CrossMark

Elliptic curves over real quadratic fields are modular

Nuno Freitas · Bao V. Le Hung · Samir Siksek

Abstract We prove that all elliptic curves defined over real quadratic fields
are modular.

Mathematics Subject Classification Primary 11F80; Secondary 11G05

N. Freitas
Mathematisches Institut, Universität Bayreuth, 95440 Bayreuth, Germany
e-mail: nunobfreitas@gmail.com

B. V. Le Hung
Department of Mathematics, Harvard University, One Oxford Street, Cambridge,
MA 02138, USA
e-mail: lhvietbao@googlemail.com

S. Siksek (✉)
Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK
e-mail: samir.siksek@gmail.com

Contents

1 Introduction

1.1 Summary of results

One of the great achievements of modern number theory is the proof by Breuil, Conrad, Diamond and Taylor [9] of the Shimura–Taniyama (or modularity) conjecture for elliptic curves over $\mathbb{Q}$. This built on the earlier pioneering work of Wiles [75], and of Taylor and Wiles [73], which established the modularity of semistable elliptic curves over $\mathbb{Q}$, and was historically the final step in the proof of Fermat's Last Theorem. In this paper we prove the analogous modularity statement for elliptic curves over real quadratic fields.

Let K be a totally real number field, and write $G_K := \mathrm{Gal}(\overline{\mathbb{Q}}/K)$ for its absolute Galois group. Let E be an elliptic curve over K. Recall that E is

modular if there exists a Hilbert cuspidal eigenform $\mathfrak{f}$ over K of parallel weight 2, with rational Hecke eigenvalues, such that the Hasse–Weil L-function of E is equal to the L-function of $\mathfrak{f}$. A more conceptual way to phrase this is that there is an isomorphism of compatible systems of Galois representations

$$\rho_{E,p} \cong \rho_{\mathfrak{f},p}$$

where the left-hand side is the Galois representation arising from the action of G_K on the p-adic Tate module $T_p(E)$, while the right-hand side is the Galois representation associated to $\mathfrak{f}$ by Eichler and Shimura (when $K = \mathbb{Q}$) [20], and by Carayol [12,13], Blasius and Rogawski [8], Wiles [76] and Taylor [71] (for general totally real K).

Work of Jarvis and Manoharmayum [35] establishes modularity of semi-stable elliptic curves over $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{17})$, and there are various other works [1,21], which establish modularity under local assumptions on the curve E and the field K. In this paper, we prove modularity of all elliptic curves over all real quadratic fields.

Theorem 1 *Let E be an elliptic curve over a real quadratic field K. Then E is modular.*

Before sketching the strategy of the proof of Theorem 1, we need a little more notation and terminology. Let p be a rational prime. Denote by

$$\overline{\rho}_{E,p} \ : \ G_K \to \operatorname{Aut}(E[p]) \cong \operatorname{GL}_2(\mathbb{F}_p)$$

the representation giving the action of G_K on the p-torsion of E. We say $\overline{\rho}_{E,p}$ is **modular** if there exists a Hilbert cuspidal eigenform $\mathfrak{f}$ over K of parallel weight 2, and a place $\lambda \mid p$ of $\overline{\mathbb{Q}}$ such that $\overline{\rho}_{E,p}^{\mathrm{ss}} \sim \overline{\rho}_{\mathfrak{f},\lambda}^{\mathrm{ss}}$.

For our purposes, the following is a convenient sketch of the main steps in the proof [9,15,19,73,75] of modularity of elliptic curves over $\mathbb{Q}$:

Step I (modularity lifting): one shows that if $\overline{\rho}_{E,p}$ is modular and the image of $\overline{\rho}_{E,p}$ is 'sufficiently large' (in a suitable sense), then $\rho_{E,p}$ is modular.

Step II (Langlands–Tunnell): if $\overline{\rho}_{E,3}$ is irreducible then it is modular.

Step III (3–5 modularity switching): with the help of Steps I and II, one shows that if $\overline{\rho}_{E,5}$ is irreducible then it is modular.

Step IV (dealing with exceptions): putting Steps I, II, III together, one sees that E is modular except possibly if the image of $\overline{\rho}_{E,p}$ is 'small' for $p = 3$ and 5 simultaneously. Such elliptic curves have certain mod 3 and mod 5 level structures, and so give rise to rational points on certain modular curves. By enumerating these rational points (of which there are only a handful) one shows that these elliptic curves are modular by some other means.

We are ready to outline the proof of Theorem 1 and its similarities and differences from the above; we also outline the contents of the paper. First we need a suitable modularity lifting theorem (Step I), and the following is a relatively straightforward consequence of a theorem of Breuil and Diamond [10, Théorème 3.2.2], which in turn is a consequence of deep results due to Kisin [40], Gee [31], and Barnet-Lamb, Gee and Geraghty [3,4].

Theorem 2 *Let E be an elliptic curve over a totally real number field K, and let $p \neq 2$ be a rational prime. Write $\overline{\rho} = \overline{\rho}_{E,p}$. Suppose*

 (i) *$\overline{\rho}$ is modular,*
(ii) *$\overline{\rho}(G_{K(\zeta_p)})$ is absolutely irreducible.*

   Then E is modular.

The derivation of Theorem 2 from the aforementioned theorem of Breuil and Diamond is given in Sect. 2.1. Sections 2.2 and 2.3 are devoted to background results on modular curves and images $\overline{\rho}(G_{K(\zeta_p)})$ that are needed in the remainder of the paper. In Sects. 2.4 and 3.1 we explain how Langlands–Tunnell (Step II) and Wiles' 3–5 modularity switching argument (Step III) carry over to the setting of elliptic curves over totally real fields. These, together with Theorem 2 give the following.

Theorem 3 *Let $p = 3$ or 5. Let E be an elliptic curve over a totally real field K. Suppose that $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is absolutely irreducible. Then E is modular.*

We are interested in proving modularity of elliptic curves over real quadratic fields. If we are to follow the same strategy as over $\mathbb{Q}$, we will have to examine the non-cuspidal real quadratic points on certain modular curves; these parametrize elliptic curves where $\overline{\rho}_{E,3}(G_{K(\zeta_3)})$ and $\overline{\rho}_{E,5}(G_{K(\zeta_5)})$ are simultaneously absolutely reducible. Some of these modular curves (for example the elliptic curve $X_0(15)$) have infinitely many real quadratic points, and there seems to be no simple way to prove that these points are modular. We are forced to interpolate a further step into the above strategy, by adding a 3–7 modularity switching argument (Sect. 3.2), which incorporates some ideas of Manoharmayum [44,45].

Theorem 4 *Let E be an elliptic curve over a totally real field K. Suppose that $\overline{\rho}_{E,7}(G_{K(\zeta_7)})$ is absolutely irreducible. Then E is modular.*

Thanks to the addition of Theorem 4, an elliptic curve E over a totally real field K is modular except possibly if the images $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ are simultaneously absolutely reducible for $p = 3, 5, 7$. In Sect. 3.3 we show that such an elliptic curve gives rise to a K-point on one of 27 modular curves:

$$X(u, v, w) \qquad u \in \{\text{b3, s3, ns3}\}, \quad v \in \{\text{b5, s5, ns5}\}, \quad w \in \{\text{b7, s7, ns7}\}.$$
$$(1)$$

See Sect. 2.2 for an explanation of the notation, but for now b, s and ns respectively stand for 'Borel', 'normalizer of split Cartan' and 'normalizer of non-split Cartan'.

The curves in (1) all have genera >1, and so by Faltings' Theorem [24] possess only finitely many K-points.

Theorem 5 *Let K be a totally real field. All but finitely many $\overline{K}$-isomorphism classes of elliptic curves over K are modular.*

Now a possible route to proving Theorem 1 is to attempt to enumerate all non-cuspidal real quadratic points on the 27 modular curves in (1) (or suitable quotients) and then prove that the resulting points are modular. This appears to be a monumental task, and some of the curves on this list have defeated us.

In Sects. 4.1 and 4.2 we carry out a more precise examination of cases where $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is absolutely reducible, and the corresponding modular curves. This allows us (Sect. 4.3) to reduce the proof of Theorem 1 to establishing Lemmas 1.1 and 1.2 below.

Lemma 1.1 *Elliptic curves over $\mathbb{Q}(\sqrt{5})$ are modular.*

Lemma 1.2 *All non-cuspidal real quadratic points on the modular curves*

$$X(\text{b5}, \text{b7}), \qquad X(\text{b3}, \text{s5}), \qquad X(\text{s3}, \text{s5}), \tag{2}$$

$$X(\text{b3}, \text{b5}, \text{d7}), \quad X(\text{s3}, \text{b5}, \text{d7}), \quad X(\text{b3}, \text{b5}, \text{e7}), \quad X(\text{s3}, \text{b5}, \text{e7}), \tag{3}$$

correspond to modular elliptic curves.

The notation d7 and e7 is explained in Sect. 4.2, but for now we remark that they indicate mod 7 level structures that are respectively finer than 'normalizer of split Cartan' and 'normalizer of non-split Cartan'.

The seven modular curves in (2), (3) have genera 3, 3, 4, 97, 153, 73 and 113. It seems that these modular curves are not amenable to the methods of Mazur [47], Kamienny [37] and Merel [48] (we comment on this at the end of Sect. 4.3). In the literature there are explicit methods which, in certain favourable circumstances, are capable of determining all quadratic points on a given curve (e.g. [11,61]), using an explicit model for the curve and knowledge of the Mordell–Weil group of its Jacobian. As far as we are aware, such methods have only been applied to curves of genus 2 and 3, and our computations for the three curves in (2) push these methods to their limit. In Sects. 5.1, 5.2, 5.3 we determine the quadratic points on the three modular curves in (2), and deduce modularity of the non-cuspidal real quadratic points.

The four modular curves in (3) are extremely complicated on account of their large genera, and certainly beyond the aforementioned explicit methods found in the literature. Instead we give a representation of these curves as normalizations of fibre products of curves (Sect. 5.4), and develop an analogue of the

Mordell–Weil sieve that applies to such fibre products (Sect. 5.5). Remarkably, we are able to show (Sect. 5.6) with the help of this sieve, that the non-cuspidal real quadratic points on these curves have j-invariants belonging to $\mathbb{Q}$, and so their modularity follows from the modularity of elliptic curves over $\mathbb{Q}$.

The proof Theorem 1 is completed in Sect. 5.7 by proving Lemma 1.1. This requires the computation of $\mathbb{Q}(\sqrt{5})$-points on another four modular curves.

We end the paper by briefly mentioning some further implications of this work for modularity of elliptic curves over arbitrary totally real fields (Sect. 6.1), and also possible Diophantine applications, including the Fermat equation (Sect. 6.2).

Computational Details. Some of the computational details in Sects. 5.1–5.7 were carried out in the computer algebra system Magma [7]. The reader can find the Magma scripts for verifying these computations (together with extensive commentary) at: http://homepages.warwick.ac.uk/staff/S.Siksek/progs/modularity/

Our programs make use of several Magma packages. We would like to acknowledge these packages and their authors and main contributers:

- Hyperelliptic Curves: Nils Bruin, Brendan Creutz, Steve Donnelly, Michael Harrison, David Kohel, Michael Stoll, Paul van Wamelen;
- Small Modular Curves Database: Michael Harrison;
- Algebraic Function Fields: Florian Heß, Claus Fieker, Nicole Sutherland;
- Modular Forms: William Stein, Kevin Buzzard, Steve Donnelly;
- Modular Abelian Varieties: William Stein, Jordi Quer.

2 Background

2.1 Modularity lifting: Proof of Theorem 2

We assume the hypotheses of Theorem 2. Thus K is a totally real field, E is an elliptic curve over K, and $p \neq 2$ is a rational prime so that $\overline{\rho} := \overline{\rho}_{E,p}$ is modular, and $\overline{\rho}(G_{K(\zeta_p)})$ is absolutely irreducible.

Write $\rho : G_K \rightarrow \mathrm{GL}_2(\mathbb{Z}_p)$ for the representation arising from the action of G_K on the Tate module $T_p(E)$. For any finite place υ of K, we write $D_\upsilon \subset G_K$ for the corresponding decomposition group. We apply Théorème 3.2.2 of [10] to $\overline{\rho}$. That theorem involves making certain choices and we merely make the obvious ones as follows. We let S be the set of finite places υ of K of bad reduction for E together with the places above p. We let T be the subset of places above p where E has potentially ordinary reduction or potentially multiplicative reduction. In the notation of that theorem $\epsilon : G_K \rightarrow \mathbb{Z}_p^*$ is the p-adic cyclotomic character, and we choose $\psi = \epsilon$. We let $\rho_\upsilon = \rho|_{D_\upsilon}$. For a place υ we write $[r_\upsilon, N_\upsilon]$ for the Weil–Deligne type of ρ_υ. We note [54, Sects. 5.1, 5.2] that $N_\upsilon = 0$ if and only if E has potentially good reduction at

υ. If $p \neq 5$, the conditions of [10, Théorème 3.2.2] are satisfied; that theorem asserts existence of a global modular lift of $\overline{\rho}$ having the same local properties as ρ (properties (i)–(v) in the conclusion of the theorem). It also asserts that every global lift having these properties is modular, so ρ is modular. When $p = 5$, the theorem has an extra hypothesis: namely that the projective image of $\overline{\rho}|_{G_{K(\zeta_5)}}$ is not $\mathrm{PSL}_2(\mathbb{F}_5)$. The reason for this additional hypothesis is to assure the existence of Taylor–Wiles systems (as in [40, Sect. 3.2.3]). However, in our situation, because $\det \overline{\rho}$ is the mod 5 cyclotomic character, we can still choose Taylor–Wiles systems without this extra condition, as we now explain. In the notation of [40, Sect. 3.2.3] (except we change the name of the field to K) and following the proof of Theorem 2.49 in [18], what we need to show is that for each positive integer n and non-trivial cocycle $\psi \in \mathrm{H}^1(G_{K,S}, \mathrm{ad}^0\overline{\rho}(1))$, we can find a place $\upsilon \notin S$ of K such that

(i) $|k(\upsilon)| \equiv 1 \pmod{p^n}$, and $\overline{\rho}(\mathrm{Frob}_\upsilon)$ has distinct eigenvalues.
(ii) The image of ψ under the restriction map

$$\mathrm{H}^1(G_K, \mathrm{ad}^0\overline{\rho}(1)) \to \mathrm{H}^1(D_\upsilon, \mathrm{ad}^0\overline{\rho}(1))$$

is non-trivial.

Let K_m be the extension of $K(\zeta_{p^m})$ cut out by $\mathrm{ad}^0\overline{\rho}$. The argument in [18, Theorem 2.49] then works once we know the restriction of ψ to $\mathrm{H}^1(G_{K_0}, \mathrm{ad}^0\overline{\rho}(1))$ is non-trivial. To do this we want to show that $\mathrm{H}^1(\mathrm{Gal}(K_0/K), \mathrm{ad}^0\overline{\rho}(1)^{G_{K_0}}) = 0$. Because G_{K_0} acts trivially on $\mathrm{ad}^0\overline{\rho}$, the coefficient module vanishes unless $\zeta_p \in K_0$. Because $\mathrm{H}^1(\mathrm{PSL}_2(\mathbb{F}_5), \mathrm{Sym}^2 \mathbb{F}_5^2) = \mathbb{F}_5$ does not vanish, the extra condition when $p = 5$ was used to exclude the possibility that $\mathrm{Gal}(K_0/K) = \mathrm{PSL}_2(\mathbb{F}_5)$ and $\zeta_5 \in K_0$. But using the fact that $\overline{\rho}$ has cyclotomic determinant, we can get rid of it as follows. Suppose the above happens, then $\sqrt{5} \in K$ because $\det : \overline{\rho}(G_K) \to \mathrm{PGL}_2(\mathbb{F}_5) \to \mathbb{F}_5^*/(\mathbb{F}_5^*)^2$ is induced by the cyclotomic character. But $\zeta_5 \in K_0 \setminus K$ implies that $\mathrm{Gal}(K_0/K)$ has a quotient of order 2. As $\mathrm{Gal}(K_0/K) \cong \mathrm{PSL}_2(\mathbb{F}_5) \cong A_5$, we have a contradiction.

2.2 Background on modular curves

In order to apply Theorem 2, we need a better understanding of its condition (ii). We will show that if condition (ii) fails, then the image of $\overline{\rho}_{E,p}$ lies either in a Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$, or in certain subgroups contained in the normalizers of Cartan subgroups. In this case, the elliptic curve E/K gives rise to a non-cuspidal K-point on a modular curve corresponding to one of these subgroups. We now briefly sketch the relationship between certain subgroups of $\mathrm{GL}_2(\mathbb{F}_p)$ and the corresponding modular curves. For more details we recommend the articles by Mazur [46] and Rohrlich [55].

