RESIDUAL GALOIS REPRESENTATIONS OF ELLIPTIC CURVES WITH IMAGE CONTAINED IN THE NORMALISER OF A NON-SPLIT CARTAN

Samuel Le Fourn, Pedro Lemos

RESIDUAL GALOIS REPRESENTATIONS OF ELLIPTIC CURVES WITH IMAGE CONTAINED IN THE NORMALISER OF A NON-SPLIT CARTAN

SAMUEL LE FOURN AND PEDRO LEMOS

ABSTRACT. 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

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

of E is either the whole of $\mathrm{GL}_2(\mathbb{F}_p)$, or is *contained* in the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. In this paper, we build on work of Zywina and show that when $p > 1.4 \times 10^7$, the image of $\bar{\rho}_{E,p}$ is either $\mathrm{GL}_2(\mathbb{F}_p)$, or *is* the normaliser of a non-split 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 k be a number field and p be a prime. Given an elliptic curve E defined over k, the subgroup $E[p]$ consisting of the p-torsion points of $E(\overline{k})$ has the structure of a 2-dimensional $G_k[\mathbb{F}_p]$-module, where G_k is the absolute Galois group $\mathrm{Gal}(\overline{k}/k)$ of k. The study of this Galois module has been a very active field of research in modern number theory. In this paper, we will focus on the image of the associated Galois representation

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

As we will only be interested in the image of the representation up to conjugacy, we will often assume that a basis of $E[p]$ has been chosen and identify $\mathrm{GL}(E[p])$ with $\mathrm{GL}_2(\mathbb{F}_p)$.

In the case where E does not have complex multiplication, Serre [14] showed that, for p large enough, the image of $\bar{\rho}_{E,p}$ is the whole of $\mathrm{GL}_2(\mathbb{F}_p)$. However, in principle, how large p must be depends on the elliptic curve E and the number field k. In his paper [14], Serre is asks whether it is possible to prove that this is actually only dependent on the number field. The state of the art in what concerns this question can be summarised in the following result, due to the work of several mathematicians, amongst whom we highlight Bilu, Mazur, Parent, Rebolledo and Serre (we refer the reader to the next section for the terminology).

Theorem 1.1 ([11, 2, 3, 15])**.** *Let E be an elliptic curve defined over $\mathbb{Q}$ and without complex multiplication. Let p be a prime not in the set $\{2, 3, 5, 7, 11, 13, 17, 37\}$. If the image of $\bar{\rho}_{E,p}$*

is not $\mathrm{GL}_2(\mathbb{F}_p)$*, then it is* contained *in the normaliser of a non-split Cartan subgroup of* $\mathrm{GL}_2(\mathbb{F}_p)$.

In this paper, we propose to improve Theorem 1.1. The main result of this paper is the following.

Theorem 1.2. *Let E be an elliptic curve defined over $\mathbb{Q}$ and without complex multiplication. Let $p > 1.4 \times 10^7$ be a prime. If the image of $\bar{\rho}_{E,p}$ is not $\mathrm{GL}_2(\mathbb{F}_p)$, then it* is *the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$.*

Theorem 1.1 differs from Theorem 1.2 in that the former asserts that when the image of $\bar{\rho}_{E,p}$ is not the whole of $\mathrm{GL}_2(\mathbb{F}_p)$, then it is merely *contained* in the normaliser of a non-split Cartan subgroup, while the latter says that, for p large enough, the image must actually be the whole normaliser of non-split Cartan when the representation is not surjective.

In the case of elliptic curves with non-integral j-invariant, we are able to loosen the restrictions on p.

Theorem 1.3. *Let E be an elliptic curve defined over $\mathbb{Q}$ with non-integral j-invariant. Let p be a prime not contained in the set $\{2, 3, 5, 7, 11, 13, 17, 37\}$. If the image of $\bar{\rho}_{E,p}$ is not $\mathrm{GL}_2(\mathbb{F}_p)$, then it* is *the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$.*

Remark 1.4. *In fact, the end of the proof of Theorem 1.2 (section 8), dealing with integral j-invariants, implies that if an elliptic curve E without CM admits a prime $p \notin \{2, 3, 5, 7, 11, 13, 17, 37\}$ such that the image of $\bar{\rho}_{E,p}$ is neither $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ nor the normaliser of a nonsplit Cartan subgroup, $j(E) \in \mathbb{Z}$ and $\log|j(E)| \leq \max(12000, 7\sqrt{p}) \leq 27000$ by Theorem 1.2. In particular, there are only finitely many such elliptic curves up to isomorphism. One would then hope to treat these cases algorithmically, but the authors admit they could not find a reasonably efficient way to do so.*

We describe an immediate application of this result. Let $d \geq 1$ be a positive integer. Najman [13] 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 $\phi : E \to E'$ of degree p defined over a number field K of degree $\leq d$. For instance, a celebrated result of Mazur [11] shows that

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

Najman [13] shows that

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

Assuming that Serre's uniformity question admits a positive answer—i.e., that $\bar{\rho}_{E,p}$ is surjective whenever $E/\mathbb{Q}$ does not have complex multiplication and p is a prime not in $I(1)$—, he is able to prove that

$$I(d) = I(1) \cup \{p \text{ prime} : p \leq d - 1\}. \tag{1.1}$$

Using Najman's arguments, we are able to show that Theorem 1.2 yields the following result.

Theorem 1.5. *For $d \geq 1.4 \times 10^7$, we have $I(d) = \{p \text{ prime} : p \leq d-1\}$.*

Remark 1.6. *Note that $I(1) \subseteq \{p \text{ prime} : p \leq d-1\}$ for $d \geq 1.4 \times 10^7$.*

Proof assuming Theorem 1.2. We follow Najman's arguments given in [13]. Let p be a prime in $I(d)$ and suppose that it is not in $I(1)$. We want to show that $p \leq d-1$. If $p < 1.4 \times 10^7$, then we have $p \leq d-1$. Assume then that $p > 1.4 \times 10^7$. Let $E/\mathbb{Q}$ be an elliptic curve without complex multiplication for which there is an isogeny $\phi : E_{/K} \to E'$ of degree p for some number field K of degree $\leq d$. Theorem 1.2 asserts that the image of $\bar{\rho}_{E,p}$ is either $\mathrm{GL}_2(\mathbb{F}_p)$ or is the normaliser of a non-split Cartan subgroup. Let us first assume that the image is $\mathrm{GL}_2(\mathbb{F}_p)$. Let C denote $\ker \phi$, which is a finite étale group scheme over K of order p. Let L denote the minimal number field over which C is defined. In other words, let

$$H := \{\sigma \in G_{\mathbb{Q}} : \sigma(C) = C\},$$

and define $L := \bar{\mathbb{Q}}^H$. Of course, $L \subseteq K$. Now, given that the image of $\bar{\rho}_{E,p}$ is the whole of $\mathrm{GL}_2(\mathbb{F}_p)$, then $\bar{\rho}_{E,p}$ establishes an isomorphism between $\mathrm{Gal}(\mathbb{Q}(E[p])/L)$ and a Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Therefore, $[\mathbb{Q}(E[p]) : L] = p(p-1)^2$, from where it follows that $[L : \mathbb{Q}] = p+1$. Thus, $p+1 \leq [K : \mathbb{Q}] \leq d$, and so $p \leq d-1$, as we wanted.

