the only subgroup of index 2). However, this would contradict the fact that the determinant of $\bar{\rho}_{E,p}|_I$ surjects to $\mathbb{F}_p^\times$. Therefore, the index of $\bar{\rho}_{E,p}(I)$ in $C_{\mathrm{ns}}(p)$ is odd, and so it is 1 or 3.

Now let H denote the group $\bar{\rho}_{E,p}(G_{\mathbb{Q}}) \cap C_{\mathrm{ns}}(p)$. By what we have seen, H has index 1 or 3 in $C_{\mathrm{ns}}(p)$ (it contains $\bar{\rho}_{E,p}(I)$). Note that if $H = C_{\mathrm{ns}}(p)$, then $\bar{\rho}_{E,p}(G_{\mathbb{Q}}) = N_{\mathrm{ns}}(p)$, as the image of $\bar{\rho}_{E,p}$ cannot be contained in $C_{\mathrm{ns}}(p)$ (recall that this is due to the fact that the image of complex conjugation must have trace 0 and determinant -1, and there is no element in $C_{\mathrm{ns}}(p)$ simultaneously satisfying both of these properties). It remains to treat the case where H has index 3 in $C_{\mathrm{ns}}(p)$.

Suppose that H has index 3 in $C_{\mathrm{ns}}(p)$. As $C_{\mathrm{ns}}(p)$ is cyclic, there is only one subgroup of index 3: the subgroup of cubes. As H contains $\bar{\rho}_{E,p}(I)$, the $\det(H) = \mathbb{F}_p^\times$. Note that if $p \equiv 1 \pmod 3$, then it is not possible to have H of index 3 in $C_{\mathrm{ns}}(p)$, because then $\det(H) \neq \mathbb{F}_p^\times$. Thus, if $p \equiv 1 \pmod 3$, the image of $\bar{\rho}_{E,p}$ must be $N_{\mathrm{ns}}(p)$.

Suppose then that $p \equiv 2 \pmod 3$ and that H has index 3 in $C_{\mathrm{ns}}(p)$. It is easy to check that $N_{\mathrm{ns}}(p)/H$ is isomorphic to D_3, the dihedral group of size 6. The image of $\bar{\rho}_{E,p}(G_{\mathbb{Q}})$ in $N_{\mathrm{ns}}(p)/H$ has index 3. Now, all the index 3 subgroups of D_3 are conjugate, from where it follows that the image of $\bar{\rho}_{E,p}$ is a conjugate of $G(p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$. □

Acknowledgements

This project started when Lemos was still a postdoctoral fellow at the Max Planck Institute for Mathematics (MPIM). For this reason, he would like to thank the MPIM for their financial support and for the excellent

work environment provided. The authors are also grateful to the anonymous referee, who provided thorough feedback on a previous version of this paper and pointed out a generalisation of an earlier result (this is now Proposition 3.1), helping to improve the quality of the paper.

References

[Bajolet et al. 2012] A. Bajolet, Y. Bilu, and B. Matschke, "Computing integral points on $X_{\mathrm{ns}}^{+}(p)$", preprint, 2012. To appear in *Algebra Number Theory*. arXiv

[Bilu and Parent 2011a] Y. Bilu and P. Parent, "Runge's method and modular curves", *Int. Math. Res. Not.* **2011**:9 (2011), 1997–2027. MR Zbl

[Bilu and Parent 2011b] Y. Bilu and P. Parent, "Serre's uniformity problem in the split Cartan case", *Ann. of Math.* (2) **173**:1 (2011), 569–584. MR Zbl

[Bilu et al. 2013] Y. Bilu, P. Parent, and M. Rebolledo, "Rational points on $X_0^+(p^r)$", *Ann. Inst. Fourier* (*Grenoble*) **63**:3 (2013), 957–984. MR Zbl

[Darmon and Merel 1997] H. Darmon and L. Merel, "Winding quotients and some variants of Fermat's last theorem", *J. Reine Angew. Math.* **490** (1997), 81–100. MR Zbl

[Deligne and Rapoport 1973] P. Deligne and M. Rapoport, "Les schémas de modules de courbes elliptiques", pp. 143–316 in *Modular functions of one variable, II* (Antwerp, 1972), edited by P. Deligne and W. Kuyk, Lecture Notes in Math. **349**, Springer, 1973. MR Zbl

[Gaudron and Rémond 2014] É. Gaudron and G. Rémond, "Théorème des périodes et degrés minimaux d'isogénies", *Comment. Math. Helv.* **89**:2 (2014), 343–403. MR Zbl

[Katz 1973] N. M. Katz, "p-adic properties of modular schemes and modular forms", pp. 69–190 in *Modular functions of one variable, III* (Antwerp, 1972), edited by W. Kuyk and J.-P. Serre, Lecture Notes in Math. **350**, Springer, 1973. MR Zbl

[Kubert and Lang 1981] D. S. Kubert and S. Lang, *Modular units*, Grundlehren der Math. Wissenschaften **244**, Springer, 1981. MR Zbl

[Le Fourn 2016] S. Le Fourn, "Surjectivity of Galois representations associated with quadratic $\mathbb{Q}$-curves", *Math. Ann.* **365**:1-2 (2016), 173–214. MR Zbl

[Lemos 2019] P. Lemos, "Serre's uniformity conjecture for elliptic curves with rational cyclic isogenies", *Trans. Amer. Math. Soc.* **371**:1 (2019), 137–146. MR Zbl

[Mazur 1977] B. Mazur, "Modular curves and the Eisenstein ideal", *Inst. Hautes Études Sci. Publ. Math.* **47** (1977), 33–186. MR Zbl

[Mazur 1978] B. Mazur, "Rational isogenies of prime degree", *Invent. Math.* **44**:2 (1978), 129–162. MR Zbl

[Momose 1984] F. Momose, "Rational points on the modular curves $X_{\mathrm{split}}(p)$", *Compos. Math.* **52**:1 (1984), 115–137. MR Zbl

[Najman 2018] F. Najman, "Isogenies of non-CM elliptic curves with rational j-invariants over number fields", *Math. Proc. Cambridge Philos. Soc.* **164**:1 (2018), 179–184. MR Zbl

[Serre 1972] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl

[Serre 1981] J.-P. Serre, "Quelques applications du théorème de densité de Chebotarev", *Inst. Hautes Études Sci. Publ. Math.* **54** (1981), 323–401. MR Zbl

[Zywina 2015] D. Zywina, "On the possible images of the mod ℓ representations associated to elliptic curves over $\mathbb{Q}$", preprint, 2015. arXiv

Communicated by Bjorn Poonen
Received 2020-02-05 Revised 2020-08-19 Accepted 2020-10-10

samuel.le-fourn@univ-grenoble-alpes.fr *Institut Fourier, Université Grenoble Alpes, Saint-Martin d'Hères, France*

lemos.pj@gmail.com *Mathematics Department, University College London, United Kingdom*