Invent. math. 140, 487–509 (2000)
Digital Object Identifier (DOI) 10.1007/s002220000055

Inventiones mathematicae

Torsion points on modular curves*

Matthew H. Baker

Department of Mathematics, Harvard University, Cambridge, MA 02138, USA
(e-mail: mbaker@math.harvard.edu)

Oblatum 1-VI-1999 & 19-X-1999
Published online: 29 March 2000 –

Abstract. Let $N \geq 23$ be a prime number. In this paper, we prove a conjecture of Coleman, Kaskel, and Ribet about the $\overline{\mathbb{Q}}$-valued points of the modular curve $X_0(N)$ which map to torsion points on $J_0(N)$ via the cuspidal embedding. We give some generalizations to other modular curves, and to noncuspidal embeddings of $X_0(N)$ into $J_0(N)$.

1. Introduction

Let X be an algebraic curve of genus $g \geq 1$ defined over a number field K. (For us, the word *curve* used without further qualifications will always mean a complete, nonsingular, absolutely irreducible curve over a field.) Assume, furthermore, that $X(K)$ is nonempty. Now choose an Albanese embedding defined over K of X into its Jacobian variety. In other words, choose a K-rational point Q on X and define the map $i_Q : X \hookrightarrow J$ by sending P to the divisor class $[(P) - (Q)]$.

Now define the set $T_Q(X)$ to be $\{P \in X(\bar{K}) \mid i_Q(X) \in J^{\mathrm{tor}}\}$. In other words, $T_Q(X)$ is the set of $\bar{K}$-valued points of X which map to torsion points on J via i_Q. Following [5], we call $T_Q(X)$ the *torsion packet* containing Q.

If $g = 1$ (i.e., X is an elliptic curve), then i_Q is an isomorphism, and so of course T is infinite. But if $g \geq 2$, the situation is entirely different. The Manin–Mumford conjecture (proven by M. Raynaud in 1983) says that when $g \geq 2$, $T_Q(X)$ is a finite set of points. It also follows from results of Raynaud that when $g \geq 2$, the cardinality of $T_Q(X)$ is bounded independently of $Q \in X(\bar{K})$. For a proof, see the appendix to [1].

There is a striking analogy between the Manin–Mumford conjecture, on the one hand, and the Mordell conjecture on the other. For example, in [15],

* The author's research was supported by an NDSEG Fellowship and by a Sloan Doctoral Dissertation Fellowship.

Lang conjectured that $i(X) \cap \Gamma'$ is finite whenever i is an embedding of X into J, Γ is a finitely generated subgroup of $J(\bar{K})$, and Γ' is its division group, i.e., the set of points x in $J(\bar{K})$ such $nx \in \Gamma$ for some positive integer n. This is now a theorem, as are various generalizations to higher-dimensional varieties; see [23] for references and a summary of recent results in this direction. Note that Lang's conjecture implies both the Manin–Mumford conjecture (taking $\Gamma = 0$) and the Mordell conjecture (taking $\Gamma = J(K)$ and considering $i(X) \cap \Gamma \subseteq i(X) \cap \Gamma'$).

Determining the finite set of K-rational points on X ("Explicit Mordell") for a "random" curve X is an extremely hard problem. Faltings' proof of the Mordell conjecture is ineffective, so even in principle this problem is difficult. Some of the most celebrated cases where $X(\mathbb{Q})$ has been determined include the case where X is a Fermat curve (A. Wiles) and where $X = X_0(N)$ is a modular curve (B. Mazur). There are also small industries devoted to solving this problem in the special case where X has genus 1 or 2.

Explicitly determining the set T of torsion points on X ("Explicit Manin–Mumford") is also, in general, a difficult one. In this setting as well, the appropriate test cases seem to be curves which either have small genus (see [2] for some examples when $g = 2$) or unusually rich arithmetic structure. For an example of the latter, see [7], in which the authors determine T when X is a Fermat curve embedded in J using a "cusp".

In their joint paper [6], Coleman, Kaskel, and Ribet study the set of points on the modular curve $X_0(N)$ (here $N \geq 23$ is a prime number) which map to torsion points of $J_0(N)$ under the embedding $i_\infty : P \mapsto [(P) - (\infty)]$. (Here ∞ denotes one of the cusps on $X_0(N)$.) We call the embedding i_∞ the *cuspidal embedding* of $X_0(N)$ into $J_0(N)$.