Let p be a prime, and H a subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ satisfying the condition $-I \in H$. Write $H_0 = H \cap \mathrm{SL}_2(\mathbb{F}_p)$ and denote by $\tilde{H}$ the preimage of H_0 via the natural map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{F}_p)$. Then $\tilde{H}$ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ of level p and we define the modular curve corresponding to H by

$$X(H) := \tilde{H} \backslash \mathfrak{H}^*, \qquad \mathfrak{H}^* := \mathfrak{H} \cup \mathbb{Q} \cup \{\infty\}$$

where $\mathfrak{H}$ is upper half-plane. The above construction gives $X(H)$ as compact Riemann surface, or equivalently as an algebraic curve over $\mathbb{C}$. It turns out however that $X(H)$ is defined over the field $\mathbb{Q}(\zeta_p)^{\det(H)}$; here we view $\det(H)$ as subgroup of $\mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q})$ via the cyclotomic character $\chi_p : \mathrm{Gal}(\mathbb{Q}(\zeta_p)/\mathbb{Q}) \xrightarrow{\cong} \mathbb{F}_p^*$. We shall assume henceforth that $\det(H) = \mathbb{F}_p^*$, and so $X(H)$ is defined over $\mathbb{Q}$. The curve $X(H)$ admits a natural map

$$j : X(H) \to X(1), \qquad (X(1) := \mathrm{SL}_2(\mathbb{Z}) \backslash \mathfrak{H}^* \cong \mathbb{P}^1)$$

corresponding to the inclusion $\tilde{H} \subset \mathrm{SL}_2(\mathbb{Z})$. The map j is defined over $\mathbb{Q}$, and is ramified only above 0, 1728 and ∞. We refer to the points of the fibre $j^{-1}(\infty) \subset X(H)(\overline{\mathbb{Q}})$ as the cusps of $X(H)$. For a number field K, the non-cuspidal K-points of $X(H)$ correspond to $\overline{\mathbb{Q}}$-isomorphism classes of pairs $(E, \eta \circ H)$ of elliptic curves E/K with an H-orbit of isomorphisms

$$\eta : \mathbb{F}_p^2 \xrightarrow{\cong} E[p](\overline{\mathbb{Q}})$$

such that $\eta^{-1}\sigma\eta \in H$ for all $\sigma \in G_K$ (note that H acts on the set of such η by its action on the domain of η). An elliptic curve E over K supports a non-cuspidal point on $X(H)$ if and only if the image $\overline{\rho}_{E,p}(G_K)$ is conjugate to a subgroup of H. In general, it is possible for one E to support distinct points of $X(H)$. In fact, picking an η, the number of points supported on E is the number of G_K-fixed points of the double coset space

$$\eta^{-1} \cdot \mathrm{Aut}_{\overline{\mathbb{Q}}}(E) \cdot \eta \backslash \mathrm{GL}_2(\mathbb{F}_p)/H,$$

where the action of G_K depends on the chosen η. Moreover if E/K gives rise to the K-point P on $X(H)$ then $j(P)$ is the j-invariant of E. If two elliptic curves E_1, E_2 defined over K, and having j-invariants $\neq 0$, 1728 give rise to the same K-point of $X(H)$, then they have the same j-invariant and hence are quadratic twists of each other.

2.2.1 Borel subgroups, cartan subgroups, and normalizers

Let $p \geq 3$ be a prime. Let $B(p)$ denote the Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$. Let $C_s(p)$ and $C_{ns}(p)$ respectively denote the split and non-split Cartan subgroups

of $GL_2(\mathbb{F}_p)$, and let $C_s^+(p)$ and $C_{ns}^+(p)$ respectively be their normalizers. The three groups $B(p)$, $C_s^+(p)$ and $C_{ns}^+(p)$ contain $-I$ and have determinants equal to $\mathbb{F}_p^*$. Thus they correspond to modular curves defined over $\mathbb{Q}$. The usual notation for these modular curves is $X_0(p)$, $X_{split}(p)$ and $X_{nonsplit}(p)$. We shall find it convenient to denote these modular curves by the following non-standard notation: $X(bp)$, $X(sp)$ and $X(nsp)$, so for example $X_0(5)$, $X_{split}(5)$ and $X_{nonsplit}(5)$ become $X(b5)$, $X(s5)$, $X(ns5)$.

2.2.2 Fibre products

Now let $p_1, \ldots, p_r \geq 3$ be distinct primes. Let $H_i \subseteq GL_2(\mathbb{F}_{p_i})$ be subgroups, all satisfying $-I \in H_i$ and $\det(H_i) = \mathbb{F}_{p_i}^*$. Let $\widetilde{H_i}$ be the corresponding congruence subgroups of $SL_2(\mathbb{Z})$ defined above. Then $\bigcap \widetilde{H_i}$ is a congruence subgroup of $SL_2(\mathbb{Z})$ of level $\prod p_i$. We define $X(H_1, \ldots, H_r)$ to be the modular curve

$$X(H_1, \ldots, H_r) := \left(\widetilde{H_1} \cap \ldots \cap \widetilde{H_r} \right) \backslash \mathfrak{H}^*.$$

This again is defined over $\mathbb{Q}$, and is the normalization of the fibre product

$$X(H_1) \times_{X(1)} X(H_2) \times_{X(1)} \ldots \times_{X(1)} X(H_r).$$

Again $X(H_1, \ldots, H_r)$ is naturally endowed with a j-map to $X(1)$ coming from the inclusion $\widetilde{H_1} \cap \ldots \cap \widetilde{H_r} \subseteq SL_2(\mathbb{Z})$. This agrees with the composition of the projection onto any of the $X(H_i)$ with $j : X(H_i) \to X(1)$. If K is a number field, non-cuspidal K-points correspond to (isomorphism classes of) pairs $(E, \eta \circ \prod H_i)$ of elliptic curves E/K with an $\prod H_i$-orbit of isomorphisms

$$\eta : \prod \mathbb{F}_{p_i}^2 \xrightarrow{\cong} E\left[\prod p_i \right](\overline{\mathbb{Q}})$$

such that $\eta^{-1} \overline{\rho}_{E, p_i}(G_K)\eta$ is contained in H_i, for $i = 1, \ldots, r$.

In the notation $X(H_1, \ldots, H_r)$, we shall replace the group H_i with bp, sp, nsp if it equals $B_0(p)$, $C_s^+(p)$ or $C_{ns}^+(p)$ respectively. Thus for example, $X(b5, b7)$ is isomorphic to $X(b5) \times_{X(1)} X(b7)$ and in standard notation can be denoted by $X_0(35)$.

2.3 The image $\overline{\rho}_{E, p}(G_{K(\zeta_p)})$

The following proposition is our first attempt at understanding condition (ii) of Theorem 2. It relates the failure of condition (ii) to K-points on the modular curves $X(bp)$, $X(sp)$ and $X(nsp)$.

Proposition 2.1 *Let E be an elliptic curve over a totally real field K. Let $p \geq 3$ be a rational prime, and write $\overline{\rho} = \overline{\rho}_{E,p}$. Then*

(i) $\overline{\rho}(G_{K(\zeta_p)}) = \overline{\rho}(G_K) \cap \mathrm{SL}_2(\mathbb{F}_p)$.
(ii) *If $\overline{\rho}(G_{K(\zeta_p)})$ is absolutely reducible, then $\overline{\rho}(G_K)$ is contained either in a Borel subgroup, or in the normalizer of a Cartan subgroup. In this case E gives rise to a non-cuspidal K-point on $X(\mathrm{b}p)$, $X(\mathrm{s}p)$ or $X(\mathrm{ns}p)$.*

Proof The determinant of $\overline{\rho}$ is the mod p cyclotomic character

$$\chi_p : G_K \to \mathrm{Aut}(\mu_p) \cong \mathbb{F}_p^*,$$

where μ_p is the group of p-th roots of unity. The first part follows immediately.

Write $G = \overline{\rho}(G_K)$, and suppose G is not contained in a Borel subgroup. Then G is irreducible, and as K is totally real it is absolutely irreducible. The second part follows from Lemma 2.2 below. □

Lemma 2.2 *Let $p \geq 3$ be prime. Let G be an absolutely irreducible subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ such that $G \cap \mathrm{SL}_2(\mathbb{F}_p)$ is absolutely reducible. Then G is contained in the normalizer of a Cartan subgroup.*

If moreover $\det(G) = \mathbb{F}_p^*$, then $G \cap \mathrm{GL}_2^+(\mathbb{F}_p)$ is contained in a Cartan subgroup, where $\mathrm{GL}_2^+(\mathbb{F}_p)$ is the subgroup of $\mathrm{GL}_2(\mathbb{F}_p)$ consisting of matrices with square determinant.

Proof Write $G' = G \cap \mathrm{SL}_2(\mathbb{F}_p)$. If p divides the order of G then (e.g. [57, Sect. 2.1] or [70, Lemma 2]) either G contains $\mathrm{SL}_2(\mathbb{F}_p)$ or G is contained in a Borel subgroup. In the former case, G' contains $\mathrm{SL}_2(\mathbb{F}_p)$ and so is absolutely irreducible, and in the latter case, G is reducible. We may thus suppose that p does not divide the order of G.

Write H for the image of G in $\mathrm{PGL}_2(\mathbb{F}_p)$, and let n be its order. Let $\mathcal{M}$ be the set of points $P \in \mathbb{P}^1(\overline{\mathbb{F}}_p)$ that are fixed by at least one non-trivial element of H. Swinnerton-Dyer shows [70, pp. 13–14] that $\mathcal{M}$ is stable under the action of G, and that it can be written as the union of disjoint orbits $\mathcal{N}_1, \cdots, \mathcal{N}_\nu$, with the following exhaustive list of possibilities.

(a) H is cyclic, G is contained in a Cartan subgroup: $\nu = 2$, $|\mathcal{N}_1| = |\mathcal{N}_2| = 1$;
(b) H is dihedral, G is contained in the normalizer of a Cartan subgroup: $\nu = 3$, $n \geq 4$ even, $|\mathcal{N}_1| = |\mathcal{N}_2| = n/2$, $|\mathcal{N}_3| = 2$;
(c) H is isomorphic to A_4: $\nu = 3$, $n = 12$, $|\mathcal{N}_1| = 6$, $|\mathcal{N}_2| = |\mathcal{N}_3| = 4$;
(d) H is isomorphic to S_4: $\nu = 3$, $n = 24$, $|\mathcal{N}_1| = 12$, $|\mathcal{N}_2| = 8$, $|\mathcal{N}_3| = 6$;
(e) H is isomorphic to A_5: $\nu = 3$, $n = 60$, $|\mathcal{N}_1| = 30$, $|\mathcal{N}_2| = 20$, $|\mathcal{N}_3| = 12$.

In our situation, case (a) is easy to rule out as it forces G to be absolutely reducible—indeed, the unique elements of $\mathcal{N}_1$, $\mathcal{N}_2$ are fixed by G. We would like to rule out cases (c), (d), (e).

Observe that G' is the kernel of the restriction to G of det $: G \to \mathbb{F}_p^*$. Thus G' is normal in G with cyclic quotient. If G' consists only of scalar matrices, then H is a quotient of G/G' and so is cyclic, which contradicts possibilities (b)–(e). Thus G' contains some non-scalar matrix g. The image of g in H is non-trivial.

As G' is absolutely reducible, there is some $P \in \mathbb{P}^1(\overline{\mathbb{F}}_p)$ which is fixed by G'. In particular, P belongs to one of the orbits $\mathcal{N}_i$, which we denote by $\mathcal{N}$. If Q is some other element of $\mathcal{N}$ then its stabilizer in G is conjugate to the stabilizer of P. However, G' is normal in G and is contained in the stabilizer of P. Thus G' stabilizes Q too. It follows that every element Q of $\mathcal{N}$ is an eigenvector of g. However, g is a non-scalar and p does not divide its order (as $g \in G$). Thus g has precisely two eigenvectors in $\mathbb{P}^1(\overline{\mathbb{F}}_p)$. Hence $|\mathcal{N}| \leq 2$. From the above list of cases (b)–(e), we see that we are in case (b), and $|\mathcal{N}| = 2$. This proves the first part of the lemma.

For the second part, suppose that $\det(G) = \mathbb{F}_p^*$. Let G^+ be the stabilizer of P, which by the above contains G'. Since G acts transitively on $\mathcal{N}$, which has size 2, the subgroup G^+ must have index 2. From the isomorphism det $: G/G' \to \mathbb{F}_p^*$ we see that $\det(G^+/G)$ consists of the squares in $\mathbb{F}_p^*$. Hence $G^+ = G \cap \mathrm{GL}_2^+(\mathbb{F}_p)$. Now G^+ stabilizes P, and by the argument above also stabilizes the other element of $\mathcal{N}$. Thus G^+ is contained in a Cartan subgroup. □

2.4 Langlands–Tunnell: Proof of Theorem 3 for $p = 3$

We shall need the following consequence of the well-known Langlands–Tunnell Theorem.

Theorem 6 *(Langlands–Tunnell) Let E be an elliptic curve over a totally real number field K and suppose $\overline{\rho}_{E,3}$ is irreducible. Then $\overline{\rho}_{E,3}$ is modular.*

The deduction of Theorem 6 from the Langlands–Tunnell Theorem [41,74] is well-known; a proof is given by Dieulefait and Freitas in [21, Lemma 4.2] and follows the lines of the argument by Wiles [75, Chap. 5] for classical modular forms.

The following is an immediate corollary of Langlands–Tunnell and Theorem 2. It is of course Theorem 3 with $p = 3$.

Corollary 2.3 *If E is an elliptic curve over a totally real field K and $\overline{\rho}_{E,3}(G_{K(\zeta_3)})$ is absolutely irreducible, then E is modular.*

3 Modularity switching

3.1 3–5 modularity switching: Proof of Theorem 3 for $p = 5$

In this section we prove Theorem 3 for $p = 5$. This proof follows from
Wiles' original 3–5 modularity switching argument over $\mathbb{Q}$ [75, Proof of The-
orem 5.2]. For the convenience of the reader and as a warm-up for the more
elaborate switching argument in the next section, we include a sketch.

Lemma 3.1 *Let E be an elliptic curve over a number field K. Then there is
an elliptic curve E'/K such that*

(i) $\overline{\rho}_{E,5} \sim \overline{\rho}_{E',5}$;
(ii) $\overline{\rho}_{E',3}(G_K)$ *contains* $\mathrm{SL}_2(\mathbb{F}_3)$.

We first show how the lemma implies Theorem 3 for $p = 5$.

Proof of Theorem 3 for $p = 5$ Let E be an elliptic curve over a totally real
number field K and let E' be as in the statement of Lemma 3.1. It follows from
Corollary 2.3 that E' is modular. Thus $\overline{\rho}_{E,5} \sim \overline{\rho}_{E',5}$ is modular. Applying
Theorem 2 to E with $p = 5$, we see that E is modular.   □

Let us now prove Lemma 3.1. Let K be a number field and E/K is an
elliptic curve. Let us denote by $X_E(5)$ the twisted modular curve over K which
classifies elliptic curves E' with a symplectic isomorphism of group schemes
$E[5] \to E'[5]$. Here the adjective **symplectic** means that the isomorphism is
compatible with the Weil pairings on $E[5]$ and $E'[5]$. We observe the following
properties

- $X_E(5) \cong \mathbb{P}^1$ over K. This is because $X_E(5)$ isomorphic (over $\overline{K}$) to the
  standard modular curve $X(5)$, and hence is a smooth curve of genus 0 over
  K, possessing by construction a K-rational point (coming from E itself).
- There is bijection between the non-cuspidal K-points on $X_E(5)$ and classes
  of pairs (E', u) where E' is an elliptic curve over K and u is a symplectic
  isomorphism $E[5] \to E'[5]$ of G_K-modules. Two such pairs (E'_1, u_1) and
  (E'_2, u_2) belong to the same class if there is a K-isomorphism $\phi : E'_1 \to E'_2$
  such that $u_2 = \phi \circ u_1$.

Thus a non-cuspidal K-point of $X_E(5)$ gives an elliptic curve E' satis-
fying condition (i), and we want to find such a point where (ii) holds. The
maximal subgroups of $\mathrm{GL}_2(\mathbb{F}_3)$ are $\mathrm{SL}_2(\mathbb{F}_3)$, the Borel subgroups, and the
normalizers of non-split Cartan subgroups[1]. If the image of $\overline{\rho}_{E',3}$ does not
contain $\mathrm{SL}_2(\mathbb{F}_3)$ then E' gives rise to a K-point of $X_E(5) \times_{X(1)} X(\mathrm{b}3)$ or

---

[1] In $\mathrm{GL}_2(\mathbb{F}_3)$ the normalizer of a split Cartan subgroup is contained as an index 2 subgroup in
the normalizer of a non-split Cartan subgroup, as the latter is a 2-Sylow subgroup.

$X_E(5) \times_{X(1)} X(\text{ns}3)$. Over $\mathbb{C}$, the normalizations of these two curves are isomorphic to the modular curves $X(5, \text{b}3)$ and $X(5, \text{ns}3)$, and hence are irreducible. It follows from Hilbert's irreducibility theorem [59, Theorem 3.4.1] that there is a non-cuspidal K-rational point of $X_E(5)$ which does not lift to a K-rational point of $X_E(5) \times_{X(1)} X(\text{b}3)$, $X_E(5) \times_{X(1)} X(\text{ns}3)$. This point provides the required E'.

3.2 3–7 modularity switching: Proof of Theorem 4

In this section we prove Theorem 4. The proof resembles Wiles' 3–5 modularity switching argument sketched in the previous section. In fact, similar arguments have been used to prove special cases of Serre's modularity conjecture for $\mathrm{GL}_2(\mathbb{F}_q)$ for $q = 5$, 7 and 9 by Taylor [72], Manoharmayum [44,45] and Ellenberg [22]. Our argument in this section is closest in spirit to Manoharmayum's; however we benefit from having at our disposal a more powerful modularity lifting theorem, allowing us to work in greater generality. We shall need the following lemma.

Lemma 3.2 *Let E be an elliptic curve over a totally real number field K. Then there is an extension L/K and an elliptic curve E'/L such that*

(i) *L is totally real;*
(ii) *$[L : K] = 4$;*
(iii) *$L \cap K(E[7]) = K$;*
(iv) *$\overline{\rho}_{E,7}|_{G_L} \sim \overline{\rho}_{E',7}$;*
(v) *$\overline{\rho}_{E',3}(G_L)$ contains $\mathrm{SL}_2(\mathbb{F}_3)$.*

Before proving Lemma 3.2 we use it prove Theorem 4.

Proof of Theorem 4 Let E be an elliptic curve over a totally real number field K and let L and E' be as in the statement of Lemma 3.2. It follows from Corollary 2.3 that E' is modular. Thus $\overline{\rho}_{E',7}$ is modular. Hence $\overline{\rho}_{E,7}|_{G_L}$ is modular.

From now on, we shall only be concerned with the elliptic curve E, and will write $\overline{\rho} = \overline{\rho}_{E,7}$. We will shortly show that $\overline{\rho}(G_{L(\zeta_7)})$ is absolutely irreducible. Thus we may apply Theorem 2 to deduce that E/L is modular. Now by a repeated application of cyclic base change results of Langlands [41], it follows that E/K is modular.