The same argument will work in the case when the image of $\bar{\rho}_{E,p}$ is the normaliser of a non-split Cartan N. Indeed, in this case, $\bar{\rho}_{E,p}$ will establish an isomorphism between $\mathrm{Gal}(\mathbb{Q}(E[p])/L)$ and the intersection of N with a Borel subgroup. The size of this intersection is either $p-1$ or $2(p-1)$. As the size of N is $2(p^2-1)$, it follows that $[L : \mathbb{Q}]$ is either $p+1$ or $2(p+1)$. In any case, $p+1 \leq [L : \mathbb{Q}] \leq [K : \mathbb{Q}] \leq d$, leading once again to $p \leq d-1$. So we have

$$I(d) \subseteq I(1) \cup \{p \text{ prime} : p \leq d-1\},$$

as we wanted.

If $p \in I(1)$, then, by definition, there are elliptic curves over $\mathbb{Q}$ without complex multiplication admitting a $\mathbb{Q}$-isogeny of degree p. Let p be a prime such that $p \leq d-1$ and $p \notin I(1)$. Take any elliptic curve $E/\mathbb{Q}$ such that $\bar{\rho}_{E,p}$ surjects onto $\mathrm{GL}_2(\mathbb{F}_p)$. Consider a Borel subgroup B of $\mathrm{GL}_2(\mathbb{F}_p)$. The fixed points of B constitute a subgroup of $E(\bar{\mathbb{Q}})$ of order p defined over a number field of degree $p+1 \leq d$. This finishes the proof of the theorem. □

1.1. **Plan of the proofs of Theorems 1.2 and 1.3.** In what follows, we will adopt the notation of [13] and write $I(1)$ for the set $\{2, 3, 5, 7, 11, 13, 17, 37\}$. The proofs of Theorems 1.2 and 1.3 build on the following result of Zywina [16, Proposition 1.13].

Proposition 1.7 (Zywina)**.** *Let $E/\mathbb{Q}$ be an elliptic curve without complex multiplication and let $p \notin I(1)$ be a prime. Suppose 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 non-split 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 non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, or is conjugate to the group

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

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

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 $G(p)$. In the case where the j-invariant $E/\mathbb{Q}$ is not integral, we will rule out this possibility by using Mazur's formal immersion argument (see [11]). 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 the point of Proposition 7.1). It then follows that we have 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 non-trivial quotient of the jacobian of $X_{G(p)}$ with finite Mordell–Weil group and use the standard formal immersion arguments due to Mazur to prove that such a point cannot exist. This will give us Theorem 1.3. Notice that the existence of such a quotient is unexpected at first because this is precisely what does not work with the modular curve as associated to the normaliser of $C_{\mathrm{ns}}(p)$.

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 treat this case by applying Runge's method to obtain an upper bound for $\log |j(E)|$ which is linear in $\sqrt{p}$. On the other hand, an explicit version of Serre's surjectivity theorem provides a lower bound linear in p, which gives rise to a contradiction for $p \geq 1.4 \times 10^7$. We thus deduce Theorem 1.2.

2. Subgroups of $\mathrm{GL}_2(\mathbb{F}_p)$

The sole point of this section is to set down some notation for some explicit maximal subgroups of $\mathrm{GL}_2(\mathbb{F}_p)$. For that, let p be an odd prime. Throughout these notes, we will write $C_{\mathrm{sp}}(p)$ for the split Cartan subgroup given by diagonal matrices; i.e.,

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

Its normaliser will be denoted by $N_{\mathrm{sp}}(p)$. This is the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ consisting of diagonal and anti-diagonal matrices. The conjugates of $C_{\mathrm{sp}}(p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$ are known as the *split Cartan subgroups* of $\mathrm{GL}_2(\mathbb{F}_p)$.

For each prime p, fix a generator ϵ_p of the cyclic group $\mathbb{F}_p^\times$. We define

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

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

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

The conjugates of $C_{ns}(p)$ in $\mathrm{GL}_2(\mathbb{F}_p)$ are known as the *non-split Cartan subgroups* of $\mathrm{GL}_2(\mathbb{F}_p)$.

Finally, we define

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

which is a subgroup of $N_{\mathrm{ns}}(p)$ of index 3.

This notation will be used throughout the whole text.

3. Modular curves

We give a brief review of some basic facts concerning modular curves. The reader should be warned that our definitions of *cusps at infinity* and *at zero* differ slightly from the standard ones. The main reference for this section is [5].

Let p be an odd prime. Let $X(p)$ denote the modular curve over $\mathbb{Q}$ classifying the isomorphism classes of pairs of generalised elliptic curves with full level p structure. In other words, $X(p)$ represents the functor $\mathbf{Sch}_{\mathbb{Q}}^{\mathrm{op}} \to \mathbf{Set}$ that, to a $\mathbb{Q}$-scheme S, associates the set of isomorphism classes of pairs (E, α_p) consisting of a *generalised* elliptic curve E over S whose geometric fibres are either smooth or p-sided Néron polygons (for the definitions of Néron polygons and generalised elliptic curves, see chapter II of [5]), and a full level p structure α_p, i.e., an S-isomorphism

$$\alpha_p : E^{\mathrm{sm}}[p] \to (\mathbb{Z}/p\mathbb{Z})^2_{/S},$$

where E^{sm} denotes the smooth locus of E/S. The curve $X(p)$ is a smooth, projective curve over $\mathbb{Q}$ whose base extension to $\mathbb{Q}(\zeta_p)$ has $p-1$ connected components. Moreover, it admits a left action of $\mathrm{GL}_2(\mathbb{F}_p)$ which, at the level of its moduli interpretation, can be described as follows: given $M \in \mathrm{GL}_2(\mathbb{F}_p)$, and a point $x \in X(p)(S)$ classified by the pair (E, α_p), the point Mx is classified by the pair $(E, M \circ \alpha_p)$.

The *cusps* of $X(p)$ are defined to be the $\bar{\mathbb{Q}}$-points of $X(p)$ which are classified by a pair (E, α_p) whose generalised elliptic curve E is singular (i.e., a Néron polygon). Any pair of this kind is isomorphic to a pair whose Néron p-gon is the standard one (see chapter II of [5]). Recall that if E is a standard p-gon over S, then there is a canonical S-isomorphism

$$E^{\mathrm{sm}}[p] = \mu_p \times \mathbb{Z}/p\mathbb{Z}.$$

In what follows, we will often identify $E^{\mathrm{sm}}[p]$ with $\mu_p \times \mathbb{Z}/p\mathbb{Z}$, often without further comment. The field of definition of the cusps of $X(p)$ is $\mathbb{Q}(\zeta_p)$.

There are two sets (in fact, Galois orbits) of cusps of $X(p)$ that we would like to define. For the purpose of this paper, we are going to call these two sets the set of *cusps at infinity* and the set of *cusps at zero*. A *cusp at infinity* will be any cusp of $X(p)$ which is classified by a pair (E, α_p) where E is a standard p-gon and α_p is a p-level structure such that $\alpha_p(\zeta_p, 0) \in \langle (1,0) \rangle$, where ζ_p is a primitive pth root of unity, and $\alpha_p(1,1) = (0,1)$. On the

other hand, a *cusp at zero* will be any cusp of $X(p)$ which is classified by a pair (E, α_p) where E is a standard p-gon and α_p is a p-level structure such that $\alpha_p(\zeta_p, 0) \in \langle (0,1) \rangle$, where ζ_p is a primitive pth root of unity, and $\alpha_p(1,1) = (1,0)$. Each connected component of $X(p)_{/\mathbb{Q}(\zeta_p)} := X(p) \times_{\mathbb{Q}} \mathbb{Q}(\zeta_p)$ contains exactly one cusp at infinity and one cusp at zero. Moreover, the sets of cusps at infinity and at zero of $X(p)$ are disjoint.