For the reader's convenience, we recall a few definitions. $X_0(N)$ is the (compactified) coarse moduli space for the set of (cyclic) isogenies $E \to E'$ of degree N between elliptic curves. The algebraic curve $X_0(N)$ is defined over $\mathbb{Q}$, and the assumption that $N \geq 23$ simply means that the genus of this curve is at least two. As a Riemann surface, $X_0(N)$ can be thought of as the quotient of the complex upper half plane $\mathcal{H}$ by the action of the group $\Gamma_0(N)$, at least once this quotient is suitably compactified by adding two cusps, which are $\mathbb{Q}$-rational points that we call 0 and ∞. For additional background material on $X_0(N)$ and its Jacobian $J_0(N)$, as well as a number of important results we will use in what follows, see B. Mazur's paper "Modular Curves and the Eisenstein Ideal" ([18]).

The curve $X_0(N)$ has a natural involution w_N, the Atkin–Lehner involution, whose moduli interpretation is that it takes an N-isogeny $E \to E'$ to the dual isogeny $E' \to E$. The quotient of $X_0(N)$ by w_N is denoted by $X_0^+(N)$, which is also an algebraic curve defined over $\mathbb{Q}$. We let $g_0^+(N)$ (or simply g^+) be the genus of $X_0^+(N)$.

When g^+ happens to be zero (which happens, for $N \geq 23$, if and only if $N \in \{23, 29, 31, 41, 47, 59, 71\}$), $X_0(N)$ is forced to be a hyperelliptic

curve (double cover of $\mathbf{P}^1$). It is a theorem of Ogg [21] that the converse is almost true as well: $X_0(N)$ is hyperelliptic if and only if $g^+ = 0$ or $N = 37$. The curve $X_0(37)$ is unusual in that the hyperelliptic involution and Atkin–Lehner involution do not coincide.

We call the set $T_\infty(X_0(N))$ of points Q on $X_0(N)$ such that $i_\infty(Q)$ has finite order the *cuspidal torsion packet* on $X_0(N)$. Certainly the two cusps ∞ and 0 are in this torsion packet; the image under i_∞ of latter point has order $n = \mathrm{Num}\frac{N-1}{12}$ in $J_0(N)$ by a well-known theorem of Ogg.

Furthermore, there can sometimes be other points in $T_\infty(X_0(N))$. A proof of the following proposition can be found in [6, Proposition 1.1].

Proposition 1.1. *When $g^+ = 0$, the hyperelliptic branch points on $X_0(N)$ (the points which ramify in the degree 2 covering $X_0(N) \to \mathbf{P}^1$) are in the cuspidal torsion packet $T_\infty(X_0(N))$. When $N = 37$, the hyperelliptic branch points are not in $T_\infty(X_0(N))$.*

The authors of [6] make the following guess about the cuspidal torsion packet on $X_0(N)$, which we refer to as the Coleman–Kaskel–Ribet (CKR) conjecture:

Conjecture 1.2. *For all prime numbers $N \geq 23$,*

$$T_\infty(X_0(N)) = \begin{cases} \{0, \infty\} & \text{if} \quad g^+ > 0 \\ \{0, \infty\} \cup \{\text{hyperelliptic branch points}\} & \text{if} \quad g^+ = 0\,. \end{cases}$$

They prove this result in the special case where $N = 37$ using results about $J_0(37)$ found in B. Kaskel's thesis.

In this paper we give two proofs of Conjecture 1.2. In Sect. 2, we summarize the work previously done on this problem, in particular the results of [6] and [26]. We also discuss a few technical results needed later on. In Sect. 3, we give our proofs of the CKR conjecture.

We now give a brief summary of the two proofs of the CKR conjecture. In both arguments, a key fact is the theorem of Mazur which says that the intersection (via the cuspidal embedding) of $X_0(N)$ and the cuspidal group C of $J_0(N)$ consists precisely of the cusps of $X_0(N)$. (The cuspidal group is the cyclic subgroup of $J_0(N)(\mathbb{Q})$ generated by the difference of cusps $i_\infty(0) = [(0) - (\infty)]$). Following [6], the idea of the first proof is to take a torsion point P on $X_0(N)$ and to decompose it into its primary components P_l; if one can show that all P_l are in C, then so is P and we are done. We first analyze the cases where one of P_2, P_3 is not in C. This can happen only when $X_0(N)$ is hyperelliptic or trigonal, and results of [1] and [6] prove the conjecture for all such values of N. Then, assuming that N is large enough so that P_2, P_3 are in C, we prove that there is an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which acts on the projection P^+ of P to $J_0^+(N)$ as multiplication by -1, which is impossible unless the genus of $X_0^+(N)$ is at most 2. The cases where $g^+ \leq 2$ are dealt with using the results of [6] and calculations found in [32]. A slight complication arises in this method when

N ramifies in the Hecke algebra $\mathbf{T}$, but we find a way around this based on bounds for the gonality of $X_0(N)$ obtained in [1].