All that remains to show is that $\overline{\rho}_{E,7}(G_{L(\zeta_7)})$ is absolutely irreducible. It is here that we make use of the hypothesis in Theorem 4 that $\overline{\rho}(G_{K(\zeta_7)})$ is absolutely irreducible. By Proposition 2.1,

$$\overline{\rho}(G_{K(\zeta_7)}) = \overline{\rho}(G_K) \cap \mathrm{SL}_2(\mathbb{F}_7), \qquad \overline{\rho}(G_{L(\zeta_7)}) = \overline{\rho}(G_L) \cap \mathrm{SL}_2(\mathbb{F}_7).$$

To complete the proof it is enough to show that $\overline{\rho}(G_K) = \overline{\rho}(G_L)$. The former is the Galois group of $K(E[7])/K$ and the latter is the Galois group of $L(E[7])/L$. Since $L \cap K(E[7]) = K$ by part (iii) of Lemma 3.2 we see that these groups are the same.   □

We now turn our attention to proving Lemma 3.2. As before K is a totally real number field and E/K is an elliptic curve. We shall need to work with the modular curve $X_E(7)$. For detailed expositions of the construction and properties of $X_E(7)$ we recommend the paper of Halberstadt and Kraus [32], and that of Poonen et al. [51]. For our needs, the following basic facts found in the above references are sufficient.

- $X_E(7)$ is a smooth plane quartic curve defined over K.
- $X_E(7)$ is a twist of the Klein quartic

$$X(7) \ : \ x^3 y + y^3 z + z^3 x = 0. \tag{4}$$

- For any extension L/K, there is bijection between the non-cuspidal L-points on $X_E(7)$ and classes of pairs (E', u) where E' is an elliptic curve over L and u is a symplectic isomorphism $E[7] \to E'[7]$ of G_L-modules. Two such pairs (E_1', u_1) and (E_2', u_2) belong to the same class if there is an L-isomorphism $\phi : E_1' \to E_2'$ such that $u_2 = \phi \circ u_1$.

An elliptic curve E' as in the statement of Lemma 3.2 corresponds to a point on $X_E(7)$ over a certain totally real quartic extension L/K satisfying some additional hypotheses. We shall generate points on $X_E(7)$ defined over quartic extensions by taking the intersection of $X_E(7)$ with suitable K-lines. Write $\check{\mathbb{P}}^2$ for the dual space of $\mathbb{P}^2$—this is the parameter space of lines in $\mathbb{P}^2$, and is in fact isomorphic to $\mathbb{P}^2$. To generate a point on $X_E(7)$ that satisfies (ii)–(v) we will choose a line $\ell \in \check{\mathbb{P}}^2(K)$ that avoids certain thin subsets of $\check{\mathbb{P}}^2(K)$. To satisfy (i), in addition to avoiding these thin subsets, we want our line to belong to a certain open subset of $\check{\mathbb{P}}^2(K_\sigma)$ for each embedding $\sigma : K \hookrightarrow \mathbb{R}$.

It is helpful to recall the notion of thin subsets [58, Chap. 9], [59, Chap. 3]. Let $n \geq 1$. By a **thin subset of $\mathbb{P}^n(K)$ of type I** we mean a subset contained in $V(K)$ where V is a proper Zariski-closed K-subvariety of $\mathbb{P}^n$. By a **thin subset of $\mathbb{P}^n(K)$ of type II** we mean a subset contained in $\pi(V(K))$ where $\pi : V \to \mathbb{P}^n$ is a generically surjective K-morphism of degree ≥ 2 from an irreducible variety V of dimension $\dim(V) = n$. A **thin subset of $\mathbb{P}^n(K)$** is a subset contained in the finite union of thin subsets of types I and II. Hilbert's Irreducibility Theorem [59, Theorem 3.4.1] asserts that a thin subset cannot equal $\mathbb{P}^n(K)$.

Proof of Lemma 3.2 Fix a totally real field K and an elliptic curve E/K. Write $X = X_E(7) \subset \mathbb{P}^2$. We shall write $X^{(n)}$ for the n-th symmetric power of X (see

for example [49, Sect. 2.2]). This is a smooth n-dimensional variety over K, whose K-points correspond to effective degree n divisors on X that are stable under the action of G_K. Let ϕ be the morphism

$$\phi : \check{\mathbb{P}}^2 \hookrightarrow X^{(4)}, \qquad \ell \mapsto \ell \cdot X.$$

We define certain thin subsets $S_1, \ldots, S_6$ of $\check{\mathbb{P}}^2(K)$ and choose the line to lie outside these subsets. We let S_1 be the set of lines $\ell \in \check{\mathbb{P}}^2(K)$ passing through any point belonging to either of these two finite sets:

(a) $\overline{K}$-cusps of X_E;
(b) $X_E(K(E[7]))$ (which is certainly a finite set by Faltings' Theorem [24]).

The set S_1 is plainly a union of thin subsets of type I. Let S_2 be the set of lines $\ell \in \check{\mathbb{P}}^2(K)$ that are tangent to X. These belong to the dual curve of X, and hence S_2 is a thin subset of type I. Now any line belonging to $\check{\mathbb{P}}^2(K) \setminus S_1 \cup S_2$ intersects X either in an irreducible divisor of degree 4, or in the sum of two irreducible divisors of degree 2.

Write ϵ for the natural map

$$\epsilon : X^{(2)} \times X^{(2)} \to X^{(4)}, \qquad (D_1, D_2) \mapsto D_1 + D_2.$$

We let S_3 be the set of lines $\ell \in \check{\mathbb{P}}^2(K)$ so that $\phi(\ell)$ belongs to $\epsilon(X^{(2)}(K) \times X^{(2)}(K))$. The set S_3 certainly contains all K-lines ℓ so that $\ell \cdot X$ is the sum of two irreducible divisors of degree 2. We would like to show that S_3 is a thin set. It is known for a curve C of genus at least 2 that $C^{(2)}(K)$ is finite, unless C is hyperelliptic or bielliptic [33, Corollary 3]. In our situation X is not hyperelliptic, but might be bielliptic [5, Sect. 4], [32, Théorème 8.1]. At any rate, $X^{(2)}$ is isomorphic to its image in the Jacobian of X under an Abel-Jacobi map. It follows from a theorem of Faltings [25] that

$$X^{(2)}(K) = \{D_1, \ldots, D_n\} \cup \{E_1(K), \ldots, E_m(K)\} \tag{5}$$

where the D_i are finitely many divisors of degree 2 and the E_i are finitely many elliptic curves defined over K lying on $X^{(2)}$. Now $\epsilon(X^{(2)}(K) \times X^{(2)}(K))$ consists of a union of isolated points $D_i + D_j$, the K-points of curves $D_i \times E_j$ and the K-points of surfaces $E_i \times E_j$. The former two categories will contribute to S_3 thin sets of type I. To show that S_3 is a thin set, it is enough to show for any i, j that there is no K-rational map $\psi : \check{\mathbb{P}}^2 \dashrightarrow E_i \times E_j$, birational onto its image, making the following diagram commute:

$$
\begin{array}{ccc}
 & & E_i \times E_j \\
 & \nearrow & \uparrow \\
\psi & & X^{(2)} \times X^{(2)} \\
 & & \downarrow \epsilon \\
\check{\mathbb{P}}^2 & \xrightarrow{\phi} & X^{(4)}
\end{array}
$$

Any $E_i \times E_j$ is an abelian variety and the image of any rational map $\check{\mathbb{P}}^2 \dashrightarrow E_i \times E_j$ must be constant [49, Corollary 3.9], therefore S_3 is thin.

Now if $\ell \in \check{\mathbb{P}}^2(K) \setminus \cup_{1 \leq i \leq 3} S_i$ then $\ell \cdot X$ is an irreducible effective divisor of degree 4, and so there is some quartic extension L/K and a point $P \in X(L)$ such that ℓ passes through P. Moreover, P is a non-cusp, and corresponds to a pair (E', u) where E' is an elliptic curve over L and $u : E[7] \to E'[7]$ is a symplectic isomorphism of G_L-modules. This supplies an extension L and elliptic curve E' satisfying (ii) and (iv). To obtain condition (iii), we need to remove another thin set. When (iii) fails, $K(E[7]) \cap L$ is either equal to L or is a quadratic extension M/K contained in $K(E[7])$. The former is impossible by (b) above. To eliminate the latter possibility, note that there are at most finitely many $M \subset K(E[7])$ that are quadratic extensions of K, and for each of these we want our line to avoid the image of the natural map

$$
X^{(2)}(M) \to X^{(4)}(K), \qquad D \mapsto D + D^\sigma,
$$

where σ denotes conjugation. Of course we have a decomposition of $X^{(2)}(M)$ similar to (5), where the M replaces K, and the elliptic curves E_i are defined over M. By mimicking the above argument with $E_i \times E_i^\sigma$ replacing $E_i \times E_j$, we deduce the existence of a thin set S_4 such that for $\ell \in \check{\mathbb{P}}^2$ avoiding $S_1, \ldots, S_4$, the conditions (ii), (iii) and (iv) hold.

We now look at condition (v). As before, we use the fact that the maximal subgroups of $GL_2(\mathbb{F}_3)$ are $SL_2(\mathbb{F}_3)$, the Borel subgroups, and the normalizers of non-split Cartan subgroups. If the image of $\overline{\rho}_{E',3}$ does not contain $SL_2(\mathbb{F}_3)$ then E' gives rise to an L-point on either of the geometrically irreducible curves Y_1 or Y_2, which are the normalizations of $X \times_{X(1)} X(b3)$ or $X \times_{X(1)} X(ns3)$ respectively. In particular, $\ell \cdot X$ belongs to the image of $Y^{(4)}(K) \to X^{(4)}(K)$ where $Y = Y_1$ or Y_2. By Lemma 3.3 below, such ℓ respectively belong to thin sets S_5, S_6. Choosing ℓ outside $S_1, \ldots, S_6$ yields an L and E' satisfying (ii)–(v).

Finally we want to show that we can choose ℓ so that L is totally real. Let $\sigma_1, \ldots, \sigma_n$ be the real places of K. Write $X_{\sigma_i}/\mathbb{R}$ for $X \times_K K_{\sigma_i}$. Then X_{σ_i} is a real twist of the Klein quartic (4). By Lemma 3.4 below, the X_{σ_i} are in fact

isomorphic to $X(7)$ over $\mathbb{R}$. The line $2y + z = 0$ meets the model given for $X(7)$ in (4) in four distinct real points. Thus there are non-empty open subsets $U_i \subset \check{\mathbb{P}}^2(K_{\sigma_i})$ such that any $\ell \in U_i$ meets X_{σ_i} in four real points. Let

$$W = \check{\mathbb{P}}^2(K) \cap \prod_{i=1}^{n} U_i.$$

Now if $\ell \in W \setminus \cup S_i$ then the field L is totally real. All that remains is to show that W is not thin, and so the difference $W \setminus \cup S_i$ is non-empty. Note that as $\prod_{i=1}^{n} \check{\mathbb{P}}^2(\mathbb{R})$ is compact, it is covered by finitely many translates of $\prod_{i=1}^{n} U_i$ under $\prod_{i=1}^{n} \mathrm{PGL}_3(\mathbb{R})$. As $\mathrm{PGL}_3(K)$ is dense in $\prod_{i=1}^{n} \mathrm{PGL}_3(\mathbb{R})$ by weak approximation, we can assume the translates are given by elements in $\mathrm{PGL}_3(K)$. Thus $\check{\mathbb{P}}^2(K)$ is the union of finitely many translates of W by the action of $\mathrm{PGL}_3(K)$. By Hilbert's irreducibility, W is not thin and so $W \setminus \cup S_i$ is non-empty. $\qquad\square$

Lemma 3.3 *Let $X \subset \mathbb{P}^2$ be a smooth quartic curve over a number field K and $\mu : Y \to X$ be K-morphism of degree at least 2. Write μ also for the induced map $Y^{(4)} \to X^{(4)}$. Let $\phi : \check{\mathbb{P}}^2 \hookrightarrow X^{(4)}$ be the map sending a line ℓ to the zero-cycle $\ell \cdot X$. Then the induced map $\mu^{-1}(\phi(\check{\mathbb{P}}^2)) \to \phi(\check{\mathbb{P}}^2)$ does not have a rational section.*

Proof By Riemann–Hurwitz, the genus of Y is at least 5. We argue by contradiction. Suppose the map has a rational section. Then we have a commutative triangle,

$$
\begin{array}{ccc}
 & & Y^{(4)} \\
 & \psi \nearrow & \downarrow \mu \\
\check{\mathbb{P}}^2 & \xrightarrow{\phi} & X^{(4)}
\end{array}
$$

where ψ is a K-rational map with $\check{\mathbb{P}}^2$ birational to its image. Choose $D \in Y^{(4)}(K)$ belonging to the image of $\check{\mathbb{P}}^2(K)$ under ψ. Note that $\psi(\check{\mathbb{P}}^2)$ is a rational surface and so its image in the Jacobian J_Y of Y under a suitable Abel-Jacobi map $\psi(\check{\mathbb{P}}^2) \subset Y^{(4)} \to J_Y$ must be a single point. Thus the closed points of $\psi(\check{\mathbb{P}}^2)$ (interpreted as divisors on Y) are linearly equivalent to D. It follows that $\mathbb{P}(L(D))$ has dimension at least 2, and so $l(D) \geq 3$. An easy application of Riemann–Roch shows that D is a special divisor (i.e. $l(K_Y - D) > 0$ where K_Y is the canonical divisor on Y). Clifford's Theorem [34, Theorem IV.5.4] states that

$$l(D) - 1 \leq \frac{\deg(D)}{2}, \tag{6}$$

and that if equality holds then D is either canonical, or Y is hyperelliptic. Since $\deg(D) = 4$, we see that $l(D) = 3$ and so the Clifford inequality (6) is in fact an equality. Thus D is either canonical, or Y is hyperelliptic. In the first case the genus of Y is $(\deg(D) + 2)/2 = 3$ giving a contradiction. In the second case, Y is a hyperelliptic curve dominating X, therefore X is also hyperelliptic by [33, Proposition 1], and we also have a contradiction.   □

The following lemma appears to be implicitly assumed in Manoharmayum's argument [44, p. 710], and is perhaps a known fact about the Klein quartic, although we have been unable to find it in the literature, and therefore give a proof.

Lemma 3.4 *The only real twist of the Klein quartic is itself.*

Proof Write $X = X(7)$ for the Klein quartic (4). Its real twists are classified by $\mathrm{H}^1(\mathrm{Gal}(\mathbb{C}/\mathbb{R}), \mathrm{Aut}(X))$, where $\mathrm{Aut}(X)$ is the automorphism group of $X/\mathbb{C}$. The group $\mathrm{Aut}(X)$ is the unique simple group of order 168. Generators for $\mathrm{Aut}(X)$ as a subgroup of $\mathrm{PGL}_3(\mathbb{C})$ were given by Klein, and are reproduced by Elkies in [23, p. 54]. We used these to explicitly write down the semidirect product

$$\mathrm{Gal}(\mathbb{C}/\mathbb{R}) \ltimes \mathrm{Aut}(X).$$

There are 28 complements of $\mathrm{Aut}(X)$ in this semidirect product and we verified that they form a single conjugacy class, therefore $\mathrm{H}^1(\mathrm{Gal}(\mathbb{C}/\mathbb{R}), \mathrm{Aut}(X)) = 0$.   □

3.3 On modularity over totally real fields: Proof of Theorem 5

Let K be a totally real field and E an elliptic curve over K. By Theorems 3 and 4 we know that E is modular except possibly if $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is simultaneously absolutely reducible for $p = 3, 5, 7$. It follows from Proposition 2.1 that such E gives rise to a K-point on one of the 27 modular curves in (1).

These all have genera > 1, and so by Faltings' Theorem [25] at most finitely many K-points. Theorem 5 follows immediately.

Remark. By studying the gonalities of various modular curves, the second-named author [42] proves much stronger finiteness statements than Theorem 5.

4 A more precise study of images

4.1 The image $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ again

The following proposition gives a refinement of Proposition 2.1 for $p = 3$, 5, 7. Part (a) of the proposition (corresponding to $p = 3$) is in fact proved

by Rubin in [56, Proposition 6]. The group denoted by H in part (b) appears (implicitly) in a similar setting in [15, Proof of Lemma 7.2.3].

Proposition 4.1 *Let K be a totally real number field, and let E be an elliptic curve over K. Let p be a rational prime and suppose*

$$K \cap \mathbb{Q}(\zeta_p) = \mathbb{Q}.$$

Write $\overline{\rho} = \overline{\rho}_{E,p}$. Suppose $\overline{\rho}$ is irreducible but $\overline{\rho}(G_{K(\zeta_p)})$ is absolutely reducible.

(a) *If $p = 3$ then $\overline{\rho}(G_K)$ is conjugate to $C_s^+(3)$ (the normalizer of a split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_3)$); this has order 8, and its image in $\mathrm{PGL}_2(\mathbb{F}_3)$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$.*
(b) *If $p = 5$ then $\overline{\rho}(G_K)$ is conjugate in $\mathrm{GL}_2(\mathbb{F}_5)$ to the subgroup*

$$H = \left\langle \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \right\rangle$$

of order 16. The subgroup H is contained as a subgroup of index 2 in the normalizer of a split Cartan subgroup, and as a subgroup of index 3 in the normalizer of a non-split Cartan subgroup. The image of H in $\mathrm{PGL}_2(\mathbb{F}_5)$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^2$.
(c) *If $p = 7$ then $\overline{\rho}(G_K)$ is conjugate in $\mathrm{GL}_2(\mathbb{F}_7)$ to a subgroup of either H_1 or H_2 where*

$$H_1 = \left\langle \begin{pmatrix} 3 & 0 \\ 0 & 5 \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} \right\rangle, \qquad H_2 = \left\langle \begin{pmatrix} 0 & 5 \\ 3 & 0 \end{pmatrix}, \begin{pmatrix} 5 & 0 \\ 3 & 2 \end{pmatrix} \right\rangle.$$

The group H_1 has order 36 and is contained as a subgroup of index 2 in the normalizer of a split Cartan subgroup. The group H_2 has order 48 and is contained as a subgroup of index 2 in the normalizer of a non-split Cartan subgroup. The images of H_1 and H_2 in $\mathrm{PGL}_2(\mathbb{F}_7)$ are respectively isomorphic to $D_3 \cong S_3$ and D_4.