Now let H be a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Define X_H to be the curve $H\backslash X(p)$. This is a projective curve over $\mathbb{Q}$ which can be proved to be smooth. In general, the curve X_H will *not* represent the functor $\mathbf{Sch}_{\mathbb{Q}}^{\mathrm{op}} \to \mathbf{Set}$ which to a $\mathbb{Q}$-scheme S associates the set of isomorphism classes of pairs (E, α_H) consisting of a generalised elliptic curve whose generic fibres are either smooth or p-sided Néron polygons, and α_H is a level H structure. Nevertheless, X_H is the coarse moduli space for the corresponding stack. The cusps of X_H are defined to be the $\bar{\mathbb{Q}}$-points of X_H which in the image of the set of cusps of $X(p)$ under the quotient morphism $X(p) \to X_H$. The cusps at infinity (resp. at zero) of X_H are those which are in the image of the cusps at infinity (resp. at zero) of $X(p)$. Note that, like $X(p)$, the curve X_H may have several cusps at infinity and several cusps at zero. Moreover, cusps at infinity of X_H can also be at zero (this happens, for example, when $H = \mathrm{GL}_2(\mathbb{F}_p)$).

It will be important to understand the Galois structure of the cusps of the modular curves we are going to work with. The following results will help us achieve this.

Let

$$M_p := \left((\mathbb{Z}/p\mathbb{Z})^2 - \{(0,0)\}\right) / \pm 1. \tag{3.1}$$

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 $\mathbb{F}_p^*$ via multiplication by the determinant.

Lemma 3.1. *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 $\left(\begin{pmatrix} a \\ b \end{pmatrix}, d\right)$, then ${}^\sigma c$ corresponds to*

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

where χ_p is the cyclotomic character.

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

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

where U is the group of matrices

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

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 $\mathrm{Isom}(\mu_p \times \mathbb{Z}/p\mathbb{Z}, (\mathbb{Z}/p\mathbb{Z})^2)/\pm U$ 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 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, we are done. □

Corollary 3.2. *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 3.1 and the definition of X_H. In order to prove the second one, start by noticing 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. □

Remark 3.3. *Under the bijection constructed in the proof of Lemma 3.1, the cusps at infinity of $X(p)$ are the $(p-1)$ cusps of the form*

$$\left(\begin{pmatrix} a \\ 0 \end{pmatrix}, a \right),$$

and they form a full Galois orbit.

Corollary 3.4. *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 3.2, the Galois orbit of a cusp of X_H associated to an element $\overline{\begin{pmatrix} a \\ b \end{pmatrix}} \in (H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p$ is the set of cusps of X_H associated to the elements in the set*

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

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:

$$H = \{h\gamma_\lambda : \lambda \in \mathbb{F}_p^\times, h \in H \cap \mathrm{SL}_2(\mathbb{F}_p)\}. \tag{3.2}$$

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

According to Lemma 3.1, the Galois orbit of $\left(\begin{pmatrix} a \\ b \end{pmatrix}, 1\right)$ is

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

Applying γ_λ^{-1} on the left, we conclude that, under the identification of Corollary 3.2, the Galois orbit of the cusp $\overline{\begin{pmatrix} a \\ b \end{pmatrix}}$ is

$$\left\{\lambda\gamma_\lambda^{-1} \cdot \overline{\begin{pmatrix} a \\ b \end{pmatrix}} \in (H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p : \lambda \in \mathbb{F}_p^\times\right\}. \tag{3.3}$$

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 $H/(H \cap \mathrm{SL}_2(\mathbb{F}_p))$. Observation (3.2) and the fact that the elements of the set (3.3) are fixed by the action of $\mathrm{SL}_2(\mathbb{F}_p) \cap H$ together yield that the Galois orbit of $\overline{\begin{pmatrix} a \\ b \end{pmatrix}}$ is

$$\left\{\gamma \cdot \overline{\begin{pmatrix} a \\ b \end{pmatrix}} \in (H \cap \mathrm{SL}_2(\mathbb{F}_p))\backslash M_p,\, \gamma \in H\right\},$$

as we wanted. □

In the future, if R is a Dedekind domain whose field of fractions is a number field K, we will write $X_{H/R}$ to mean the minimal regular model of $X_{H/K} := X \times_{\mathbb{Q}} K$ over R. If p is invertible in R, the structure morphism $X_{H/R} \to \mathrm{Spec}(R)$ is smooth.

Before finishing this section, we wish to highlight some of the aspects discussed in the above paragraphs in the case of some particular modular curves that are going to be used during the paper. 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 $X_0(p)$. It is the generic fibre of the coarse moduli space for the stack described in [5, V.1]: this is the stack whose fibre over a scheme S is the category whose objects are pairs (E, C), where E is a generalised elliptic curve over S whose geometric fibres are either smooth or m-gons ($m = 1, p$), and C is a finite flat subgroup of E^{sm} of rank p and intersecting every irreducible component of E. This modular curve has two distinct cusps: one cusp at infinity and one at zero. The cusp at infinity is classified by the pair (E, C), where E is the standard Néron 1-gon and C is the subgroup of $E^{\mathrm{sm}}[p](\bar{\mathbb{Q}}) = \mu_p(\bar{\mathbb{Q}})$ generated by $(\zeta_p, 0)$, where ζ_p is a primitive pth root of unity. The morphism $X(p) \to X_0(p)$ has the following moduli interpretation: if $x \in X(p)(\bar{\mathbb{Q}})$ is classified by a pair (E, α_p), let C be the subgroup of

$E^{\mathrm{sm}}[p]$ generated by the preimage of $(1,0)$; the image of x in $X_0(p)(\bar{\mathbb{Q}})$ is then classified by $(c(E), C)$, where $c(E)$ denotes the contraction of E with respect to C, i.e., the irreducible components of E that do not intersect C are contracted to a point (see [5, IV.1] for the definition of contraction).

When $H = C_{\mathrm{sp}}(p)$, we will denote X_H by $X_{\mathrm{sp}}(p)$, and when $H = N_{\mathrm{sp}}(p)$, the curve X_H will be denoted by $X^+_{\mathrm{sp}}(p)$. As $C_{\mathrm{ns}}(p)$ has index 2 in $N_{\mathrm{ns}}(p)$, the canonical morphism $X_{\mathrm{sp}}(p) \to X^+_{\mathrm{sp}}(p)$ is finite of degree 2, and the same holds for the morphism $X_{\mathrm{ns}}(p) \to X^+_{\mathrm{ns}}(p)$. The curve $X_{\mathrm{sp}}(p)$ has $p+1$ cusps, of which one is at infinity and another one is at zero. The field of definition of these two cusps is $\mathbb{Q}$, but the field of definition of all the other cusps is $\mathbb{Q}(\zeta_p)$. Under the canonical map $X_{\mathrm{sp}}(p) \to X^+_{\mathrm{sp}}(p)$, the cusps at infinity and at zero are identified, while all the other cusps are identified with their conjugate by the action of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}(\zeta_p)^+)$. Therefore, $X^+_{\mathrm{sp}}(p)$ has $(p+1)/2$ cusps, of which one is at infinity and at a zero (and is defined over $\mathbb{Q}$), while the field of definition of all the other cusps is $\mathbb{Q}(\zeta_p)^+$. On the other hand, the modular curve $X^+_{\mathrm{ns}}(p)$ has $p-1$ cusps, all at infinity. Their field of definition is $\mathbb{Q}(\zeta_p)$. The morphism $X_{\mathrm{ns}}(p) \to X^+_{\mathrm{ns}}(p)$ identifies the cusps of $X_{\mathrm{ns}}(p)$ with their conjugate by the action of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}(\zeta_p)^+)$. Therefore, $X^+_{\mathrm{ns}}(p)$ has $(p-1)/2$ cusps, all at infinity, and their field of definition is $\mathbb{Q}(\zeta_p)^+$.