Our second proof of the CKR conjecture, which came into being shortly after the first (due to insights of Ken Ribet), seems more ripe for generalization. It is conceptually simper than the first proof; in particular, the trigonal modular curves no longer play an exceptional role in this approach, and Coleman's theory of p-adic integration as developed in [5] and [6] is not needed. The idea here is to first deal with the case where the torsion point P is annihilated by the Eisenstein ideal of $\mathbf{T}$, which can be done by using results from [6] or by exploiting the Galois-module structure of $J_0(N)[\mathfrak{J}]$ determined in [18] and [26]. In other cases, we use a result of Ribet which says that a point not in $J_0(N)[\mathfrak{J}]$ must be ramified at N. We then find an element σ in an inertia group at N which acts nontrivially on P, but such that $(\sigma - 1)^2 P = 0$. For example, when the order of P is prime to N, the existence of σ follows from Grothendieck's Galois criterion for semistable reduction (see [10, Proposition 3.5]). We then have $(\sigma^2 P) + (P) = 2(\sigma P)$ as divisors on $X_0(N)$, which forces $X_0(N)$ to be hyperelliptic and σP (and hence P) to be a hyperelliptic branch point.

We have recently learned that A. Tamagawa has independently proved Conjecture 1.2 by methods somewhat similar to those used in our second proof.

Finally, we give proofs in Sect. 4 of some generalizations of the Coleman–Kaskel–Ribet conjecture. For example, we determine the set of torsion points on the modular curve $X_0^+(N)$ for all primes N, where the embedding is via the unique cusp. We also study non-cuspidal embeddings of these curves, and determine the complete set of torsion packets on $X_0(N)$ and $X_0^+(N)$ when N is sufficiently large.

Acknowledgements I would like to thank my thesis adviser Robert Coleman for his continual support and encouragement. Special acknowledgements are due to Ken Ribet — the second proof I give of the CKR conjecture is based on emails from and discussions with him, and Remarks 3.11 and 3.16 are derived from his ideas. Ribet also introduced the idea of utilizing the interplay between torsion points on $X_0(N)$ and $X_0^+(N)$ while thinking about the specific case $N = 389$. In addition, he pointed out the elementary but very useful Lemma 3.5, and taught me a number of things about modular curves and Galois representations. I would also like to thank William Stein for helping me with a number of computations, Barry Mazur for suggesting the application of my results to Mordell–Weil ranks, and the referee for many helpful comments on my original manuscript. Typesetting in this paper was done in LaTeX.

2. Torsion points on $X_0(N)$

We begin with a summary of the paper [6] by Coleman, Kaskel, and Ribet.

The basic approach in [6] is to use the Chinese remainder theorem to decompose the image P in $J_0(N)$ of a torsion point Q on the modular curve $X_0(N)$ (N prime) as a sum of its l-primary components, $P := i_\infty(Q) = \sum P_l$ (where $P_l \in J_0(N)$), and to try to show that P_l is in the cuspidal group for as many primes l as possible. The cuspidal group is the cyclic group C

of order $n = \mathrm{Num}(\frac{N-1}{12})$ generated by $i_\infty(0)$, which Mazur proves in [18, Theorem (1)] is the full group of rational torsion points on $J_0(N)$.

The following proposition follows from the main result of [19]; see [6, Proof of Proposition 1.2] for a proof.