Proof Write $G = \overline{\rho}(G_K)$. We observe that G satisfies the following:

  (i) G is irreducible, and $G \cap \mathrm{SL}_2(\mathbb{F}_p)$ is absolutely reducible;
 (ii) G odd (in other words, there is some element $c \in G$ with eigenvalues 1 and -1);
(iii) $\det(G) = \mathbb{F}_p^*$.

Indeed, (i) follows by assumption and Proposition 2.1, (ii) is a consequence of K having a real embedding, and (iii) follows from $K \cap \mathbb{Q}(\zeta_p) = \mathbb{Q}$. One proof of the proposition is to enumerate all subgroups G of $\mathrm{GL}_2(\mathbb{F}_p)$ with

$p = 3, 5, 7$, and check that those satisfying (i), (ii), (iii), also satisfy (a), (b), (c). We can simplify the enumeration by using Lemma 2.2 and its proof, from which we know that $G \cap \mathrm{GL}_2^+(\mathbb{F}_p)$ is contained in a Cartan subgroup $C_*(p)$ and G is contained in its normalizer $C_*^+(p)$. If $p = 5$ then the element $c \in G$ that represents complex conjugation belongs to $\mathrm{GL}_2^+(\mathbb{F}_p)$ and therefore to $G \cap \mathrm{GL}_2^+(\mathbb{F}_p)$. Now an element of $C_{\mathrm{ns}}(5)$ has eigenvalues α, α^p for some $\alpha \in \mathbb{F}_{25}^*$. This shows that the non-split Cartan case does not arise for $p = 5$. □

Remark. It follows from the proof that $\overline{\rho}(G_L)$ is contained in a Cartan subgroup, where $L = K(\sqrt{(-1)^{(p-1)/2}p})$.

4.2 Modular curves again

Let H_1, H_2 be the two subgroups of $\mathrm{GL}_2(\mathbb{F}_7)$ specified in part (c) of Proposition 4.1. We shall denote the corresponding modular curves $X(H_1)$ and $X(H_2)$ by $X(\mathrm{d}7)$ and $X(\mathrm{e}7)$ respectively. We extend our earlier conventions regarding the naming of modular curves in an obvious way. Thus, for a number field K, non-cuspidal K-points on $X(\mathrm{s}3, \mathrm{b}5, \mathrm{d}7)$ give elliptic curves E/K such that the image of $\overline{\rho}_{E,3}$ is contained in a conjugate of the normalizer of a split Cartan subgroup of $\mathrm{GL}_2(\mathbb{F}_3)$, the image of $\overline{\rho}_{E,5}$ is contained in a conjugate of the Borel subgroup of $\mathrm{GL}_2(\mathbb{F}_5)$, and the image of $\overline{\rho}_{E,7}$ is contained in a conjugate of the group H_1.

Corollary 4.2 *Let K be a real quadratic field, and let E be an elliptic curve defined over K.*

(a) *If $\overline{\rho}_{E,3}(G_{K(\zeta_3)})$ is absolutely reducible then E gives rise to a non-cuspidal K-point on either $X(\mathrm{b}3)$ or $X(\mathrm{s}3)$.*
(b) *If $K \neq \mathbb{Q}(\sqrt{5})$ and $\overline{\rho}_{E,5}(G_{K(\zeta_5)})$ is absolutely reducible then E gives rise to a non-cuspidal K-point on either $X(\mathrm{b}5)$ or $X(\mathrm{s}5)$.*
(c) *If $\overline{\rho}_{E,7}(G_{K(\zeta_7)})$ is absolutely reducible then E gives rise to a non-cuspidal K-point on either $X(\mathrm{b}7)$ or $X(\mathrm{d}7)$ or $X(\mathrm{e}7)$.*

Proof The field K is real quadratic, so for $p = 3, 7$, we have $K \cap \mathbb{Q}(\zeta_p) = \mathbb{Q}$. For $p = 5$ this holds once we have imposed the condition $K \neq \mathbb{Q}(\sqrt{5})$. The corollary follows from Proposition 4.1.   □

  Remarks

(i) In (b) we could have replaced $X(\mathrm{s}5)$ with the curve $X(H)$, where H is as given in the statement of Proposition 4.1. In fact $X(H)$ is a double cover of $X(\mathrm{s}5)$, so this would give a more precise statement. However, for our later purpose of proving modularity of elliptic curves over real quadratic fields, we have found it sufficient and simpler to work with $X(\mathrm{s}5)$.

(ii) Parts (a), (b) improve over Proposition 2.1 by eliminating $X(\text{ns3})$ and $X(\text{ns5})$. This substantially reduces the amount of work we have to do later in enumerating quadratic points on modular curves.

(iii) In (b) we really need the condition $K \neq \mathbb{Q}(\sqrt{5})$. To see this let $K = \mathbb{Q}(\sqrt{5})$. Let $G = \overline{\rho}(G_K)$ where $\overline{\rho} = \overline{\rho}_{E,5}$ and E/K is an elliptic curve. Then

$$\det(G) = \chi_5 \left(\text{Gal}(\mathbb{Q}(\zeta_5)/\mathbb{Q}(\sqrt{5})) \right) = \{\overline{1}, \overline{4}\} \subset \mathbb{F}_5^*.$$

Now, in the proof of Proposition 4.1, if we replace the condition $\det(G) = \mathbb{F}_5^*$ with $\det(G) = \{\overline{1}, \overline{4}\}$ we obtain three subgroups of $\text{GL}_2(\mathbb{F}_5)$ of orders 6, 8 and 12. The subgroup of order 8 is contained in the normalizer of a split Cartan subgroup, but not in the normalizer of any non-split Cartan subgroup. The subgroups of orders 6 and 12 are contained in the normalizer of a non-split Cartan subgroup but not in the normalizer of any split Cartan subgroup.

4.3 Theorem 1 follows from Lemmas 1.1 and 1.2

Most of the rest of this paper is devoted to the proof of Lemma 1.2, which states that non-cuspidal real quadratic points on the modular curves in (2) and (3) are modular. Let X be any of the modular curves in (2), (3). Recall that X is a moduli space of elliptic curves with level structures. A non-cuspidal point on X gives rise to a well-defined j-invariant, and when $j \neq 0$, 1728 this determines the underlying elliptic curve up to quadratic twists. When we say that a point on X is modular, we mean that one (and hence any) elliptic curve supporting it is modular. Lemma 1.2 is in fact a summary of results established separately in the following Lemmas: 5.1, 5.2, 5.3 and 5.9.

Lemma 1.1 asserts modularity of elliptic curves over $\mathbb{Q}(\sqrt{5})$. Its proof is relegated to Sect. 5.7. We now explain how Lemmas 1.1 and 1.2 are enough to imply the modularity of all elliptic curves over real quadratic fields.

Proof of Theorem 1 Let E be an elliptic curve over a real quadratic field K. We would like to show that E is modular. By Lemma 1.1 we may suppose that $K \neq \mathbb{Q}(\sqrt{5})$. We know from Theorems 3 and 4 that if $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is absolutely irreducible with $p = 3$ or 5 or 7 then E is modular. Thus we may restrict our attention to elliptic curves E such that $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is simultaneously absolutely reducible for $p = 3$, 5 and 7. Now appealing to Corollary 4.2 we see that E gives rise to a K-point on one of the twelve modular curves

$$X(\text{b3}, \text{b5}, \text{b7}), \quad X(\text{b3}, \text{b5}, \text{d7}), \quad X(\text{b3}, \text{b5}, \text{e7}),$$
$$X(\text{b3}, \text{s5}, \text{b7}), \quad X(\text{b3}, \text{s5}, \text{d7}), \quad X(\text{b3}, \text{s5}, \text{e7}),$$

$$X(\mathrm{s3}, \mathrm{b5}, \mathrm{b7}), \qquad X(\mathrm{s3}, \mathrm{b5}, \mathrm{d7}), \qquad X(\mathrm{s3}, \mathrm{b5}, \mathrm{e7}),$$
$$X(\mathrm{s3}, \mathrm{s5}, \mathrm{b7}), \qquad X(\mathrm{s3}, \mathrm{s5}, \mathrm{d7}), \qquad X(\mathrm{s3}, \mathrm{s5}, \mathrm{e7}).$$

It is clear that every one of these twelve covers one of the seven modular curves in Lemma 1.2, which completes the proof. □

The proof of Lemma 1.2 will involve the determination of all non-cuspidal real quadratic points on some of the seven modular curves. If these points correspond to CM elliptic curves or $\mathbb{Q}$-curves then they are modular (see the remark at the beginning of Sect. 5.1 for more on the latter). The following lemma will sometimes be useful in proving the modularity of the remaining points.

Lemma 4.3 *Let E be an elliptic curve over a totally real field K, with j-invariant j. Let $p = 3$ or 5 or 7, and suppose that the p-th division polynomial of E is irreducible. Let $\mathfrak{q}$ be a prime of K satisfying $\upsilon_{\mathfrak{q}}(j) < 0$ and $p \nmid \upsilon_{\mathfrak{q}}(j)$. Then E is modular.*

Proof We shall show that $\overline{\rho}_{E,p}(G_K)$ contains $\mathrm{SL}_2(\mathbb{F}_p)$, which is enough to imply modularity by Theorems 3 and 4.

By the theory of the Tate curve [63, Proposition 6.1], the two conditions $\upsilon_{\mathfrak{q}}(j) < 0$ and $p \nmid \upsilon_{\mathfrak{q}}(j)$ together imply that p divides the order of $\overline{\rho}_{E,p}(G_K)$. It follows from the classification of maximal subgroups of $\mathrm{GL}_2(\mathbb{F}_p)$ (see for example [70, Lemma 2]) that $\overline{\rho}_{E,p}(G_K)$ is either contained in a Borel subgroup or contains $\mathrm{SL}_2(\mathbb{F}_p)$. Since the p-th division polynomial is irreducible, $\overline{\rho}_{E,p}(G_K)$ is not contained in a Borel subgroup. □

Remarks (I) See [42] for an alternative study of modular curves and quadratic points on them related to elliptic curves with small residual image at 3, 5, and 7.

(II) It is natural to ask whether the quadratic points on the seven modular curves in Lemma 1.2 can determined by the methods of Mazur [47], Kamienny [37] and Merel [48]. Let X be any of these modular curves. To apply such methods to our problem we appear to require:

(a) an abelian variety $\mathcal{A}/\mathbb{Q}$ having Mordell–Weil rank 0, that is a quotient of the Jacobian J of X;
(b) a prime p and a morphism $X^{(2)} \to \mathcal{A}$ (defined over $\mathbb{Q}$) that is injective on some small p-adic neighbourhood $U \subset X^{(2)}(\mathbb{Q}_p)$ (where $X^{(2)}$ is the symmetric square of X);
(c) some argument that shows that $X^{(2)}(\mathbb{Q}) \subset U$.

Usually U is the p-adic residue class of $\{\infty, \infty\}$, and property (b) is a consequence of $X^{(2)} \to \mathcal{A}$ being a formal immersion at $\{\infty, \infty\}$ above p. For the modular curves that we consider, we have been unable to demonstrate (c)

without actually determining the quadratic points.[2] In fact, as we will see in due course, $X(\mathrm{b}5, \mathrm{b}7)$ and $X(\mathrm{s}3, \mathrm{s}5)$ have infinitely many quadratic points, so (b) and (c) cannot simultaneously hold. Likewise, $X(\mathrm{b}3, \mathrm{s}5)$ has many non-cuspidal quadratic points, suggesting that these methods are inapplicable.

5 Real quadratic points on certain modular curves

5.1 Real quadratic points on $X(\mathrm{b}5, \mathrm{b}7)$

In this section we parametrize the non-cuspidal real quadratic points on $X(\mathrm{b}5, \mathrm{b}7)$ (commonly known as $X_0(35)$), showing that they are all modular. In fact, we shall show that all except two correspond to $\mathbb{Q}$-curves. By a $\mathbb{Q}$-curve, we mean an elliptic curve defined over $\overline{\mathbb{Q}}$ that is isogenous to all its Galois conjugates. It is known that $\mathbb{Q}$-curves are modular. Indeed, Ribet [53, Corollary 6.2] had shown that modularity of $\mathbb{Q}$-curves is a consequence of Serre's modularity conjecture, which is now a theorem of Khare and Wintenberger [38,39].

Lemma 5.1 *Let $X = X(\mathrm{b}5, \mathrm{b}7)$. The non-cuspidal real quadratic points on X are modular.*

Proof The curve X is hyperelliptic of genus 3. Galbraith [30, Sect 4.4] derives the following model for X,

$$X \; : \; y^2 = (x^2 + x - 1)(x^6 - 5x^5 - 9x^3 - 5x - 1). \tag{7}$$

Write J for $J_0(35)$—the Jacobian of X. Using 2-descent [67] we find that $J(\mathbb{Q})$ has Mordell-Weil rank 0. As J has good reduction at 3, the map $J(\mathbb{Q}) \to J(\mathbb{F}_3)$ is injective. The group generated by differences of the four obvious rational points $(0, \pm 1)$, $\infty_\pm$ (where the latter are the points at infinity) surjects onto $J(\mathbb{F}_3) \cong \mathbb{Z}/24\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, and is therefore equal to $J(\mathbb{Q})$. We find

$$J(\mathbb{Q}) = \frac{\mathbb{Z}}{24\mathbb{Z}} \cdot \big[\infty_- - \infty_+\big] + \frac{\mathbb{Z}}{2\mathbb{Z}} \cdot \big[3(0, -1) - 3\infty_+\big].$$

Let P be a quadratic point of X, and let P^σ denote its conjugate. Then $P + P^\sigma - \infty_+ - \infty_-$ is a rational divisor of degree 0 and hence defines an element of $J(\mathbb{Q})$. It follows that $P + P^\sigma$ is linearly equivalent to $T + \infty_+ + \infty_-$ where T is a divisor of degree 0 representing one of the 48 elements of $J(\mathbb{Q})$. As X is hyperelliptic, for $[T] \neq 0$, the divisor $T + \infty_+ + \infty_-$ is linearly

---

[2] Indeed, the absence of (c) is the reason why we do not yet have a parametrization of quadratic points on the family $X(\mathrm{b}N)$, even though (a) and (b) are known [37] for prime $N > 71$, where $\mathcal{A}$ is the Eisenstein quotient of $J_0(N)$, and p is any prime $\neq 2, 3, 5$.

equivalent to at most one effective divisor of degree 2. Enumerating these 47 possibilities for $[T] \neq 0$, we find that all the effective degree 2 divisors they yield are sums of two rational points, except for $\left(\frac{-1\pm\sqrt{5}}{2}, 0\right)$. We would like to show that this pair of points is modular. By Ogg [52, Corollary 2], the hyperelliptic involution on $X = X(\text{b5}, \text{b7})$ is the Fricke involution w_{35}. The points $\left(\frac{-1\pm\sqrt{5}}{2}, 0\right)$ are fixed by w_{35} and so correspond to CM elliptic curves, which are therefore modular.

We now turn to the case $[T] = 0$, and so $P + P^\sigma$ is linearly equivalent to $\infty_+ + \infty_-$. Thus $P = (x, \pm\sqrt{f(x)})$ with $x \in \mathbb{Q}\backslash\{0\}$, where f is the polynomial on the right-hand side of (7). Suppose now that P is a quadratic point of the form $(x, \sqrt{f(x)})$ with $x \in \mathbb{Q}\backslash\{0\}$. The Galois conjugate P^σ of P is related to P by the hyperelliptic involution. Therefore the elliptic curve E represented by $P \in X$ is related to its Galois conjugate E^σ by a 35-isogeny, and hence is a $\mathbb{Q}$-curve. $\square$

5.2 Real quadratic points on $X(\text{b3}, \text{s5})$

Lemma 5.2 Let $X = X(\text{b3}, \text{s5})$. *All non-cuspidal real quadratic points on X are modular.*

Proof The curve X is studied briefly in [15, pp. 554–556] where it is shown that $X \cong X_0(75)/w_{25}$. That paper does not give a model for X but mentions a computation by Elkies where he finds X to have genus 3 and determines its rational points. It turns out there is precisely one non-cuspidal rational point and that this has j-invariant 0.

In order to write down a model for X, and to compute the Mordell–Weil group of its Jacobian, we shall make use of Stein's Magma implementation of modular symbols algorithms [16,66]. We find that an echelon basis for the $+1$-eigenspace of the Atkin–Lehner operator w_{25} acting on $S_2(\Gamma_0(75))$ is given by

$$
\begin{aligned}
f_1 &= q - q^4 - 2q^5 + q^6 + q^9 + 2q^{10} - 4q^{11} + \cdots \\
f_2 &= q^2 - q^4 + q^6 - q^7 - q^8 - 2q^{11} + \cdots \\
f_3 &= q^3 + q^4 + 2q^5 - q^6 + q^7 - 2q^8 - 2q^{10} + 2q^{11} + \cdots
\end{aligned}
$$

Thus X has genus 3; a basis for the holomorphic differentials is given by $f_i \, dq/q$. Following Galbraith [30, Chap. 3], we compute the image of this curve under the canonical map, and find that it is a plane quartic curve. It follows that the canonical map is indeed an embedding of X in $\mathbb{P}^2$. The equation given by the canonical embedding is

$$X \; : \; 3x^3z - 3x^2y^2 + 5x^2z^2 - 3xy^3 - 19xy^2z - xyz^2$$
$$+ 3xz^3 + 2y^4 + 7y^3z - 7y^2z^2 - 3yz^3 = 0. \tag{8}$$

We shall also need the map $j : X \to X(1)$. The details of finding this map explicitly are similar to those in a recent paper of Banwait and Cremona [2, Sect. 2.4], and so we only give a sketch. The isomorphism $X_0(75)/w_{25} \to X$ pulled back to the upper half-plane $\mathfrak{H}$ is given by $\tau \mapsto 5\tau$. Thus the q-expansion of j on X is given by

$$j(q^5) \; = \; \frac{1}{q^5} + 744 + 196884q^5 + 21493760q^{10}$$
$$+ 864299970q^{15} + 20245856256q^{20} + \cdots.$$

We can express $j = F(x, y, z)/G(x, y, z)$ where F, G are homogeneous forms of the same degree with coefficients in $\mathbb{Z}$. We seek such F, G so that the identity of modular forms

$$F(f_1, f_2, f_3) = j(q^5)G(f_1, f_2, f_3) \tag{9}$$

holds. To find suitable F, G, we run through the degrees $d = 1, 2, \ldots$ and express F, G as linear combinations of the monomials of degree d with unknown coefficients. By comparing the coefficients of the q-expansions in (9) we obtain simultaneous equations for the unknown coefficients of F, G. Solving these we obtain the first non-zero F, G with $d = 15$. For this particular F, G we checked the equality of coefficients on both sides of (9) beyond the Sturm bound [66, Theorem 9.18]. This is sufficient to show that $j = F(x, y, z)/G(x, y, z)$. The expressions for F and G are complicated and can be found in the accompanying Magma script (see p. 6).