4. The modular curve $X_{G(p)}$

Let p be an odd prime and recall the terminology introduced in sections 2 and 3. Let $H(p)$ denote the group $N_{\mathrm{ns}}(p) \cap N_{\mathrm{sp}}(p)$.

Lemma 4.1. *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 $\begin{pmatrix} a & 0 \\ 0 & \pm a \end{pmatrix}$, $a \in \mathbb{F}_p^\times$, 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 $\begin{pmatrix} 0 & \epsilon_p b \\ \pm b & 0 \end{pmatrix}$, $b \in \mathbb{F}_p^\times$, are in $G(p)$. □

From here onwards, we will assume that p is an odd prime satisfying $p \equiv 2 \pmod 3$. If $\pi_{\mathrm{ns}} : X_{H(p)} \to X^+_{\mathrm{ns}}(p)$ denotes the canonical projection from $X_{H(p)}$ to $X^+_{\mathrm{ns}}(p)$, and if $\pi : X_{G(p)} \to X^+_{\mathrm{ns}}(p)$ denotes the canonical projection from $X_{G(p)}$ to $X^+_{\mathrm{ns}}(p)$, Lemma 4.1

implies that π_{ns} factors through π:

$$\begin{array}{ccc} & & X_{H(p)} \\ & \overset{\pi'}{\swarrow} & \big\downarrow{\scriptstyle \pi_{\mathrm{ns}}} \\ X_{G(p)} & \xrightarrow[\pi]{} & X_{\mathrm{ns}}^+(p) \end{array}, \tag{4.1}$$

where $\pi_{\mathrm{ns}} = \pi \circ \pi'$.

The curve $X_{H(p)}$ also has a natural projection $\pi_{\mathrm{sp}} : X_{H(p)} \to X_{\mathrm{sp}}^+(p)$. By pulling back by π_{ns} and then pushing forward along π_{sp}, one obtains a homomorphism

$$\phi : \mathrm{Div}(X_{\mathrm{ns}}^+(p)_{/\bar{\mathbb{Q}}}) \to \mathrm{Div}(X_{\mathrm{sp}}^+(p)_{/\bar{\mathbb{Q}}}).$$

Note that ϕ is defined over $\mathbb{Q}$, in the sense that if $\sigma \in G_{\mathbb{Q}}$, we have

$$^{\sigma}(\phi(D)) = \phi(^{\sigma}D)$$

for every divisor $D \in \mathrm{Div}(X_{\mathrm{ns}}^+(p)_{/\bar{\mathbb{Q}}})$.

Lemma 4.2. *Let $c \in X_{\mathrm{ns}}^+(p)(\bar{\mathbb{Q}})$ be a cusp of $X_{\mathrm{ns}}^+(p)$. Then*

$$\mathrm{ord}_\infty(\phi(c)) = 1,$$

where $\mathrm{ord}_\infty(D)$ stands for the order of the divisor D at the cusp at infinity of $X_{\mathrm{sp}}^+(p)$.

Proof. Note that the morphism π_{ns} is unramified at the cusps of $X_{H(p)}$. Therefore, the pullback of any cusp of $X_{\mathrm{ns}}^+(p)$ to $X_{H(p)}$ is the sum of $(p+1)/2$ distinct cusps of $X_{H(p)}$, because $\deg(\pi_{\mathrm{ns}}) = (p+1)/2$.

Let us fix $m = (p-1)/2$ and denote $c_1, \cdots, c_m$ the cusps of $X_{\mathrm{ns}}^+(p)$, and $c'_1, \cdots, c'_{m+1}$ the cusps of $X_{\mathrm{sp}}^+(p)$. We thus have

$$\sum_{i=1}^{m} \phi(c_i) = \frac{p-1}{2} \sum_{i=1}^{m+1} c'_i,$$

because $\deg \pi_{\mathrm{sp}} = m$ and π_{sp} is also unramified at the cusps. We also know that the field of definition of each cusp of $X_{\mathrm{ns}}^+(p)$ is $\mathbb{Q}(\zeta_p)^+ := \mathbb{Q}(\zeta_p + \zeta_p^{-1})$ and that the Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_p)^+/\mathbb{Q})$ acts transitively on them. As ϕ is defined over $\mathbb{Q}$, we find

$$\sum_{\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)^+/\mathbb{Q})} {}^{\sigma}(\phi(c)) = \sum_{i=1}^{m} \phi(c_i) = \frac{p-1}{2} \sum_{i=1}^{m+1} c'_i.$$

Set $n := \mathrm{ord}_\infty(\phi(c))$. As the cusp at infinity of $X_{\mathrm{sp}}^+(p)$ is defined over $\mathbb{Q}$, we conclude that

$$\frac{p-1}{2} = \mathrm{ord}_\infty \left(\sum_{\sigma \in \mathrm{Gal}(\mathbb{Q}(\zeta_p)^+/\mathbb{Q})} {}^{\sigma}(\phi(c)) \right) = n \cdot \frac{p-1}{2},$$

from where it follows that $n = 1$, as we wanted. □

Let now $\phi' : \mathrm{Div}(X_{G(p)/\bar{\mathbb{Q}}}) \to \mathrm{Div}(X_{\mathrm{sp}}^+(p)_{/\bar{\mathbb{Q}}})$ be the homomorphism $\pi_{\mathrm{sp},*} \circ \pi'^*$. Just like ϕ, this homomorphism is defined over $\mathbb{Q}$. The following result is a consequence of Lemma 4.2.

Corollary 4.3. *Let $c \in X_{\mathrm{ns}}^+(p)(\overline{\mathbb{Q}})$ be a cusp. There are exactly three cusps of $X_{G(p)}$ lying over c, of which exactly one of them is at infinity. If we denote by c_1 the cusp at infinity over c and by c_2 and c_3 the two other ones, we have*

$$\mathrm{ord}_\infty(\phi'(c_i)) = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{if } i = 2, 3. \end{cases}$$

Proof. Start by observing that

$$\phi(c) = \phi'(c_1) + \phi'(c_2) + \phi'(c_3).$$

It follows that $\mathrm{ord}_\infty \phi'(c_k) = 1$ for some $k \in \{1, 2, 3\}$, and that $\phi'(c_j) = 0$ for $j \neq k$. We may assume, without loss of generality, that $k = 1$. Of course, this means that c_1 is a cusp at infinity and that c_2 and c_3 are not. □

The following lemma is a result on the Galois orbits of the cusps of $X_{G(p)}$ that will play a crucial role later in the paper.