Write J for the Jacobian of X. Computing the eigenbasis for the space spanned by f_1, f_2, f_3 yields three rational eigenforms corresponding to the elliptic curves E_1, E_2, E_3 with Cremona labels $15A1, 75B1, 75C1$. It follows that J is isogenous to $E_1 \times E_2 \times E_3$.

We use this to compute the rank of $J(\mathbb{Q})$ and of $J(\mathbb{Q}(\sqrt{5}))$ (we will see in due course why the latter rank is convenient to have). The three curves E_i have rank 0 over $\mathbb{Q}$ and so do their quadratic twists by by 5. Thus E_i have rank 0 over $\mathbb{Q}$ and over $\mathbb{Q}(\sqrt{5})$. It follows that J has rank 0 over those two fields. Thus $J(\mathbb{Q})$ is equal to its torsion subgroup. Let $L = \mathbb{Q}(\sqrt{5})$. We shall denote the torsion subgroups of $J(\mathbb{Q})$ and of $J(L)$ by $J(\mathbb{Q})_{\text{tors}}$ and $J(L)_{\text{tors}}$. To compute $J(\mathbb{Q})_{\text{tors}}$ we shall in fact compute $J(L)_{\text{tors}}$ and recover $J(\mathbb{Q})_{\text{tors}}$ using the inclusion $J(\mathbb{Q})_{\text{tors}} \subseteq J(L)_{\text{tors}}$. A short search for points on X defined over L reveals the following points:

$$P_0 = (0 : 0 : 1), \qquad P_1 = (1 : 0 : 0), \qquad P_2 = (1 : -1 : 2),$$
$$P_3 = \left(-2 : \frac{-1 + \sqrt{5}}{2} : 1\right), \quad P_4 = \left(\frac{1 - \sqrt{5}}{2} : -3 + \sqrt{5} : 1\right), \quad P_5 = \left(\frac{1 + \sqrt{5}}{2} : 2 + \sqrt{5} : 1\right).$$

As $J(L)$ is torsion, we see that $[P_i - P_0]$ belong to $J(L)_{\text{tors}}$. Write H for the subgroup of $J(L)_{\text{tors}}$ generated by the differences $[P_i - P_0]$, $[P_i^\sigma - P_0]$ where σ denotes conjugation in L. Write $\mathcal{O}_L$ for the ring of integers of L and let $\mathfrak{p} = 19\mathcal{O}_L + (9 + \sqrt{5})\mathcal{O}_L$. Then $\mathfrak{p}$ is a prime ideal of norm 19. Thus we have a natural injection

$$\pi_{\mathfrak{p}} : J(L)_{\text{tors}} \hookrightarrow J(\mathbb{F}_{19})$$

induced by $\mathfrak{p}$. We find that

$$J(\mathbb{F}_{19}) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/200\mathbb{Z},$$
$$\pi_{\mathfrak{p}}(H) \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z} \times \mathbb{Z}/40\mathbb{Z}.$$

It follows that either $J(L)_{\text{tors}} = H$ or H has index 5 in $J(L)_{\text{tors}}$. In the latter case, the order of $J(L)_{\text{tors}}$ is divisible by 25.

Let $\mathfrak{q} = 7\mathcal{O}_L$. This is a prime ideal of norm 49, and so we have another injection

$$\pi_{\mathfrak{q}} : J(L)_{\text{tors}} \hookrightarrow J(\mathbb{F}_{49}) \cong (\mathbb{Z}/8\mathbb{Z})^3 \times \mathbb{Z}/440\mathbb{Z}.$$

As the order of $J(\mathbb{F}_{49})$ is not divisible by 25, the order of $J(L)_{\text{tors}}$ is not divisible by 25. It follows that $J(L)_{\text{tors}} = H$. Taking $\text{Gal}(L/\mathbb{Q})$-invariants, we find that

$$J(\mathbb{Q}) = J(\mathbb{Q})_{\text{tors}} = (\mathbb{Z}/40\mathbb{Z}) \cdot [P_2 - P_0] + (\mathbb{Z}/2\mathbb{Z}) \cdot [P_3 + P_3^\sigma + 3P_0 - 5P_2].$$

We now enumerate the real quadratic points of X. As X is a plane quartic (and therefore not hyperelliptic), the map

$$X^{(2)}(\mathbb{Q}) \to J(\mathbb{Q}), \qquad D \mapsto [D - 2P_0]$$

is injective. Note that any element $D \in X^{(2)}(\mathbb{Q})$ is linearly equivalent to $D' + 2P_0$, where $[D'] \in J(\mathbb{Q})$. For each of the 80 elements $[D']$ of $J(\mathbb{Q})$ we computed the Riemann–Roch space of $D' + 2P_0$ and checked if this divisor is linearly equivalent to a positive divisor of degree 2 (in which case the positive divisor is unique). This allowed us to enumerate $X^{(2)}(\mathbb{Q})$ and hence all the rational and quadratic points on X. It turns out that

$$X(\mathbb{Q}) = \{P_0, P_1, P_2\}.$$

Table 1 The table gives the real quadratic points on the curve $X(\mathrm{b3}, \mathrm{s5})$ given by the model in (8)

Point	j-Invariant	Discriminant of quadratic order of CM
$(-2 + \sqrt{2} : \sqrt{2} : 2)$	$2417472 - 1707264\sqrt{2}$	-24
$(-1 : -2 + \sqrt{3} : 1)$	1728	-4
$(-2 : -1 + \sqrt{3} : 2)$	$76771008 + 44330496\sqrt{3}$	-36
$(1 + \sqrt{5} : -6 - 2\sqrt{5} : 2)$	0	-3
$(1 + \sqrt{5} : 4 + 2\sqrt{5} : 2)$	∞	Not applicable
$(-4 : -1 + \sqrt{5} : 2)$	∞	Not applicable
$(-10 + 2\sqrt{17} : 3 - \sqrt{17} : 4)$	$-2770550784 - 671956992\sqrt{17}$	-51
$(-3 + \sqrt{33} : 6 : 0)$	-32768	-11
$(-9 + \sqrt{33} : -9 + \sqrt{33} : 8)$	$-18808030478336 + 3274057859072\sqrt{33}$	-99
$(-8 : -11 + \sqrt{41} : 4)$	$(3674587331712451 - 573874126975281\sqrt{41})/65536$	Non-CM
$(-5 + \sqrt{41} : -2 + 2\sqrt{41} : 8)$	$(-251512901 - 39335769\sqrt{41})/64$	Non-CM
$(-23 + \sqrt{145} : -12 : 24)$	$(3494273 - 283495\sqrt{145})/64$	Non-CM
$(11 + \sqrt{145} : -11 - \sqrt{145} : 2)$	$(839560577 - 84563815\sqrt{145})/65536$	Non-CM
$(-91 + \sqrt{185} : 6 - 2\sqrt{185} : 88)$	$(572227476834375 - 42070984346475\sqrt{185})/65536$	Non-CM

It also gives the j-invariants of these points, and, where applicable, the discriminant of the order of complex multiplication of the corresponding elliptic curve

The points P_0 and P_1 are cusps; the point P_2 has j-invariant 0. This is consistent with Elkies' computation mentioned in [15, page 555]. The complex quadratic points are (up to conjugation)

$$(-5 + \sqrt{-11} : 0 : 6), \qquad (-6 + 2\sqrt{-15} : 1 + \sqrt{-15} : 8),$$

with j-invariants -32768 and 0 respectively. Table 1 gives the real quadratic points (up to conjugation), and their j-invariants; the ones with j-invariant $= \infty$ are of course cusps.

From the table we see that there are 12 non-cuspidal real quadratic points (up to conjugation). Of these, seven correspond to elliptic curves with complex multiplication and are therefore modular. For the remaining five modularity follows from Lemma 4.3 with $p = 7$ and $\mathfrak{q}$ some prime above 2. □

5.3 Real quadratic points on $X(s3, s5)$

Lemma 5.3 *Non-cuspidal real quadratic points on $X(s3, s5)$ are modular.*

Proof Let $X = X(s3, s5)$. As in the proof of Lemma 5.2, we use the canonical map to write down a model for $X \cong X_0(225)/\langle w_9, w_{25} \rangle$. The subspace of $S_2(\Gamma_0(225))$ invariant under w_9 and w_{25} has dimension 4 and echelon basis

$$
\begin{aligned}
f_1 &= q - 4q^5 - 4q^6 + 5q^7 - 4q^9 + 4q^{10} - 8q^{11} + \cdots, \\
f_2 &= q^2 - 2q^5 - 4q^6 + 4q^7 - q^8 - 2q^9 + 2q^{10} - 6q^{11} + \cdots, \\
f_3 &= q^3 - 2q^5 - 2q^6 + 2q^7 + q^8 - q^9 + 2q^{10} - 3q^{11} + \cdots, \\
f_4 &= q^4 - 2q^5 - 2q^6 + 5q^7 - 2q^9 + 2q^{10} - 4q^{11} + \cdots
\end{aligned}
$$

Thus X has genus 4. Instead of using the basis $f_1, \ldots, f_4$, we shall use the following basis:

$$
\begin{aligned}
g_1 &= -f_1 + 2f_4, \quad g_2 = f_1 - f_2 + 2f_3 - f_4, \\
g_3 &= f_1 + f_2 - 2f_3 - f_4, \quad g_4 = f_1 - f_3.
\end{aligned}
$$

The motivation behind this change of basis is that the automorphisms of the model for X given by the canonical map induced by $g_1, \ldots, g_4$ are easier to describe as we will soon see. The canonical map $X \to \mathbb{P}^3$ induced by $g_1, \ldots, g_4$ is an embedding and yields the following model:

$$
X \; : \; \begin{cases} 3x_1^2 - x_2^2 - 2x_3^2 + 2x_3x_4 - 3x_4^2 = 0 \\ x_2^2 x_3 - x_3^3 + 4x_3^2 x_4 - 12x_3 x_4^2 + 12x_4^3 = 0. \end{cases}
$$

Note that X has the following pair of involutions,

$$
\begin{aligned}
\alpha &: (x_1, x_2, x_3, x_4) \mapsto (-x_1, x_2, x_3, x_4), \\
\beta &: (x_1, x_2, x_3, x_4) \mapsto (x_1, -x_2, x_3, x_4).
\end{aligned}
$$

Write $\mathcal{A} := \{1, \alpha, \beta, \alpha\beta\}$. (It is possible to show that $\mathcal{A}$ is the automorphism group of X, but we shall not need this.) Write $\mathbb{K}$ for the function field of X. The subfield fixed by $\mathcal{A}$ is $\mathbb{Q}(w)$ where $w = x_3/x_4$. Since $\mathcal{A} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, it follows that $\mathbb{K} = \mathbb{Q}(w)(\sqrt{h_1}, \sqrt{h_2})$ for some $h_1, h_2 \in \mathbb{Q}(w)$. By inspection, we see that X has the following more convenient model:

$$
X \; : \; \begin{cases} u^2 = w(w-1)(w^2 - w + 4) \\ v^2 = w(w^3 - 4w^2 + 12w - 12), \end{cases} \tag{10}
$$

where

$$u = \frac{x_1 x_3}{x_4^2}, \qquad v = \frac{x_2 x_3}{x_4^2}, \qquad w = \frac{x_3}{x_4}.$$

We are interested in determining the real quadratic points of X, and for this we shall consider the following three obvious quotients

$$D_1 \; : \; u^2 = w(w-1)(w^2 - w + 4),$$
$$D_2 \; : \; v^2 = w(w^3 - 4w^2 + 12w - 12),$$
$$D_3 \; : \; z^2 = (w-1)(w^2 - w + 4)(w^3 - 4w^2 + 12w - 12).$$

We shall write

$$h_1 = w(w-1)(w^2 - w + 4), \qquad h_2 = w(w^3 - 4w^2 + 12w - 12),$$
$$h_3 = (w-1)(w^2 - w + 4)(w^3 - 4w^2 + 12w - 12).$$

The quotient maps $\phi_i : X \to D_i$ are all of degree 2, given by

$$\phi_1(w, u, v) = (w, u), \qquad \phi_2(w, u, v) = (w, v), \qquad \phi_3(w, u, v) = (w, uv/w).$$

The curve D_2 is isomorphic to the Weierstrass elliptic curve with Cremona label $225A1$, with Mordell–Weil rank 1 over $\mathbb{Q}$ (and trivial torsion). It follows that X has infinitely many quadratic points obtained by pulling back the rational points on D_2 via the degree 2 map ϕ_2. We claim that all non-cuspidal real quadratic points on X are obtained thus.

The curve D_3 has genus 2. Write J for its Jacobian. Using 2-descent we find that $J(\mathbb{Q})$ has Mordell–Weil rank 0, and by a similar calculation to those in the previous sections we find that

$$J(\mathbb{Q}) = \frac{\mathbb{Z}}{10\mathbb{Z}} \cdot \left[\left(\frac{1 + \sqrt{5}}{2}, \frac{-15 + 5\sqrt{5}}{2} \right) \right.$$
$$\left. + \left(\frac{1 - \sqrt{5}}{2}, \frac{-15 - 5\sqrt{5}}{2} \right) - \infty_+ - \infty_- \right].$$

Here ∞_+, ∞_- are the two points at infinity on D_3. As in the proof of Lemma 5.1 we can determine all rational and quadratic points on D_3. In particular,

$$D_3(\mathbb{Q}) = \{(1, 0), \infty_+, \infty_-\}.$$

The real quadratic points (w, z) on D_3 are, up to Galois conjugation and the action of the hyperelliptic involution, equal to $((1+\sqrt{5})/2, (-15+5\sqrt{5})/2)$, or satisfy $w \in \mathbb{Q}$ and $z = \pm\sqrt{h_3(w)} \notin \mathbb{Q}$.

Clearly a real quadratic point on X maps to either a rational or a real quadratic point on D_3. We deduce that the only real quadratic points on X satisfy

(a) either $(w, u, v) = ((1-\sqrt{5})/2, (-5-\sqrt{5})/2, \sqrt{5})$ or its conjugates under the action of Galois and the automorphisms $u \mapsto -u$, $v \mapsto -v$;

(b) or $w \in \mathbb{Q}$ and $uv/w = \pm\sqrt{h_3(w)} \notin \mathbb{Q}$.

The points in (a) satisfy $w^2 - w - 1 = 0$, and are cusps as can be deduced from the expression for j-map on X given below in (11).

We are interested in proving the modularity of the non-cuspidal real points $P = (w, u, v)$ satisfying (b). Write $d = h_3(w)$. Let $K = \mathbb{Q}(\sqrt{d})$, which is the field of the definition of the point P. If u is rational, then $\phi_1(P) = (w, u) \in D_1(\mathbb{Q})$. However, D_1 is isomorphic to the Weierstrass elliptic curve with Cremona label 15A8, with rank 0 and precisely four rational points. On the model D_1 the four points are the two points at infinity, and the points $(0, 0)$ and $(1, 0)$. None of these are images of real quadratic points on X. We can conclude that $u = \pm\sqrt{h_1(w)}$ is not rational. As $h_1(w) \in \mathbb{Q}$ and $u \in K = \mathbb{Q}(\sqrt{d})$ we have $h_1(w) = a^2 d$ for some non-zero $a \in \mathbb{Q}$. Thus

$$h_2(w) = \frac{w^2 h_3(w)}{h_1(w)} = \frac{w^2 d}{a^2 d} = \frac{w^2}{a^2}.$$

Therefore $v \in \mathbb{Q}$. Hence, if $P = (w, u, v)$ is a non-cuspidal real quadratic point on X then $\phi_2(P) = (w, v) \in D_2(\mathbb{Q})$, proving our claim.

To prove that the non-cuspidal real quadratic points on X are modular we need an expression for the j-map. The parametrization of X via $g_1, \ldots, g_4$ allows us to write down the j-map as a rational function on X using the same method as described in the proof of Lemma 5.2. On the model (10), the j-map is given by

$$j = \frac{(R_1 + R_2 \cdot v)^3}{8\, R_3^{15}} \tag{11}$$

where

$$R_1 = w^{15} - 10w^{14} + 45w^{13} - 115w^{12} + 155w^{11} - 54w^{10} + 185w^9 - 2395w^8 + 10465w^7$$
$$- 27055w^6 + 47072w^5 - 56090w^4 + 43405w^3 - 19485w^2 + 4020w - 132;$$
$$R_2 = w^{13} - 8w^{12} + 25w^{11} - 35w^{10} + 5w^9 + 94w^8 - 467w^7 + 1805w^6 - 5175w^5$$
$$+ 9685w^4 - 11216w^3 + 7698w^2 - 2715w + 315;$$
$$R_3 = w^2 - w - 1.$$

The expression for j involves only w and v. Thus the map $j : X \to X(1)$ factors via $\phi_2 : X \to D_2$. If P is a non-cuspidal real quadratic point on X then $\phi_2(P) \in D_2(\mathbb{Q})$, and so $j(P) \in \mathbb{Q}$. The modularity of P follows from the modularity of elliptic curves over the rationals [9]. □

Remark. The reader might be surprised by the fact—crucial to our proof—that the j-map (11) on X factors via the quotient D_2. Also surprising is the fact that the expression for j is a perfect cube. To explain this, we start by the observation that the normalizer $C_s^+(3)$ of a split Cartan subgroup in $\mathrm{GL}_2(\mathbb{F}_3)$ is conjugate to a subgroup of the normalizer $C_{ns}^+(3)$ of a non-split Cartan subgroup. Thus the map $X(s3) \to X(1)$ factors via $X(ns3)$. Moreover, $C_s^+(3) \cap \mathrm{SL}_2(\mathbb{F}_3)$ has index 2 in $C_{ns}^+(3) \cap \mathrm{SL}_2(\mathbb{F}_3)$. It follows that the map $X(s3) \to X(ns3)$ has degree 2. Thus the map $X \to X(1)$ factors via a degree 2 map $X \to X(ns3, s5)$. The curves $X(ns3)$ and $X(s5)$ have genus 0. In [14, Sect. 5.1], Chen derives the following expressions for the j-maps on $X(ns3)$ and $X(s5)$,

$$j = n^3, \qquad j = \frac{(s^2 - 5)^3 (s^2 + 5s + 10)^3 (s + 5)^3}{(s^2 + 5s + 5)^5},$$

where n and s are respectively Hauptmoduln on $X(ns3)$ and $X(s5)$. Since $X \to X(1)$ factors via $X(ns3)$, it is clear that j must be a cube on X, and this provides a useful check on our computation of j. A (singular) model for $X(ns3, s5)$ is given by

$$n^3 = \frac{(s^2 - 5)^3 (s^2 + 5s + 10)^3 (s + 5)^3}{(s^2 + 5s + 5)^5}.$$

It is easy to see that this is birational to the elliptic curve

$$E : s^2 + 5s + 5 = m^3.$$

The curve E has Cremona label 225A1, and is in fact isomorphic to D_2. This explains why $X \to X(1)$ factors through D_2.

5.4 Models for some modular curves of genus 1

We shall shortly study the quadratic points on the four modular curves in (3).

Lemma 5.4 *The four modular curves in* (3) *respectively have genus* 97, 153, 73 *and* 113.

The lemma follows from a standard application of Riemann–Hurwitz applied to $X \to X(1)$, where X runs through the four modular curves. For

this, we need the ramification indices of points above 0, 1728 and ∞, and we explain how to compute these in Sect. 5.5.1. We omit the details.

In view of the very great genera, it is unlikely that methods used to compute the quadratic points on the curves in (2) will succeed for the curves in (3). Indeed, it is entirely impractical to write down smooth projective models for any of these four curves. Instead, we shall view them as normalizations of the fibre products

$$X(\mathrm{b}3, \mathrm{b}5) \times_{X(1)} X(\mathrm{d}7), \qquad X(\mathrm{s}3, \mathrm{b}5) \times_{X(1)} X(\mathrm{d}7),$$
$$X(\mathrm{b}3, \mathrm{b}5) \times_{X(1)} X(\mathrm{e}7), \qquad X(\mathrm{s}3, \mathrm{b}5) \times_{X(1)} X(\mathrm{e}7).$$

We therefore write down models for

$$X(\mathrm{b}3, \mathrm{b}5), \qquad X(\mathrm{s}3, \mathrm{b}5), \qquad X(\mathrm{d}7), \qquad X(\mathrm{e}7) \tag{12}$$

and the corresponding j-maps to $X(1)$. We will also compute the Mordell–Weil groups of their Jacobians over $\mathbb{Q}$. We then use this information to study the quadratic points on the four curves in (3). The curves in (12) are all elliptic curves (genus 1 curves with $\mathbb{Q}$-points). Rather than identify elements of the Mordell–Weil groups with points on the curves (as is usual for elliptic curves), we shall view them as elements of Pic^0; we have found this to be more convenient for studying quadratic points.

5.4.1 The modular curves $X(\mathrm{d}7)$ and $X(\mathrm{e}7)$

Lemma 5.5 *A model for the curve $X(\mathrm{d}7)$ is given by*

$$X(\mathrm{d}7) \; : \; y^2 = -7(x^4 - 10x^3 + 27x^2 - 10x - 27).$$

The j-map $X(\mathrm{d}7) \to X(1)$ is given by

$$j = \frac{x(x+1)^3(x^2-5x+8)^3(x^2-5x+1)^3(x^4-5x^3+8x^2-7x+7)^3}{(x^3-4x^2+3x+1)^7}. \tag{13}$$

The $\mathbb{Q}$-points on $X(\mathrm{d}7)$ are $(5/2, \pm 7/4)$. The Mordell–Weil group over $\mathbb{Q}$ is given by

$$\frac{\mathbb{Z}}{2\mathbb{Z}} \cdot [(5/2, -7/4) - (5/2, 7/4)].$$

A model for the curve $X(\mathrm{e}7)$ is given by

$$X(\mathrm{e}7) \; : \; y^2 = 7(16x^4 + 68x^3 + 111x^2 + 62x + 11).$$

The j-map $X(\mathrm{e}7) \to X(1)$ is given by

$$j = \frac{(3x + 1)^3(4x^2 + 5x + 2)^3(x^2 + 3x + 4)^3(x^2 + 10x + 4)^3}{(x^3 + x^2 - 2x - 1)^7}. \quad (14)$$

The $\mathbb{Q}$-points on $X(\mathrm{e}7)$ are $(-1/3, \pm 14/9)$. The Mordell–Weil group over $\mathbb{Q}$ is given by

$$\frac{\mathbb{Z}}{2\mathbb{Z}} \cdot [(-1/3, -14/9) - (-1/3, 14/9)].$$

Remarks.

- The models given for $X(\mathrm{d}7)$ and $X(\mathrm{e}7)$ are 'quartic' genus 1 curves. As both have rational points, we can also give Weierstrass models. In fact, a Weierstrass model for $X(\mathrm{d}7)$ is

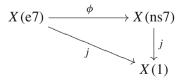$$y^2 + xy = x^3 - x^2 - 107x + 552$$

  with Cremona label 49A3. A Weierstrass model for $X(\mathrm{e}7)$ is

$$y^2 + xy = x^3 - x^2 - 1822x + 30393$$

  with Cremona label 49A4. We prefer to work with the models in the lemma, as the j-maps are expressible simply in terms of the x-coordinates, a fact that is crucial for the proof of Lemma 5.9.
- A model for $X(\mathrm{d}7)$ together with the j-map has been computed by Elkies and is quoted by Sutherland [69, Sect. 3]. Elkies' model and and j-map are equivalent to ours.

Proof of Lemma 5.5 Recall that $C_\mathrm{s}^+(7)$ and $C_\mathrm{ns}^+(7)$ respectively denote the normalizers of the split and non-split Cartan subgroups in $\mathrm{GL}_2(\mathbb{F}_7)$. By definition, curves $X(\mathrm{d}7)$ and $X(\mathrm{e}7)$ are in fact $X(H_1)$ and $X(H_2)$ where H_1, H_2 are as in Proposition 4.1. Then $H_i \cap \mathrm{SL}_2(\mathbb{F}_7)$ with $i = 1, 2$ is a subgroup of index 2 respectively in $C_\mathrm{s}^+(7) \cap \mathrm{SL}_2(\mathbb{F}_7)$ and $C_\mathrm{ns}^+(7) \cap \mathrm{SL}_2(\mathbb{F}_7)$. Thus the inclusion $H_2 \subset C_\mathrm{ns}^+(7)$ induces the following commutative diagram:

$$
\begin{array}{ccc}
X(\mathrm{e}7) & \xrightarrow{\ \phi\ } & X(\mathrm{ns}7) \\
& {\scriptstyle j} \searrow & \downarrow {\scriptstyle j} \\
& & X(1)
\end{array}
$$

where ϕ has degree 2. We also obtain a similar diagram in which $X(\mathrm{d}7)$ and $X(\mathrm{s}7)$ respectively replace $X(\mathrm{e}7)$ and $X(\mathrm{ns}7)$. The curves $X(\mathrm{s}7)$ and

$X(\text{ns}7)$ are rational. In [14, Sect. 5.1], Chen derives expressions for the maps
$j : X(\text{s}7) \to X(1)$ and $j : X(\text{ns}7) \to X(1)$. These are given above in (13)
and (14) where x is respectively a Hauptmodul for $X(\text{s}7)$, $X(\text{ns}7)$. We used
these to derive the models given for $X(\text{d}7)$ and $X(\text{e}7)$. We only give the details
for $X(\text{e}7)$; the details for $X(\text{d}7)$ are an easy adaptation of what follows.

From (14), the map $X(\text{ns}7) \to X(1)$ has degree 21. Subtracting 1728 from
both sides of (14) and factoring gives

$$j - 1728 = \frac{(16x^4 + 68x^3 + 111x^2 + 62x + 11) \cdot G(x)^2}{(x^3 + x^2 - 2x - 1)^7}, \qquad (15)$$

where

$$G(x) = 7(9x^4 + 26x^3 + 34x^2 + 20x + 4)(x^4 + 6x^3 + 17x^2 + 10x + 2).$$

It is convenient to view both $j : X(\text{ns}7) \to X(1)$ and $j : X(\text{e}7) \to X(1)$ as
Belyi covers branched at 0, 1728 and ∞. The factorizations in (14) and (15)
yield the ramification data for $X(\text{ns}7) \to X(1)$:

- there are 7 points above 0, say $p_1, \ldots, p_7$, all having ramification degree 3;
- there are 13 points above 1728, say $q_1, \ldots, q_{13}$, with $q_1, \ldots, q_5$ unramified
  and $q_6, \ldots, q_{13}$ having ramification degree 2;
- there are 3 points above ∞, say r_1, r_2, r_3, all having ramification degree 7.

The corresponding ramification data for $X(\text{e}7) \to X(1)$ can be deduced
from the group H_2; for this we applied the formulae of Ligozat [43], which
give the following information:

- there are 14 points above 0, say $P_1, \ldots, P_{14}$, all having ramification degree
  3;
- there are 22 points above 1728, say $Q_1, \ldots, Q_{22}$, with Q_1, Q_2 unramified,
  and $Q_3, \ldots, Q_{22}$ having ramification degree 2;
- there are 6 points above ∞, say $R_1, \ldots, R_6$, all having ramification degree
  7.

It is clear from the above data that the degree 2 cover $\phi : X(\text{e}7) \to X(\text{ns}7)$
is branched precisely at four of the five points $q_1, \ldots, q_5$ and nowhere else.
One of these five points is $\infty \in X(\text{ns}7)(\mathbb{Q})$ and the other four correspond to
the roots of the irreducible factor $16x^4 + 68x^3 + 111x^2 + 62x + 11$ in (15).
Since the set of branch points is stable under the action of Galois, we see that
ϕ is branched precisely at the roots of $16x^4 + 68x^3 + 111x^2 + 62x + 11$. It
follows that $X(\text{e}7)$ has an equation of the form

$$D_c : y^2 = c(16x^4 + 68x^3 + 111x^2 + 62x + 11),$$

where c is a non-zero, squarefree integer. The map ϕ is then given by $(x, y) \mapsto x$, and the map $j : X(\text{e7}) \to X(1)$ is given by the expression in (14). As H_2 is a congruence group of level 7, the modular curve $X(\text{e7})$ has a model over $\text{Spec}(\mathbb{Z}[1/7])$, and so its Jacobian elliptic curve has good reduction outside 7. Using the formulae in [27], the Jacobian elliptic curve of D_c is given by

$$E_c : Y^2 = X^3 + 111cX^2 + 3512c^2x + 34224.$$

The model E_c has discriminant $2^{12} \cdot 7^3 \cdot c^6$. If $p \mid c$ is a prime $\neq 7$ then the minimal model for E_c has bad reduction at p. Thus c is not divisible by any prime $p \neq 7$. It follows that $c = \pm 1, \pm 7$. We checked that the only values for which the minimal model of E_c has good reduction at 2 are $c = -1$ and $c = 7$. Thus one of the two curves D_{-1}, D_7 is a model for $X(\text{e7})$. The curve D_{-1} has no real points. It is however easy to see that $X(\text{e7})$ must have real points. For example if $E/\mathbb{R}$ is any real elliptic curve, then the image of $\overline{\rho}_{E/\mathbb{R},7}$ has order 2 and every subgroup of $\text{GL}_2(\mathbb{F}_7)$ of order 2 is conjugate to a subgroup of H_2, thus $E/\mathbb{R}$ gives rise to a real point on $X(\text{e7})$. It follows that D_7 is a model for $X(\text{e7})$.

This model has two obvious rational points $(-1/3, \pm 14/9)$ (both with j-invariant 0). Taking $X(\text{e7})$ into Weierstrass form gives the elliptic curve with Cremona label 49A4, which has Mordell–Weil group isomorphic to $\mathbb{Z}/2\mathbb{Z}$. Thus $(-1/3, \pm 14/9)$ are all the rational points of $X(\text{e7})$. Finally the divisor $(-1/3, -14/9) - (-1/3, 14/9)$ cannot be principal as $X(\text{e7})$ has genus 1, and so no two points are linearly equivalent. Hence it must generate $\text{Pic}^0 X(\text{e7})$. $\square$

5.4.2 The modular curves $X(\text{b3}, \text{b5})$ and $X(\text{s3}, \text{b5})$

In Lemmas 5.6 and 5.7 below we give models for $X(\text{b3}, \text{b5})$ and $X(\text{s3}, \text{b5})$ as well as their rational points and the Mordell–Weil groups of their Jacobians. In fact most of this is known (for example [56, Sect. 2]) but we know of no reference for the j-maps, which are essential for our later proofs. The method of Galbraith (sketched in the proof of Lemma 5.2) applies only to non-hyperelliptic modular curves of genus ≥ 3, for which weight 2 cusp forms can be used to define an embedding into projective space. We now sketch an alternative based on a paper of Muić [50, Theorem 3-3]. Let X be a modular curve of genus g corresponding to a congruence subgroup $\Gamma \subseteq \text{SL}_2(\mathbb{Z})$. Let $k \geq 3$ and let $S_k(\Gamma)$ be the space of cusp forms for Γ of weight k. Write t for the dimension $S_k(\Gamma)$, and let $f_1, \ldots, f_t$ be a basis. If $t \geq g + 2$ then $f_1, \ldots, f_t$ define an embedding $|\mathfrak{c}|$ of X into $\mathbb{P}^{t-1}$ given by a very ample divisor $\mathfrak{c}$ of degree $t + g - 1$. In particular, if $t \geq g + 3$, and so $\deg(\mathfrak{c}) \geq 2g + 2$, the image of the embedding is a scheme-theoretic intersection of quadrics [17, page 130]. By using higher weight cusp forms where the space has sufficiently

large dimension one can straightforwardly adapt the method sketched in the proof of Lemma 5.2 to write down a projective model for X and the associated j-map $X \to X(1)$.

For $X(b3, b5)$ and $X(s3, b5)$ the spaces of cusp forms of weight 4 have dimensions 4 and 8 respectively, and this allowed us to write models for these curves as intersections of quadrics in $\mathbb{P}^3$ and $\mathbb{P}^7$ respectively, as well as expressions for the j-maps. Transforming these into Weierstrass models gives Lemmas 5.6 and 5.7 below; the expressions for the j-maps on these two curves are complicated but can be found in the accompanying Magma script (see page 6).

Lemma 5.6 *The curve $X(b3, b5) = X_0(15)$ is an elliptic curve with the Weierstrass model*

$$X(b3, b5) \ : \ y^2 + xy + y = x^3 + x^2 - 10x - 10$$

and Cremona label 15A1. Its $\mathbb{Q}$-points are

$$\infty, \ (-1, 0), \ (-2, -2), \ (8, -27), \ (3, -2), \ (-13/4, 9/8), \ (-2, 3), \ (8, 18).$$

The Mordell–Weil group over $\mathbb{Q}$ is

$$\frac{\mathbb{Z}}{4\mathbb{Z}} \cdot [(-1, 0) - \infty] \oplus \frac{\mathbb{Z}}{2\mathbb{Z}}[(-2, -2) - \infty].$$

Lemma 5.7 *The curve $X(s3, b5)$ is an elliptic curve with the Weierstrass model*

$$X(s3, b5) \ : \ y^2 + xy + y = x^3 + x^2 - 5x + 2, \tag{16}$$

and Cremona label 15A3. The $\mathbb{Q}$-points of this model are

$$\infty, \ (3/4, -7/8), \ (0, -2), \ (2, -4), \ (1, -1), \ (-3, 1), \ (0, 1), \ (2, 1).$$

The Mordell–Weil group of the Jacobian over $\mathbb{Q}$ is

$$\frac{\mathbb{Z}}{4\mathbb{Z}} \cdot [(0, -2) - \infty] \oplus \frac{\mathbb{Z}}{2\mathbb{Z}} \cdot [(3/4, -7/8) - \infty].$$

5.5 A sieve for the symmetric square of a fibre product

Let X_1, X_2 be curves of genus ≥ 1 defined over $\mathbb{Q}$, and let J_1, J_2 be their Jacobians. Let $\mu_k \ : \ X_k \to \mathbb{P}^1$ be morphisms defined over $\mathbb{Q}$ for $k = 1$, 2. Suppose that the fibre product $X_1 \times_{\mathbb{P}^1} X_2$ is irreducible and let X be its

normalization. Let $\pi_k : X \to X_k$ be the induced projection maps. Fix rational degree 2 divisors D_k on X_k and let

$$\iota_k \ : \ X_k^{(2)} \to J_k, \qquad D \mapsto [D - D_k] \tag{17}$$

be the corresponding Abel–Jacobi maps for the symmetric squares. Let

$$\alpha \ : \ X^{(2)} \to J_1 \times J_2, \qquad D \mapsto (\iota_1(\pi_1 D), \iota_2(\pi_2 D)) \,.$$

Let $p_1, \ldots, p_r$ be primes of good reduction for X. For each $p = p_i$ we have the commutative diagram

$$
\begin{array}{ccc}
X^{(2)}(\mathbb{Q}) & \xrightarrow{\ \alpha\ } & J_1(\mathbb{Q}) \times J_2(\mathbb{Q}) \\
{\scriptstyle \text{red}}\downarrow & & \downarrow{\scriptstyle \text{red}} \\
X^{(2)}(\mathbb{F}_p) & \xrightarrow{\ \alpha\ } & J_1(\mathbb{F}_p) \times J_2(\mathbb{F}_p)
\end{array}
$$

where the vertical maps denote reduction modulo p. The following lemma is an obvious consequence of the diagram's commutativity.