Lemma 4.4. *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.*

Proof. By Corollary 3.4, 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 $C_{\mathrm{ns}}(p)$. 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 $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ corresponds, under θ to the action of the Frobenius automorphism on $\mathbb{F}_{p^2}^\times$. With this in mind, it is easy to see that the set

$$\mathcal{O}_{\mathrm{cubes}} := \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \in M_p : a + \sqrt{\epsilon_p} b \text{ is a cube in } \mathbb{F}_{p^2}^\times \right\}$$

is a Galois orbit for the action of $G(p)$ on M_p. Indeed, if $\gamma \in G(p) \cap C_{\mathrm{ns}}(p)$, 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 $\mathcal{O}_{\text{cubes}}$ in M_p, that we shall denote by $\mathcal{O}'$, is also a Galois orbit because any non-cube of $\mathbb{F}_{p^2}^\times$ can be obtained from a specified non-cube by multiplying by an appropriate cube and applying the Frobenius automorphism if needed.

The set $\mathcal{O}_{\text{cubes}}$ is precisely the set of cusps at infinity of $X_{G(p)}$. □

5. The jacobian of $X_{G(p)}$

The jacobian of $X_{G(p)}$ is going to be denoted by $J_{G(p)}$. The aim of this section is to prove the following key proposition.

Proposition 5.1. *Let p be a prime satisfying the following two conditions:*
(i) $p = 11$ or $p \geq 17$;
(ii) $p \equiv 2 \pmod 3$.
Then the jacobian $J_{G(p)}$ of $X_{G(p)}$ admits a non-trivial optimal quotient A such that
(1) $A(\mathbb{Q})$ is finite;
(2) the kernel of the canonical projection $J_{G(p)} \to A$ is stable under the action of the Hecke operators T_ℓ, where $\ell \neq p$ is a prime.

We start by recalling a morphism from $X_{\text{sp}}^+(p)$ to $J_0(p)$, the jacobian of $X_0(p)$, that was first introduced by Momose [12] in his study of the rational points of $X_{\text{sp}}^+(p)$. This will be defined in terms of two degeneracy morphisms from $X_{\text{sp}}(p)$ to $X_0(p)$ as follows. Let $\mathfrak{X}_{\text{sp}}(p)$ denote the stack over $\mathbb{Z}[1/p]$ associated to the the split Cartan problem; that is, given a $\mathbb{Z}[1/p]$-scheme S, the fibre $\mathfrak{X}_{\text{sp}}(p)(S)$ over S is the category consisting of pairs $(E, (A, B))$, where E is a generalised elliptic curve over S whose geometric fibres are either smooth or p-gons, and A and B are two distinct finite flat (and so étale) subgroup schemes of E^{sm} of rank p. Let also $\mathfrak{X}_0(p)$ denote the stack over $\mathbb{Z}[1/p]$ corresponding to the Borel subgroup problem as described in [5, V.1] (see also section 3 above). The first degeneracy morphism, denoted by d_1, is the rule mapping an object $(E, (A, B))$ in $\mathfrak{X}_{\text{sp}}(p)(S)$ to $(c(E), A)$, where $c(E)$ denotes the generalised elliptic curve obtained by contracting to one point all the irreducible components of the geometric fibres of E that are not intersected by A (see [5, IV.1] for the definition of contraction). The second degeneracy map, denoted by d_p, is the rule that maps $(E, (A, B))$ to $(c(E/B), E[p]/B)$, where, as before, $c(E/B)$ is the generalised elliptic curve obtained by contracting to one point the irreducible components of the geometric fibres of E/B that do not intersect $E[p]/B$.

Remark 5.2. *To be rigorous, we should actually write $(c(E), c(A))$ and $(c(E/B), c(E[p]/B))$ instead of $(c(E), A)$ and $(c(E/B), E[p]/B)$, but we will allow ourselves this abuse of notation, confident that it will cause no confusion to the reader.*

We have thus obtained two $\mathbb{Z}[1/p]$-morphisms of stacks,

$$d_1 : \mathfrak{X}_{\text{sp}}(p) \to \mathfrak{X}_0(p) \quad \text{and} \quad d_p : \mathfrak{X}_{\text{sp}}(p) \to \mathfrak{X}_0(p),$$

and these naturally induce morphisms of their coarse moduli spaces, which we shall also denote by d_1 and d_p. We can now define $h : X_{\text{sp}}(p) \to J_0(p)$ by setting

$$h : x \mapsto \text{cl}(d_1(x) - d_p(x)). \tag{5.1}$$

The cusps at infinity and at zero of $X_0(p)$ will be denoted by ∞ and 0, respectively. The following lemma is well-known, but we will nevertheless give a proof, for the benefit of the reader.

Lemma 5.3. *Let $c \in X_{\mathrm{sp}}(p)(\bar{\mathbb{Q}})$ be a cusp. We have*

$$h(c) = \begin{cases} 0 & \text{if } c \text{ is the cusp at infinity or at zero} \\ \mathrm{cl}(0-\infty) & \text{otherwise.} \end{cases}$$

Proof. If c is the cusp at infinity of $X_{\mathrm{sp}}(p)$, then it is classified by the pair

$$(E, (\langle(\zeta_p, 0)\rangle, \langle(1,1)\rangle)),$$

where E is the standard Néron p-gon and ζ_p is a primitive pth root of unity. Under d_1, the cusp c is mapped to the cusp of $X_0(p)$ classified by $(c(E), \langle(\zeta_p, 0)\rangle)$, which is ∞, the cusp at infinity of $X_0(p)$. On the other hand, d_p maps c to the cusp of $X_0(p)$ classified by $(c(E/\langle(1,1)\rangle), E[p]/\langle(1,1)\rangle)$. But this is once again the cusp at infinity of $X_0(p)$. Therefore, $h(c) = \mathrm{cl}(d_1(c) - d_p(c)) = 0$. A similar argument yields the same conclusion in the case where c is the cusp at zero of $X_{\mathrm{sp}}(p)$.

Suppose now that c is neither the cusp at infinity nor the cusp at zero. It is then classified by a pair

$$(E, (\langle(\zeta_p^a, r)\rangle, \langle(\zeta_p^b, s)\rangle)),$$

where, as before, E is a Néron p-gon and now $p \nmid rs$. The image of c under d_1 is the cusp of $X_0(p)$ classified by $(E, \langle(\zeta_p^a, r)\rangle)$ (as r is not divisible by p, there is no contraction), which is the cusp at zero of $X_0(p)$. Under d_p, the image of c is classified by $(c(E/\langle(\zeta_p^b, s)\rangle), E[p]/\langle(\zeta_p^b, s)\rangle))$. As p does not divide s, this contracts to a Néron 1-gon and so the cusp of $X_0(p)$ in question is the cusp at infinity. Therefore,

$$h(c) = \mathrm{cl}(d_1(c) - d_p(c)) = \mathrm{cl}(0-\infty),$$

as we wanted. □

Let w_p denote the Atkin–Lehner involution of $X_0(p)$ and denote by ω_p the involution of $X_{\mathrm{sp}}(p)$ whose moduli description is $(E, (A, B)) \mapsto (E, (B, A))$. It is easy to check that

$$w_p \circ d_p = d_1 \circ \omega_p \quad \text{and} \quad w_p \circ d_1 = d_p \circ \omega_p. \tag{5.2}$$

It follows that

$$h \circ \omega_p = -w_p \circ h. \tag{5.3}$$

We conclude that there exists a morphism $\tilde{h} : X_{\mathrm{sp}}(p)^+ \to J_0(p)/(1+w_p)J_0(p)$ making the following diagram commute:

$$\begin{array}{ccc} X_{\mathrm{sp}}(p) & \xrightarrow{\quad h \quad} & J_0(p) \\ \downarrow & & \downarrow \\ X_{\mathrm{sp}}^+(p) & \xrightarrow[\tilde{h}]{} & J_0(p)/(1+w_p)J_0(p) \end{array}.$$

Abusing notation, we will denote the image of $\mathrm{cl}(0-\infty)$ in $J_0(p)/(1+w_p)J_0(p)$ by $\mathrm{cl}(0-\infty)$ as well. As ω_p interchanges the cusps at infinity and at zero of $X_{\mathrm{sp}}(p)$, Lemma 5.3 yields the following result.