Lemma 5.8 *With the above notation and assumptions,*

$$\alpha\left(X^{(2)}(\mathbb{Q})\right) \subseteq \bigcap_{p=p_i} \text{red}^{-1}\, \alpha\left(X^{(2)}(\mathbb{F}_p)\right) .$$

In particular, if the intersection on the right-hand side is empty, then $X^{(2)}(\mathbb{Q}) = \emptyset$ and so the curve X does not have any rational or quadratic points.

Remark. Let $X/\mathbb{Q}$ be a curve of genus ≥ 2 and let J be its Jacobian. Fix some $P_0 \in X(\mathbb{Q})$ and let $\iota : X \to J$ be the corresponding Abel–Jacobi map $P \mapsto [P - P_0]$. For any prime p of good reduction for X, we have a commutative diagram

$$
\begin{array}{ccc}
X(\mathbb{Q}) & \xrightarrow{\ \iota\ } & J(\mathbb{Q}) \\
{\scriptstyle \text{red}}\downarrow & & \downarrow{\scriptstyle \text{red}} \\
X(\mathbb{F}_p) & \xrightarrow{\ \iota\ } & J(\mathbb{F}_p)
\end{array}
$$

If $p_1, \ldots, p_r$ we have

$$\iota\left(X(\mathbb{Q})\right) \subseteq \bigcap_{p=p_i} \text{red}^{-1}\, \iota\left(X(\mathbb{F}_p)\right) .$$

This sieve is known as the Mordell–Weil sieve (e.g. [69, Sect. 3]) and gives information on the image of $X(\mathbb{Q})$ inside $J(\mathbb{Q})$. The Mordell–Weil sieve can be extended in the obvious manner to the symmetric square of X (see for example [61, Sect. 5]), and then it gives information on the image of $X^{(2)}(\mathbb{Q})$ inside $J(\mathbb{Q})$. Now suppose X is the normalization of a fibre product $X_1 \times_{\mathbb{P}^1} X_2$ as above. The Jacobian J is in general much larger (in dimension) than $J_1 \times J_2$, and the information provided by Lemma 5.8 will be weaker than that given by the Mordell–Weil sieve. However, in the Sect. 5.6 where we apply Lemma 5.8, it is realistic to carry out explicit computations with J_1, J_2 but not with J. Indeed, in our applications J_1 and J_2 will be elliptic curves, whereas the possible dimensions for J will be 97, 153, 73 and 113.

5.5.1 Points on the normalization of a fibre product

To apply Lemma 5.8, we need to enumerate the points of $X^{(2)}(\mathbb{F}_p)$ for several primes p. To enumerate $X^{(2)}(\mathbb{F}_p)$ it is enough to know the points of $X(\mathbb{F}_p)$ and $X(\mathbb{F}_{p^2})$, and this is easy in practice if p is sufficiently small and one has a convenient model (say smooth and projective) for X. For the four curves (3) we will apply the lemma to, it seems impractical to write down such a model in view of the very large genera of these curves (Lemma 5.4). In what follows we will use a description of the points of the normalization X of $X_1 \times_{\mathbb{P}^1} X_2$ in terms of the points of X_1, X_2. This will allow us to enumerate the points of X over finite fields.

We continue with the above notation, and we let $\mathbb{F}$ be either an extension of $\mathbb{Q}$ or an extension of $\mathbb{F}_p$ where p is a prime of good reduction for X, X_1, X_2 and the projection maps, and such that the characteristic of $\mathbb{F}$ does not divide the ramification indices of the maps μ_k. By definition of $X_1 \times_{\mathbb{P}^1} X_2$, if $P \in X(\mathbb{F})$, and $P_k = \pi_k(P)$ for $k = 1, 2$ then $P_k \in X_k(\mathbb{F})$ and $\mu_1(P_1) = \mu_2(P_2)$. Now suppose $P_1 \in X_1(\mathbb{F})$, $P_2 \in X_2(\mathbb{F})$ satisfy $\mu_1(P_1) = \mu_2(P_2)$. We would like to describe the points $P \in X(\mathbb{F})$ such that $P_k = \pi_k(P)$ for $k = 1$, 2. Here we follow the description of Flynn and Testa [28, Sect. 2.1]. Write $\gamma = \mu_1(P_1) = \mu_2(P_2) \in \mathbb{P}^1(\mathbb{F})$. Suppose first that $\gamma \neq \infty$, so we may think of $\gamma \in \mathbb{A}^1(\mathbb{F})$. We consider μ_1, μ_2 as elements of the completions of the local rings of X_k at P_k, which we denote by $\widehat{\mathcal{O}}_k$. Then

$$\mu_1 = \gamma + \epsilon s^m, \qquad \mu_2 = \gamma + \delta t^n,$$

where $s \in \widehat{\mathcal{O}}_1$, $t \in \widehat{\mathcal{O}}_2$ are elements of valuation 1, the coefficients ϵ and δ belong to $\mathbb{F}^*$, and m, n are respectively the ramification indices for μ_1, μ_2 at P_1, P_2. A local equation for $X_1 \times_{\mathbb{P}^1} X_2$ in the neighbourhood of (P_1, P_2) is given by

$$\epsilon s^m = \delta t^n.$$

Of course, on this local equation, the point (P_1, P_2) is represented by $(0, 0)$. Now let $d = \gcd(m, n)$. Over $\overline{\mathbb{F}}$, the local equation has d components: $\lambda s^{m'} = t^{n'}$ where $m' = m/d$, $n' = n/d$ and λ runs through the d-th roots of ϵ/δ. We normalize these components by letting $N = m' \cdot n'$ and making the substitutions

$$s = \lambda^a r^{n'}, \qquad t = \lambda^b r^{m'},$$

with $a, b \in \mathbb{Z}$ chosen so that $am' - bn' = -1$. It follows that there are precisely d points $P \in X$ above (P_1, P_2). We may represent these by (P_1, P_2, λ) where $\lambda^d = \epsilon/\delta$. Such a point belongs to $X(\mathbb{F})$ if and only if $P_1 \in X_1(\mathbb{F})$, $P_2 \in X_2(\mathbb{F})$ and $\lambda \in \mathbb{F}$. Moreover, the parameter r has valuation 1 in the completion of the local ring of X at P. The ramification index for $X \to \mathbb{P}^1$ at each of these P is $mn' = m'n = mn/\gcd(m, n)$. To see this, note that the composition of maps

$$X \to X_1 \times_{\mathbb{P}^1} X_2 \to X_1 \to \mathbb{P}^1$$

viewed as an element of the completion of the local ring of X at P can be written as

$$\mu_1 = \gamma + \epsilon s^m = \gamma + \epsilon(\lambda^a r^{n'})^m = \gamma + \epsilon \cdot \lambda^a \cdot r^{mn'}.$$

The above argument needs an easy adaptation if $\gamma = \infty \in \mathbb{P}^1$. When representing the points of X as $P = (P_1, P_2, \lambda)$ the projection maps are obvious: $\pi_k(P) = P_k$.

If $\mathbb{F} = \mathbb{F}_p$ or $\mathbb{F}_{p^2}$, the above gives an algorithm for constructing the $\mathbb{F}$-points of X from the $\mathbb{F}$-points of X_1, X_2 and so it enables us to write down $X^{(2)}(\mathbb{F}_p)$.

5.6 Real quadratic points on $X(b3, b5, d7)$, $X(s3, b5, d7)$, $X(b3, b5, e7)$, and $X(s3, b5, e7)$.

Lemma 5.9 *The curve $X(b3, b5, e7)$ does not have any rational or quadratic points. The non-cuspidal real quadratic points on the curves*

$$X(b3, b5, d7), \qquad X(s3, b5, d7), \qquad X(s3, b5, e7) \tag{18}$$

are modular.

Proof Let $X = X(b3, b5, d7)$ which we think of as the normalization of $X_1 \times_{\mathbb{P}^1} X_2$ where $X_1 = X(b3, b5)$, $X_2 = X(d7)$ with the maps $\mu_k : X_k \to \mathbb{P}^1$ being just the j-maps to $X(1)$. Write J_1, J_2 for the Jacobians of X_1, X_2. In Sect. 5.4 we determined models for the genus 1 curves X_1, X_2 as well as the Mordell–Weil groups $J_1(\mathbb{Q})$, $J_2(\mathbb{Q})$. We fix degree 2 divisors

$$D_1 = 2\infty, \qquad D_2 = 2 \cdot (5/2, 7/4) \tag{19}$$

on X_1, X_2 respectively, which we use to define Abel–Jacobi maps $X_k^{(2)} \to J_k$ as in (17). We now apply Lemma 5.8 with $11 \le p < 100$. We note that these primes are of good reduction for X, and that they do not divide the ramification indices for the maps $X_k \to X(1)$. We find that

$$
\alpha \left(X^{(2)}(\mathbb{Q}) \right) \subseteq \left\{ \Big([0], \ [(5/2, -7/4) - (5/2, 7/4)] \Big), \right.
$$
$$
\Big([(3, -2) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)] \Big),
$$
$$
\left. \Big([(8, 18) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)] \Big) \right\}
$$
$$
\subseteq J_1(\mathbb{Q}) \times J_2(\mathbb{Q}). \tag{20}
$$

Now let P be a quadratic point on X and let P^σ be its conjugate. Then $D = P + P^\sigma \in X^{(2)}(\mathbb{Q})$. From (20) and the definition of α, we have

$$
\pi_2(P) + \pi_2(P^\sigma) - D_2 \ \sim \ (5/2, -7/4) - (5/2, 7/4).
$$

Thus

$$
\pi_2(P) + \pi_2(P^\sigma) = \operatorname{div}(f) + (5/2, 7/4) + (5/2, -7/4)
$$

for some non-zero function f on $X_2 = X(\mathrm{d}7)$. The function f belongs to the Riemann–Roch space of $(5/2, 7/4) + (5/2, -7/4)$. A $\mathbb{Q}$-basis for this Riemann–Roch space is

$$
\frac{1}{2x - 5}, \qquad \frac{x}{2x - 5}.
$$

Thus

$$
f = \frac{a + bx}{2x - 5}
$$

for some $a, b \in \mathbb{Q}$, not both zero. Thus the x-coordinate of $\pi_2(P)$ and $\pi_2(P^\sigma)$ is $-a/b \in \mathbb{P}^1(\mathbb{Q})$. From Lemma 5.5, we know that the j-value of $\pi_2(P)$ is expressible in terms of its x-coordinate as given by (13), hence it belongs to $\mathbb{P}^1(\mathbb{Q})$. Now suppose P is a non-cuspidal real quadratic point on X. It follows that any elliptic curve supporting P has j-invariant belonging to $\mathbb{Q}$ and so is modular by [9].

We apply the same sieve to $X = X(\mathrm{s}3, \mathrm{b}5, \mathrm{d}7)$, where we take $X_1 = X(\mathrm{s}3, \mathrm{b}5)$ and $X_2 = X(\mathrm{d}7)$, and we take D_1, D_2 as in (19), where of course, now ∞ is the point at infinity on the model for X_1 given in (16). We obtained

$$\alpha\left(X^{(2)}(\mathbb{Q})\right) \subseteq \Big\{ \big([0], \ [(5/2, -7/4) - (5/2, 7/4)]\big),$$

$$\big([(2, -4) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)]\big),$$

$$\big([(1, -1) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)]\big)\Big\}$$

$$\subseteq J_1(\mathbb{Q}) \times J_2(\mathbb{Q}). \tag{21}$$

By exactly the same argument as before, quadratic points have rational j-invariants, and thus non-cuspidal real quadratic points are modular.

Applying the sieve to $X = X(\text{b3}, \text{b5}, \text{e7})$ shows that $\alpha(X^{(2)}(\mathbb{Q}))$ is empty, so these this curve have no rational or quadratic points.

Finally, we apply the sieve to $X = X(\text{s3}, \text{b5}, \text{e7})$. Of course, we take $X_1 = X(\text{s3}, \text{b5})$, $X_2 = X(\text{e7})$. We fix the degree 2 divisors

$$D_1 = 2\infty, \qquad D_2 = 2 \cdot (-1/3, 14/9)$$

of X_1 and X_2. The sieve yields

$$\alpha\left(X^{(2)}(\mathbb{Q})\right) \subseteq \Big\{ \big([(0, 1) - \infty], \ [(-1/3, -14/9) - (-1/3, 14/9)]\big)\Big\}.$$

If P is a quadratic point on X and $\pi_2(P)$ is its image on $X_2 = X(\text{e7})$ then

$$\pi_2(P) + \pi_2(P^\sigma) = \text{div}(f) + (-1/3, 14/9) + (-1/3, -14/9).$$

As before, the x-coordinate of $\pi_2(P)$ on the model for $X(\text{e7})$ belongs to $\mathbb{P}^1(\mathbb{Q})$ and so the j-invariant (given by (14)) is rational. This completes the proof. $\qquad\square$

Remarks.

(I) In the above, we have morphisms $\alpha : X^{(2)} \to J_1 \times J_2$. The points $X^{(2)}(\mathbb{Q})$ belong to the fibres $\alpha^{-1}(J_1(\mathbb{Q}) \times J_2(\mathbb{Q}))$. Since $J_1(\mathbb{Q})$ and $J_2(\mathbb{Q})$ are finite and explicitly known, the reader is perhaps puzzled as to why we do not use this to recover $X^{(2)}(\mathbb{Q})$. We are in fact able to write down equations for these fibres as schemes in $\mathbb{P}^1 \times \mathbb{P}^1$. Most of these are zero-dimensional, but are of such high degree that we have found it impractical to decompose them using Gröbner basis computations. A few of the fibres are however 1-dimensional. These are so complicated that we are unable to say anything about their rational points.

(II) Our sieve cannot prove that $X^{(2)}(\mathbb{Q})$ is empty for the three modular curves (18). It is natural to ask whether these curves have rational or quadratic

points. The genera of these three modular curves are 97, 153 and 113 respectively by Lemma 5.4. As these are so large, it is reasonable to suspect that the only quadratic points lie above 0, 1728 or ∞ (and Remark (III) below shows that this is indeed the case). Using the description of the points of $X = X(b3, b5, d7)$ as the fibre product of $X_1 = X(b3, b5)$ and $X_2 = X(d7)$ we find that there are precisely 8 quadratic cusps, and no quadratic points above 0 and 1728. We shall denote the quadratic cusps by $P_1^{\pm}, P_2^{\pm}, P_3^{\pm}, P_4^{\pm}$. These respectively lie above

$$\left(\infty, \infty_{\pm}\right), \ \left((-1, 0), \infty_{\pm}\right), \ \left((-2, 3), \infty_{\pm}\right), \ \left((8, 18), \infty_{\pm}\right) \ \in X_1 \times_{\mathbb{P}^1} X_2.$$

Here $\infty_+, \infty_- \in X_2$ are conjugates defined over $\mathbb{Q}(\sqrt{-7})$. Thus $P_i^+ + P_i^-$ belong to $X^{(2)}(\mathbb{Q})$. We find that

$$\alpha(P_1^+ + P_1^-) = \alpha(P_2^+ + P_2^-) = \left([0], \ [(5/2, -7/4) - (5/2, 7/4)]\right),$$
$$\alpha(P_3^+ + P_3^-) = \alpha(P_4^+ + P_4^-) = \left([(3, -2) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)]\right).$$

This is consistent with the result of the sieve given in (20), and provides a useful check on the correctness of our computations.

We carried the same computation for $X = X(s3, b5, d7)$ and found quadratic cusps $P_1^{\pm}, P_2^{\pm}, P_3^{\pm}, P_4^{\pm}$ and no other quadratic points above 0, 1728 and ∞. The four quadratic cusps respectively lie above

$$\left(\infty, \infty_{\pm}\right), \ \left((0, 1), \infty_{\pm}\right), \ \left((2, -4), \infty_{\pm}\right), \ \left((-3, 1), \infty_{\pm}\right) \ \in X_1 \times_{\mathbb{P}^1} X_2,$$

where of course $X_1 = X(s3, b5)$ and $X_2 = X(d7)$. We find that

$$\alpha(P_1^+ + P_1^-) = \alpha(P_4^+ + P_4^-) = \left([0], \ [(5/2, -7/4) - (5/2, 7/4)]\right),$$
$$\alpha(P_2^+ + P_2^-) = \alpha(P_3^+ + P_3^-) = \left([(1, -1) - \infty], \ [(5/2, -7/4) - (5/2, 7/4)]\right).$$

This is consistent with the result of the sieve given in (21).

As a final check on our computations, we verified that there are no quadratic points on $X(b3, b5, e7)$ and $X(s3, b5, e7)$ above 0, 1728 and ∞.

(III) Although we do not need this for the proof of Lemma 5.9, it is interesting to ask if it can be shown that all quadratic points P on $X_1 := X(b3, b5, d7)$, $X_2 := X(s3, b5, d7)$, $X_3 := X(s3, b5, e7)$ do satisfy $j(P) \in \{0, 1728, \infty\}$. We explain how to solve this problem by reducing it to the determination of *rational* points on some modular curves with 'milder' level structure. Indeed, let P be a non-cuspidal quadratic point on X, where X is one of these three modular curves X_1, X_2, X_3. We know by the proof of Lemma 5.9 that $j(P) \in \mathbb{Q}$.

Thus the point P is supported by some elliptic curve E defined over $\mathbb{Q}$, even though the level structure is defined over some quadratic field K. We know in particular that $\overline{\rho}_{E,5}(G_K)$ is contained in a Borel subgroup $B(5)$ of $\mathrm{GL}_2(\mathbb{F}_5)$. Now $\overline{\rho}_{E,5}(G_K)$ is a subgroup of $\overline{\rho}_{E,5}(G_{\mathbb{Q}})$ of index 1 or 2. By a straightforward analysis of the subgroups of $\mathrm{GL}_2(\mathbb{F}_5)$ it follows that $\overline{\rho}_{E,5}(G_{\mathbb{Q}})$ is contained either in $B(5)$ or $C_s^+(5)$ (the normalizer of a split Cartan subgroup). Similarly, for $X = X_1$, we have $\overline{\rho}_{E,3}(G_{\mathbb{Q}})$ is contained either in $B(3)$ or $C_s^+(3)$ and for $X = X_2, X = X_3$ it is contained in $C_{ns}^+(3)$. Finally, for $X = X_1, X_2$, the image $\overline{\rho}_{E,7}(G_{\mathbb{Q}})$ is contained in $C_s^+(7)$ and for $X = X_3$ it is contained in $C_{ns}^+(7)$. It is now not difficult to show in all cases, with the help of the results of this paper and of [42], that $j(E) \in \{0, 1728\}$. Consider $X = X_1$ for example. Then E gives rise to a $\mathbb{Q}$-point of one of the curves $X(\mathrm{b}3, \mathrm{b}5)$, $X(\mathrm{b}3, \mathrm{s}5)$, $X(\mathrm{s}3, \mathrm{b}5)$ or $X(\mathrm{s}3, \mathrm{s}5)$. We have already determined the $\mathbb{Q}$-points on these four curves in the proofs of Lemmas 5.2, 5.3, and in Lemmas 5.6, 5.7. Enumerating the j-invariants of these points Q, we are able to show for $j(Q) \neq 0, 1728$ that they do not pull back to non-cuspidal quadratic points on $X(\mathrm{d}7)$.

5.7 Modularity of elliptic curves over $\mathbb{Q}(\sqrt{5})$

To complete the proof of Theorem 1 we merely have to prove Lemma 1.1, which states that every elliptic curve over $K = \mathbb{Q}(\sqrt{5})$ is modular. Again we reduce this problem to showing that the K-points on certain modular curves are in fact modular (in the sense that they correspond to modular elliptic curves). We give fewer details in this section as compared to previous sections. This is because the task is much simpler: determining the points on a curve over one quadratic field is usually easier than determining the points over all real quadratic fields.

Proof of Lemma 1.1 Let $K = \mathbb{Q}(\sqrt{5})$, and let E be an elliptic curve over K. We would like to show that E is modular. By Theorems 3 and 4, if $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is absolutely irreducible with $p = 3$ or 7 then E is modular. Thus we may restrict our attention to elliptic curves E such that $\overline{\rho}_{E,p}(G_{K(\zeta_p)})$ is simultaneously absolutely reducible for $p = 3$ and 7. Now appealing to Corollary 4.2 we see that E gives rise to a K-point on one of the following four modular curves

$$X(\mathrm{d}7), \quad X(\mathrm{e}7), \quad X(\mathrm{b}3, \mathrm{b}7), \quad X(\mathrm{s}3, \mathrm{b}7).$$

We determined models for the first two curves in Lemma 5.5 and found that they are isomorphic (over $\mathbb{Q}$) to elliptic curves 49A3 and 49A4 respectively. The third curve is isomorphic (over $\mathbb{Q}$) to the elliptic curve 21A1. Let X be any of the first three curves. By considering both X and its $\sqrt{5}$-twist it is easy

to show that $X(K) = X(\mathbb{Q})$ in all cases. It follows that the K-points on the first three curves are modular.

Finally let $X = X(\text{s3}, \text{b7}) \cong X_0(63)/\langle w_9 \rangle$. This curve has the obvious quotient $Y = X_0(63)/\langle w_7, w_9 \rangle$. The curve Y has genus 1 and is in fact isomorphic to the elliptic curve 21A1. By the above, $Y(K) = Y(\mathbb{Q})$. Thus if P is a non-cuspidal K-point on X then its image on Y belongs to $Y(\mathbb{Q})$. Let σ be the non-trivial automorphism of K. Thus P and P^σ map to the same point on Y. Hence $P^\sigma = P$ or $P^\sigma = w_7(P)$. In the former case P is a rational point and hence modular. In the latter case, P corresponds to a $\mathbb{Q}$-curve, and is therefore modular. $\qquad\square$

6 Further directions

6.1 On modularity over totally real fields (again)

Most of the theory we have developed in Sects. 2, 3 and 4 applies to elliptic curves over arbitrary totally real fields. If one is interested in modularity of elliptic curves over a particular totally real field K then it might be possible to prove this by the methods of this paper, together with explicit methods for computing rational points on curves (e.g. [62,68]). However, one can considerably simplify this task by imposing suitable restrictions on the field K and elliptic curve E.

Theorem 7 *Let $p = 5$ or 7. Let K be a totally real field having some unramified place υ above p. Let E be an elliptic curve semistable at υ and suppose $\overline{\rho}_{E,p}$ is irreducible. Then E is modular.*

Proof As υ/p is unramified, we have $K \cap \mathbb{Q}(\zeta_p) = \mathbb{Q}$. Let I_υ be the inertia subgroup of G_K at υ. By [57, Propositions 11, 12, 13], the image of $\overline{\rho}_{E,p}(I_\upsilon)$ in $\mathrm{PGL}_2(\mathbb{F}_p)$ contains a cyclic subgroup of order $p-1$ or $p+1$. It follows from Proposition 4.1 that $\overline{\rho}(G_{K(\zeta_p)})$ is absolutely irreducible. Thus E is modular by Theorems 3 and 4. $\qquad\square$

In [26], the first-named author proves a version of Theorem 7 with $p = 3$ (and without the irreducibility assumption) using modularity lifting theorems due to Skinner and Wiles [64,65].

6.2 Diophantine applications

Wiles' proof of Fermat's Last Theorem provides a template for attacking many other Diophantine equations via Frey curves and modularity (see [6,60] for recent surveys). Modularity results over totally real fields extend the applicability of this template, as for example in [21,36]. We expect the results and

methods of the current paper to play an important rôle in the development of this area. Indeed, a forthcoming paper [29] uses Theorem 1 to study the Fermat equation over real quadratic fields. In particular, it is shown that if $d > 6$ is a squarefree integer satisfying $d \equiv 3 \pmod 8$ or $d \equiv 6, 10 \pmod{16}$, then there is an effectively computable constant B_d such that for all primes $p > B_d$, every solution to the Fermat equation

$$x^p + y^p = z^p, \qquad x, y, z \in \mathbb{Q}(\sqrt{d})$$

satisfies $xyz = 0$.

References

1. Allen, P.B.: Modularity of nearly ordinary 2-adic residually dihedral Galois representations. arXiv:1301.1113v2, 11 Sep 2013
2. Banwait, B.S., Cremona, J.: Tetrahedral elliptic curves and the local-to-global principle for isogenies. arXiv:1306.6818v2, 16 Aug 2013
3. Barnet-Lamb, T., Gee, T., Geraghty, D.: Congruences between Hilbert modular forms: constructing ordinary lifts. Duke Math. J. **161**, 1521–1580 (2012)
4. Barnet-Lamb, T., Gee, T., Geraghty, D.: Congruences between Hilbert modular forms: constructing ordinary lifts II. Math. Res. Lett. **20**, 81–86 (2013)
5. Bars, F., Kontogeorgis, A., Xarles, X.: Bielliptic and hyperelliptic modular curves $X(n)$ and the group $Aut(X(n))$, Acta Arith. (to appear)
6. Bennett, M.A., Chen, I., Dahmen, S.R., Yazdani, S.: Generalized Fermat equations: a miscellany. preprint (2013)
7. Bosma, W., Cannon, J., Playoust, C.: The Magma Algebra System I: the user language. J. Symb. Comp. 24, 235–265 (1997) (See also http://magma.maths.usyd.edu.au/magma/)
8. Blasius, D., Rogawski, J.D.: Motives for Hilbert modular forms. Invent. Math. **114**(1), 55–87 (1993)
9. Breuil, C., Conrad, B., Diamond, F., Taylor, R.: On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises. J. Am. Math. Soc. **14**, 843–939 (2001)
10. Breuil, C., Diamond, F.: Formes modulaires de Hilbert modulo p et valeurs d'extensions galoisiennes. Ann. Sci. l'École Norm. Supér. (to appear)
11. Bruin, P., Najman, F.: On isogenies of elliptic curves over quadratic fields. preprint (2013)
12. Carayol, H.: Sur les représentations ℓ- adiques associées aux formes modulaires de Hilbert. Ann. Sci. Econ. Norm. Sup. **4**, 409–468 (1986)
13. Carayol, H.: Sur les représentations galoisiennes modulo attachées aux formes modulaires. Duke Math. J. **59**, 785–801 (1989)
14. Chen, I.: The Jacobian of modular curves associated to cartan subgroups, DPhil thesis, University of Oxford (1996)
15. Conrad, B., Diamond, F., Taylor, R.: Modularity of certain potentially Basotti–Tate Galois representations. J. Am. Math. Soc. **12**(2), 521–567 (1999)

16. Cremona, J.E.: Algorithms for Modular Elliptic Curves. Cambridge University Press, Cambridge (1992)
17. Danilov, V.I., Shokurov, V.V.: Algebraic Curves, Algebraic Manifolds and Schemes. Springer, Berlin (1998)
18. Darmon, H., Diamond, F., Taylor, R.: Fermat's last theorem. In: Coates, J., Yau, S.T. (eds.) Elliptic Curves, Modular Forms and Fermat's Last Theorem, pp. 2–140, 2nd edn. International Press, Cambridge (1997)
19. Diamond, F.: On deformation rings and Hecke rings. Ann. Math. **144**, 137–166 (1996)
20. Diamond, F., Shurman, J.: A First Course on Modular Forms, GTM 228, Springer (2005)
21. Dieulefait, L., Freitas, N.: Fermat-type equations of signature (13,13, p) via Hilbert cuspforms. Math. Ann. (to appear) (2013)
22. Ellenberg, J.: Serre's conjecture over $\mathbb{F}_9$. Ann. Math. **161**(2), 1111–1142 (2005)
23. Elkies, N.: The Klein quartic in number theory. In: Levy, S. (ed.) The Eightfold Way: The Beauty of Klein's Quartic Curve. vol 35, MSRI Publications (1998)
24. Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math. **73**(3), 349–366 (1983)
25. Faltings, G.: The general case of S. Langs conjecture, pages 175182 of Barsotti Symposium in Algebraic Geometry (Abano Terme, 1991), Perspect. Math. 15, Academic Press, San Diego, CA (1994)
26. Freitas, N.: Recipes for Fermat-type equations of the form $x^r + y^r = Cz^p$
27. Fisher, T.: The invariants of a genus one curves. Proc. Lond. Math. Soc. **97**, 753–782 (2008)
28. Flynn, E.V., Testa, D.: Finite weil restrictions of curves. arXiv:1210.4407v3, 19 May 2013
29. Freitas, N., Siksek, S.: Modularity and the Fermat equation over totally real fields. arXiv:1307.3162, 11 July 2013
30. Galbraith, S.D.: Equations for modular curves, DPhil thesis, University of Oxford (1996)
31. Gee, T.: Automorphic lifts of prescribed types. Math. Ann. **350**, 107–144 (2011)
32. Halberstadt, E., Kraus, A.: Sur la courbe modulaire $X_E(7)$. Exp. Math. **12**(1), 27–40 (2003)
33. Harris, J., Silverman, J.H.: Bielliptic curves and symmetric products. Proc. Am. Math. Soc. **112**(2), 347–356 (1991)
34. Hartshorne, R.: Algebraic Geometry, GTM 52, Springer (1977)
35. Jarvis, F., Manoharmayum, J.: On the modularity of supersingular elliptic curves over certain totally real number fields. J. Number Theory **128**(3), 589–618 (2008)
36. Jarvis, F., Meekin, P.: The Fermat equation over $\mathbb{Q}(\sqrt{2})$. J. Number Theory **109**(1), 182–196 (2004)
37. Kamienny, S.: Torsion points on elliptic curves and q-coefficients of modular forms. Invent. Math. **109**(2), 221–229 (1992)
38. Khare, C., Wintenberger, J.-P.: Serre's modularity conjecture. I. Invent. Math. **178**(3), 485–504 (2009)
39. Khare, C., Wintenberger, J.-P.: Serre's modularity conjecture. II. Invent. Math. **178**(3), 505–586 (2009)
40. Kisin, M.: Moduli of finite flat group schemes, and modularity. Ann. Math. Second Ser. **170**(3), 1085–1180 (2009)
41. Langlands, R.: Base change for GL(2). Ann. Math. Studies 96, Princeton University Press (1980)
42. Le Hung, B.: Modularity of elliptic curves over some totally real fields. arXiv:1309.4134, 16 Sep 2013
43. Ligozat, G.: Courbes modulaires de niveau 11. In: Serre, J.-P., Zagier, D.B. (eds.) Modular Functions of One Variable V, Lecture Notes in Mathematics, vol. 601, pp. 147–237. Springer, Berlin (1977)
44. Manoharmayum, J.: On the modularity of certain $GL_2(\mathbb{F}_7)$ Galois representations. Math. Res. Lett. **8**(5–6), 703–712 (2001)

45. Manoharmayum, J.: Serre's conjecture for mod 7 Galois representations. In: Cremona, J., Lario, J.C., Quer, J., Ribet, K. (eds.) Modular curves and Abelian varieties. Progr. Math., vol. 224, pp. 141–149. Birkhäuser, Basel (2004)
46. Mazur, B.: Rational points on modular curves. In: Serre, J.-P., Zagier, D.B. (eds.) Modular Functions of One Variable V, Lecture Notes in Mathematics, vol. 601, pp. 147–237. Springer, Berlin (1977)
47. Mazur, B.: Modular curves and the Eisenstein ideal. Inst. Hautes Études Sci. Publ. Math. **47**(1977), 33–186 (1978)
48. Merel, L.: Bornes pour la torsion des courbes elliptiques sur les corps de nombres. Invent. Math. **124**(1–3), 437–449 (1996)
49. Milne, J.S.: Abelian varieties. In: Cornell, G., Silverman, J.H. (eds.) Arithmetic Geometry, pp. 103–150. Springer, Berlin (1986)
50. Muić, G.: On embeddings of modular curves in projective space. preprint, 17 March 2012
51. Poonen, B., Schaefer, E.F., Stoll, M.: Twists of X(7) and primitive solutions to $x^2+y^3 = z^7$. Duke Math. J. **137**, 103–158 (2007)
52. Ogg, A.: Hyperelliptic modular curves. Bull. Soc. Math. Fr. **102**, 449–462 (1974)
53. Ribet, K.A.: Abelian varieties over $\mathbb{Q}$ and modular forms. In: Cremona, J., Lario, J.C., Quer, J., Ribet, K. (eds.) Modular curves and Abelian varieties. Progr. Math., vol. 224, pp. 241–261. Birkhäuser, Basel (2004)
54. Rohrlich, D.E.: Elliptic curves and the Weil-Deligne group. In: Kisilevsky, H., Murty, M.R. (eds.) Elliptic curves and related topics. CRM Proceedings and Lecture Notes 4, pp. 125–158, American Mathematical Society (1994)
55. Rohrlich, D.E.: Modular curves, Hecke correspondences, and L-functions. In: Cornell, G., Silverman, J.H., Stevens, G. (eds.) Modular Forms and Fermat's Last Theorem, pp. 41–100. Springer, Berlin (1997)
56. Rubin, K.: Modularity of mod 5 representations. In: Cornell, G., Silverman, J.H., Stevens, G. (eds.) Modular Forms and Fermat's Last Theorem, pp. 463–474. Springer, Berlin (1997)
57. Serre, J.-P.: Properiétés galoisiennes des points d'ordre fini des courbes elliptiques. Invent. Math. **15**, 259–331 (1972)
58. Serre, J.-P.: Lectures on the Mordell-Weil Theorem, Aspects of Mathematics 15, 3rd edn. Vieweg (1997)
59. Serre, J.-P.: Topics in Galois Theory. 2nd edn. A K Peters/CRC Press 2007
60. Siksek, S.: The modular approach to Diophantine equations. In: Cohen, H. (ed.) Number Theory, Volume II: Analytic and Modern Tools, GTM 240, pp. 495–527. Springer, Berlin (2007)
61. Siksek, S.: Chabauty for symmetric powers of curves. Algebra Number Theory **3**(2), 209–236 (2009)
62. Siksek, S.: Explicit Chabauty over number fields. Algebra Number Theory **7**(4), 765–793 (2013)
63. Silverman, J.H.: Advanced topics in the arithmetic of elliptic curves, GTM 151, Springer (1994)
64. Skinner, C., Wiles, A.: Residually reducible representations and modular forms. Publications mathématiques de l'I.H.É.S. 89, 5–126 (1999)
65. Skinner, C., Wiles, A.: Nearly ordinary deformations of irreducible residual representations. Ann. Fac. Sci. Toulouse Math. **10**, 185–215 (2001)
66. Stein, W.A.: Modular forms: a computational approach. Am. Math. Soc. (2007)
67. Stoll, M.: Implementing 2-descent for Jacobians of hyperelliptic curves. Acta Arith. **98**, 245–277 (2001)
68. Stoll, M.: Rational points on curves. J. Théorie Nombres Bordx. **23**, 257–277 (2011)
69. Sutherland, A.V.: A local-global principle for rational isogenies of prime degree. J. Théorie Nombres Bordx. **24**(2), 475–485 (2012)

70. Swinnerton-Dyer, H.P.F.: On ℓ- adic representations and congruences for coefficients of modular forms. In: Kuyk, W., Serre, J.-P. (eds.): Modular Functions of One Variable III, Lecture Notes in Mathematics, vol. 350, pp. 1–55. Springer, Berlin (1973)
71. Taylor, R.: On Galois representations associated to Hilbert modular forms. Invent. Math. **98**(2), 265–280 (1989)
72. Taylor, R.: On icosahedral Artin representations II. Am. J. Math. **125**, 549–566 (2003)
73. Taylor, R., Wiles, A.: Ring-theoretic properties of certain Hecke algebras. Ann. Math. **141**(3), 553–572 (1995)
74. Tunnell, J.: Artin's conjecture for representations of octahedral type. Bull. AMS. **5**, 173–175 (1981)
75. Wiles, A.: Modular elliptic curves and Fermat's last theorem. Ann. Mathe. **141**(3), 443–551 (1995)
76. Wiles, A.: On ordinary λ- adic representations associated to modular forms. Invent. Math. **94**(3), 529–573 (1988)