Corollary 5.4. *Let $c \in X_{\mathrm{sp}}^+(p)(\bar{\mathbb{Q}})$ be a cusp. We have*

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

Let B denote the Eisenstein quotient of $J_0(p)$ (see [10, II.10 Definitions 10.4]). It can be shown (see [10, Proposition 17.10]) that the quotient map $J_0(p) \to B$ factors through a morphism

$$q : J_0(p)/(1+w_p)J_0(p) \to B.$$

Let c be a cusp at infinity of $X_{G(p)}$ and let

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

denote the Abel–Jacobi map centred at c (we need to extend the base field because the cusps of $X_{G(p)}$ are not defined over $\mathbb{Q}$). Using the notation of the previous section, define

$$\phi' : J_{G(p)} \to J_{\mathrm{sp}}^+(p)$$

to be the homomorphism of abelian varieties obtained from pulling back by π' and pushing forward by π_{sp}. Note that we are reusing the symbol ϕ' for this homomorphism when we have already used it for a homomorphism $\mathrm{Div}(X_{G(p)/\bar{\mathbb{Q}}}) \to \mathrm{Div}(X_{\mathrm{sp}}^+(p)_{/\bar{\mathbb{Q}}})$. This is because the homomorphism induced by the latter on the rational equivalence classes of degree zero divisors coincides with the homomorphism defined by the former on the $\bar{\mathbb{Q}}$-points of the jacobians.

We can now define

$$f : X_{G(p)/\mathbb{Q}(\zeta_p)} \to B_{/\mathbb{Q}(\zeta_p)}$$

as $f := q \circ \tilde{h} \circ \phi' \circ \iota_c$.

Lemma 5.5. *The image of the morphism f in $B_{/\mathbb{Q}(\zeta_p)}$ is not reduced to a point.*

Proof. Recall that one of the constraints on our prime p is that $p = 11$ or $p \geq 17$. In particular, $X_0(p)$ has positive genus.

We know that $f(c) = 0$. If c is a cusp at infinity, let c' be a cusp of $X_{G(p)}$ not at infinity. Using Corollary 5.4, it is easy to conclude that

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

Now, a result due to Mazur [10, III.1 Corollary 1.4] implies that $\mathrm{cl}(0-\infty)$ is not zero in B, and so we conclude that the image of f contains, at least, two different points.

If c is not at infinity, the same argument can be used by requiring c' to be a cusp at infinity now. (Or we can just observe that changing the choice of cusp changes f by a translation.) □

Of course, it follows from this lemma that the image of f in $B_{/\mathbb{Q}(\zeta_p)}$ is in fact a 1-dimensional subscheme of $B_{/\mathbb{Q}(\zeta_p)}$.

Consider now the morphism $g := q \circ \tilde{h} \circ \phi' : J_{G(p)} \to B$. Let K denote the kernel of g, which is a subgroup scheme of $J_{G(p)}$. Let K^0 denote the connected component of the identity. This is an abelian subvariety of $J_{G(p)}$.

Lemma 5.6. *Let ℓ be a prime different from p. The abelian subvariety K^0 of $J_{G(p)}$ is stable under the action of the Hecke operator T_ℓ.*

Proof. By the work of Mazur [11], we already know that T_ℓ acts on B. It is evident from the definition of T_ℓ as a correspondence that $T_\ell \circ g = g \circ T_\ell$. As T_ℓ is continuous and an endomorphism of abelian varieties, it is also clear that $T_\ell(K^0) \subseteq K^0$. □

Proof of Proposition 5.1. Define $A := J_{G(p)}/K^0$. According to Lemma 5.5, the image of f in $B_{/\mathbb{Q}(\zeta_p)}$ is not reduced to a single point. Therefore, the image of the homomorphism $g : J_{G(p)} \to B$ defined above is not trivial, from where it follows that the quotient A is not trivial either. By construction, this quotient is also optimal.

As B is the Eisenstein quotient of $J_0(p)$, $B(\mathbb{Q})$ is a finite set. Now, g is defined over $\mathbb{Q}$, and so its image, which is isogenous to A, must have only finitely many $\mathbb{Q}$-rational points. Property (2) in the statement of Proposition 5.1 is the content of Lemma 5.6. □

6. A formal immersion

Let $p \equiv 2 \pmod 3$ be a prime not in the set $\{2, 3, 5, 7, 13\}$, and let A be a quotient of $J_{G(p)}$ satisfying the conditions of Proposition 5.1. Let $p_A : J_{G(p)} \to A$ denote the canonical projection. Let $c \in X_{G(p)}(\mathbb{Q}(\zeta_p))$ be a cusp and let $\iota_c : X_{G(p)/\mathbb{Q}(\zeta_p)} \to J_{G(p)/\mathbb{Q}(\zeta_p)}$ denote the Abel–Jacobi map centered at a cusp at c. Define $f_c := p_A \circ \iota_c : X_{G(p)/\mathbb{Q}(\zeta_p)} \to A_{/\mathbb{Q}(\zeta_p)}$. For any Dedekind domain R where p is invertible and whose fraction field F has characteristic 0, we will write $X_{G(p)/R}$ for the minimal regular model of $X_{G(p)}$ over R (as we are assuming that p is invertible in R, this model is actually smooth over R), by $A_{/R}$ the Néron model of A over R, and by $f_{c/R} : X_{G(p)/R} \to A_{/R}$ the R-morphism induced by f_c. Also, if $x \in X_{G(p)}(F)$ is an F-point of $X_{G(p)}$ and $\mathfrak{p}$ is a maximal ideal of R, we will write $k(\mathfrak{p})$ for the residue field $R/\mathfrak{p}$ and $x_{/k(\mathfrak{p})}$ for the special fibre over $\mathfrak{p}$ of the closure of x. This is naturally identified with a closed point in $X_{G(p)/R}$ lying over $\mathfrak{p}$. Similar considerations will be valid for $A_{/R}$.

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.

From here to the end of the paper, let R stand for the Dedekind domain $\mathbb{Z}[\zeta_p, 1/p]$. If $\mathfrak{p}$ is a prime ideal of R, we will denote by $R_\mathfrak{p}$ the $\mathfrak{p}$-adic completion of R. The following is the main result of this section.

Proposition 6.1. *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), and let $c \in X_{G(p)}(\mathbb{Q}(\zeta_p))$ be a cusp. The morphism f_{c/R_λ} is a formal immersion at $c_{/k(\lambda)}$.*

As $c_{/k(\lambda)}$ and $f(c_{/k(\lambda)}) = 0_{/k(\lambda)}$ are both defined over $k(\lambda)$, proving this proposition is equivalent to showing that the induced $k(\lambda)$-linear map of cotangent spaces

$$f^*_{c/k(\lambda)} : \mathrm{Cot}(A_{/k(\lambda)}) \to \mathrm{Cot}_c(X_{/k(\lambda)}) \tag{6.1}$$

is surjective. As $X_{/k(\lambda)}$ is 1-dimensional, it is enough to show that $f^*_{c/k(\lambda)}$ is not trivial. In order to be ready to prove this, we first recall a result of Mazur [11, Lemma 2.1].

6.1. **Mazur's lemma.** The content of this subsection is completely contained in a more general form in Mazur's paper [11]. Let $J_{G(p)/R}$ be the Néron model of the jacobian of $X_{G(p)}$ over R. As $X_{G(p)/R}$ is smooth over R, its jacobian $J_{G(p)/R}$ is proper over R. Let $\iota_{c/R} : X_{G(p)/R} \to J_{G(p)/R}$ be the Abel–Jacobi map centred at c. Given a cusp $c \in X_{G(p)}(\mathbb{Q}(\zeta_p))$, we obtain a homomorphism

$$\iota^*_{c/R} : \mathrm{Cot}(J_{G(p)/R}) \to \mathrm{Cot}_c(X_{G(p)/R}) \tag{6.2}$$

of free R-modules.

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

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

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

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

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

Lemma 6.2 (Mazur [11])**.** *We have $\iota^*_{c/R} = \pm v \circ \Theta$.*

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 $X_{G(p)}$: the point is that we can identify $\mathrm{Cot}_c(X_{G(p)/R})$ with R in such a way that the diagram

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

commutes.

6.2. **Formal immersion at the cusps.** The last results we need before we are ready to prove Proposition 6.1 are the following two lemmas.

Lemma 6.3. *Let ω be an element of $\mathrm{H}^0(X_{G(p)/\mathbb{Z}[1/p]}, \Omega^1_{X_{G(p)}/\mathbb{Z}[1/p]})$. Let c be a cusp of $X_{G(p)}$ 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 analogue assertion for $X(p)$, so we show that the lemma holds in this case. Indeed, recall (see [6, 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 6.4. *Let r be a prime number different from p, and let ω be as in the statement of Lemma 6.3. 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 [6, 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 [6, 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 [6] yields that if c is a cusp of $X(p)$ correponding 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 $X(p)$ correponding 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 $\mathbb{Z}/p\mathbb{Z}$).

In the notation of Lemma 3.1, if c is a cusp of $X_{G(p)}$ represented by $\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, 1\right)$ (they all are represented by an element of this form because $\det G(p) = \mathbb{F}_p^{\times}$), then c' will be represented by $\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, r^{-1}\right)$. As $\det G(p) = \mathbb{F}_p^{\times}$, this means that if c is a cusp represented by $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, then c' will be represented by $\gamma_r \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, where γ_r is an element of $G(p)$ of determinant r. Corollary 3.4 now yields that c' is in the Galois orbit of c. □

Proof of Proposition 6.1. The proof is standard. As $A_{/k(\lambda)}$ is not trivial, there exists a non-zero $\omega \in \mathrm{Cot}(A_{/k(\lambda)})$. By specialisation results due to Raynaud (as stated, for example,

in [11, 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 $\mathrm{Cot}(J_{G(p)/k(\lambda)})$.

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 6.2 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 6.4, there exists $\sigma \in G_{\mathbb{Q}}$ such that $a_1(c, T_r\omega) = a_r({}^{\sigma}c, \omega)$. By Lemma 6.3, 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 $G(p)$, it is fixed under the action of the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ generated by $G(p)$ 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 $x \in X_{G(p)}(\mathbb{Q}(\zeta_p)^+)$ and a maximal ideal λ of R, define

$$B_\lambda(x) := \{y \in X_{G(p)}(\mathbb{Q}) : y_{/k(\lambda)} = x_{/k(\lambda)}\},$$

i.e., $B_\lambda(x)$ is the set of $\mathbb{Q}$-rational point in the residue class of x modulo λ.

Corollary 6.5. *Let c be a cusp of $X_{G(p)}$ 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 $B_\lambda(c) = \emptyset$.*

Proof. Say that the characteristic of λ is ℓ. Suppose, for contradiction, that $B_\lambda(c)$ is non-empty, and let $x \in B_\lambda(c)$. Consider $f_c : X_{G(p)/\mathbb{Q}(\zeta_p)} \to A_{/\mathbb{Q}(\zeta_p)}$. 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 [4]. It goes as follows. We first show that a multiple of $\iota_c(x)$ is a $\mathbb{Q}$-rational point in $J_{G(p)}$. 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 $m_\sigma \cdot \mathrm{cl}(c - {}^\sigma c) \in J_{G(p)}(\mathbb{Q})$. 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 6.1. Consider the R_λ sections of $X_{G(p)}$ defined by x and c. Let $h_x : \hat{\mathcal{O}}_{X_{G(p)},x_{/k(\lambda)}} \to R_\lambda$ be the homomorphism of completed local rings at the special fibres induced by x, and let $h_c : \hat{\mathcal{O}}_{X_{G(p)},c_{/k(\lambda)}} \to R_\lambda$ be that induced by c. Note that as $c_{/k(\lambda)} = x_{/\lambda}$, we have $\hat{\mathcal{O}}_{X_{G(p)},x_{/k(\lambda)}} = \hat{\mathcal{O}}_{X_{G(p)},x_{/k(\lambda)}}$. 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 6.1 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 $\mathbb{Q}(\zeta_p)$. □

7. The image of the residual Galois representation in the non-integral case

We need only one last result in order to be ready to prove Theorem 1.3.

Proposition 7.1. *Let E be an elliptic curve defined over $\mathbb{Q}$. Suppose that there exists a prime number $p \geq 5$ such that the image of $\bar{\rho}_{E,p}$ is contained in the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Then E has potentially good reduction at every prime ℓ satisfying $\ell \not\equiv \pm 1 \pmod{p}$.*

Proof. This result is Proposition 2.2 in [9] and a proof can be found there. □

Proof of Theorem 1.3. Let p be a prime not in the set $\{2, 3, 5, 7, 11, 13, 17, 37\}$. Let $E/\mathbb{Q}$ be an elliptic curve such that $j(E) \notin \mathbb{Z}$. Suppose that the image of $\bar{\rho}_{E,p}$ is neither $\mathrm{GL}_2(\mathbb{F}_p)$ nor the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Zywina's result 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 $\mathrm{GL}_2(\mathbb{F}_p)$. 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 $v_\ell(j(E)) < 0$. Since the image of $\bar{\rho}_{E,p}$ is contained in the normaliser of a non-split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, Proposition 7.1 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 $x_{/k(\lambda)} = c_{/k(\lambda)}$. In other words, $x \in B_\lambda(c)$. But this contradicts Corollary 6.5. □

We have thus obtained that the j-invariant of any non-cuspidal point in $X_{G(p)}(\mathbb{Q})$ is integral. The treatment of the rational points of $X_{G(p)}$ with integral j-invariant will be done in the next section.

8. 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. We will prove the following.

Proposition 8.1. *For any prime $p \equiv 2 \bmod 3$ not in $I(1)$, if $P \in X_{G(p)}(\mathbb{Q})$ and $j(P) \in \mathbb{Z}$,*

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

Corollary 8.2. *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 8.2. Assume that Proposition 8.1 holds and keep the notations of the Corollary. By the explicit surjectivity theorem of [8, (7) and Theorem 5.4] (only making use of the fact that the image is included in the normaliser of a nonsplit Cartan), 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 8.1. In the case $\log |j(E)| \leq 12 \cdot 985$, we get by the same surjectivity theorem an absolute upper bound on p^2 giving a smaller upper bound on p. □

To prove Proposition 8.1, we use the following conventions:

• Due to Lemma 4.4, 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. To do it in practice, one needs to define properly a modular unit. We follow the results of [7] and lemmas of Bajolet, Bilu and Matschke [1] here.

• $e(x) := e^{2i\pi x}$ for any complex number x.

• $\mathcal{H}$ is the Poincaré 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 notations) with common denominator p, we define a modular function $g_{\underline{a}}$ on $\mathcal{H}$ whose q-expansion is $\mathcal{H}$ is

$$g_{\underline{a}}(\tau) = q^{B_2(a_1)/2} e(a_2(a_1-1)) \prod_{n=0}^{+\infty} (1 - q^{n+a_1} e(a_2))(1 - q^{n+1-a_1} e(-a_2)), \tag{8.1}$$

where $B_2(X) = X^2 - X + 1/6$ 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 $(\mathbb{Z}/p\mathbb{Z})^2 \backslash \{(0,0)\}$ (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 $(a_1, a_2) := \frac{(\widetilde{a},\widetilde{b})}{p}$, and for $m \in \mathbb{N}^*$, the product

$$U_{\mathcal{O},m} := \prod_{(a,b)\in\mathcal{O}} g^m_{(a_1,a_2)} \tag{8.2}$$

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

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

and 6 divides $m|\mathcal{O}|$ [1, Theorem 5.1].

For our orbit $\mathcal{O}_{\text{cubes}}$, the vanishing of the three sums holds, because the set of cubes is stable by multplication by scalars of $\mathbb{F}_p^*$ so each of the three sums above should be equal to itself times an in $\mathbb{F}_p^*$, hence 0. We just have to choose $m = 3$ then, and define

$$U := \zeta \cdot U_{\mathcal{O}_{\text{cubes}},3} \tag{8.3}$$

our modular unit from now on (where ζ a root of unity making U belong to $\mathbb{Q}(X_{G(p)})$, as mentioned before). As for the integrality, every g_{a_1,a_2} is integral over $\mathbb{Z}[j]$, and it is easily seen that

$$\prod_{(a,b)\in M_p} g^3_{a_1,a_2} = \pm p^3,$$

so $p^3 U^{-1}$ is also integral over $\mathbb{Z}[j]$. Consequently, for every $P \in X_{G(p)}(\mathbb{Q})$ with $j(P) \in \mathbb{Z}$,

$$U(P) \in \mathbb{Z} \text{ and } 0 \leq \log|U(P)| \leq 3\log(p), \tag{8.4}$$

which is the whole point of considering this modular unit.

Using the expansion at infinity, we can write that for every $\tau \in \mathcal{H}$,

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

where

$$\mathrm{Ord}_q(U) = 3 \sum_{(a,b)\in\mathcal{O}_{\text{cubes}}} B_2(a_1)/2, \quad \rho_U = \prod_{(a,b)\in\mathcal{O}_{\text{cubes}}} \rho^3_{a_1,a_2}$$

with from (8.1)

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

Finally,

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

where

$$\log|R_{a_1,a_2}(\tau)| = \sum_{n\geq 0} \log|1-q^{n+a_1}e(a_2)| + \log|1-q^{n+a_1}e(a_2)|$$

and we then use $|\log|1-z|| \leq -\log|1-|z||$ for $|z|\leq 1$ for $n=0$, and $|\log|1-z|| \leq \frac{|z|}{1-|z|}$ for the other terms (if $a_1 = 0$, the first $n=0$ term is put into $\rho_{(0,a_2)}$. We thus get for $a_1 \neq 0$

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

and

$$|\log|R_{0,a_2}(\tau)|| \leq \frac{2|q|}{1-|q|}.$$

• For $a=0$, all nonzero b's 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.

• The orbit $\mathcal{O}_{\text{cubes}}$ is stable by scalar multiplication by $\mathbb{F}_p^*$, which means that all of fibers of $(a,b)\mapsto a$ have the same cardinality except above 0. They are thus of cardinality $(p-2)/3$.

We now have

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

after computation. 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, gathering the previous inequalities for the product expansion,

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

after [3, 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 $a+\varepsilon_p b$ 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:

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

where

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

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 $\mathcal{O}_{\mathrm{cubes}}$ does not have 0 in its image, and each other element of $\mathbb{F}_p^*$ has $(p+1)/3$ elements in its fiber (again by stability by multiplication by $\mathbb{F}_p^*$).

Putting this together, we obtain

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

and for the choice of γ above,

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

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 uniformization 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 $C_{\mathrm{ns}}(p)$ 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 (8.4) to bound the corrresponding term in one of the two previous inequalities. The first case gives

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

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

$$|\log|q|| \leq 1.2 + \frac{13p}{|\log|q||}.$$

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

$$|\log|q|| \leq 1.2 + \frac{27p}{|\log|q||}.$$

Both cases give rise to second-degree polynomial inequalities which we can readily solve, and using then the estimates explained in [3, p.969], after simplification,

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

We can retrieve the remaining cases $p < 100$ by refining the estimates above (or by using the main theorem of [1]), and the case $\log|q| \leq \sqrt{p}$ by [3, p.969] again, which concludes the proof.

References

[1] A. Bajolet, Y. Bilu, and B. Matschke. Computing integral points on $X_n s^+(p)$, 2018.

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

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

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

[5] P. Deligne and M. Rapoport. Les schémas de modules de courbes elliptiques. pages 143–316. Lecture Notes in Math., Vol. 349, 1973.

[6] N. M. Katz. p-adic properties of modular schemes and modular forms. pages 69–190. Lecture Notes in Mathematics, Vol. 350, 1973.

[7] D. Kubert and S. Lang. *Modular units*. Springer-Verlag, 1981.

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

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

[10] B. Mazur. Modular curves and the Eisenstein ideal. *Inst. Hautes Études Sci. Publ. Math.*, (47):33–186 (1978), 1977.

[11] B. Mazur. Rational isogenies of prime degree (with an appendix by D. Goldfeld). *Invent. Math.*, 44(2):129–162, 1978.

[12] F. Momose. Rational points on the modular curves $X_{\text{split}}(p)$. *Compositio Math.*, 52(1):115–137, 1984.

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

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

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

[16] D. Zywina. On the possible images of the mod ell representations associated to elliptic curves over Q. *arXiv e-prints*, page arXiv:1508.07660, Aug 2015.

Univ. Grenoble Alpes, CNRS, IF, 38000 Grenoble, France and Department of Mathematics, University College London, 25 Gordon Street, London WC1H 0AY, United Kingdom

Email address: Samuel.Le-Fourn@univ-grenoble-alpes.fr and lemos.pj@gmail.com