Proposition 2.1. *The set of points on $X_0(N)$ mapping under i_∞ to C is just the set $\{0, \infty\}$ of cusps.*

Let $\mathbf{T} = \mathbf{T}_0(N)$ denote the full Hecke algebra for $X_0(N)$; it is precisely the ring of endomorphisms of $J_0(N)$ (see [18, II, Proposition 9.5]). The main general result in [6] is the following theorem ([6, Theorem 1.3]), which handles "most" l-primary components:

Theorem 2.2. *Let Q be an element of $T_\infty(X_0(N))$, and let $l \neq 2, 3$ be a prime for which P_l does not belong to the cuspidal group C. Then at least one of the following holds: (i) $l = N$; (ii) l satisfies $5 \leq l < 2g$, $X_0(N)$ does not have ordinary reduction at l, and l is ramified in $\mathbf{T}$ (in the sense that $\mathbf{T}/l\mathbf{T}$ is not a product of fields).*

The proof is based on Coleman's theory of p-adic integration (see [6] and [5]) plus the following theorem [6, Theorem 2.2] proved using the techniques of [18]:

Theorem 2.3. *Suppose $l \neq 2$, and that P is a torsion point on $J_0(N)$. If P is unramified at l, then $P_l \in C$.*

Ken Ribet's papers [26], [27] suggest additional techniques for tackling the Coleman–Kaskel–Ribet conjecture. Many of Ribet's results involve a certain hypothesis (*), which we now explain. We recall that the Hecke algebra $\mathbf{T} = \mathbf{T}_0(N)$ has the property that $\mathbf{T} \otimes \mathbb{Q}$ is a product of totally real number fields K_i, and $\mathbf{T}$ itself has finite index in its normalization $\tilde{\mathbf{T}}$, which is the product of the maximal orders $\mathcal{O}_i$ of K_i. By the discriminant of $\mathbf{T}$, we mean the product of the discriminants of the K_i multiplied by the square of the index of $\mathbf{T}$ in $\tilde{\mathbf{T}}$. By definition, $\mathbf{T}$ is unramified at a prime l if l does not divide the discriminant of $\mathbf{T}$; this is equivalent to saying that $\mathbf{T}/l\mathbf{T}$ is a product of finite fields. Ribet's auxiliary hypothesis is:

(*) The prime N is unramified in the Hecke algebra $\mathbf{T}$.

William Stein has done computer-aided computations (see [32]) which establish the following proposition:

Proposition 2.4. *Condition (*) is satisfied by all $N < 5000$ except for $N = 389$. The prime number 389 is ramified in $\mathbf{T}_0(389)$, but unramified in $\mathbf{T}_0^+(389)$ (which we will define shortly).*

One result (Theorem 1.6 of [26]) which involves hypothesis (*) is:

Theorem 2.5. *Let N and $l \neq N$ be prime numbers. Suppose $P \in J_0(N)^{\mathrm{tor}}$ is such that its l-primary component P_l is not contained in the cuspidal group C (which is equivalent, by [18, Theorem (1)], to supposing that $P_l \notin J_0(N)(\mathbb{Q})$). Assume either that N does not divide the order of P or that N satisfies hypothesis (*). Then there is an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order l in $J_0(N)$.*

It is useful to exploit the interplay between torsion points on $X_0(N)$ and its quotient curve $X_0^+(N)$. The intuitive reason why $X_0^+(N)$ is in many ways simpler than $X_0(N)$ is that it is "non-Eisenstein"; we will make this more precise in a moment.

We discuss now some facts about $X_0^+(N)$ that we will use in what follows.

There is a unique cusp at infinity on $X_0^+(N)$, which we denote by ∞^+, or simply ∞ if no confusion is likely to arise, and the fiber of the map $\pi : X_0(N) \to X_0^+(N)$ over ∞^+ is just $\{0, \infty\}$. Let $J_0^+(N)$ be the Picard (Jacobian) variety of $X_0^+(N)$. The fact that $J_0^+(N)$ is also the Albanese variety of $X_0^+(N)$ implies there is a commutative diagram

$$\begin{array}{ccc} X_0(N) & \xrightarrow{i_\infty} & J_0(N) \\ \pi \downarrow & & \downarrow \pi_* \\ X_0^+(N) & \xrightarrow{i_\infty} & J_0^+(N) \end{array}$$

where $i_\infty : X_0^+(N) \to J_0^+(N)$ is the map which on closed points takes Q to $[(Q) - (\infty^+)]$. The map π_* takes a point in $J_0(N)$ represented by the degree zero divisor $\sum P_i - \sum Q_i$ to the class of the divisor $\sum \pi(P_i) - \sum \pi(Q_i)$, thought of as a point of $J_0^+(N)$. Note that a point of $X_0(N)$ mapping to a torsion point of $J_0(N)$ is sent by π to a point of $X_0^+(N)$ mapping to a torsion point of $J_0^+(N)$.

The map $\pi^* : J_0^+(N) \to J_0(N)$ induced by Picard functoriality is a closed immersion (since w_N is a degree 2 automorphism with fixed points), so π^* identifies $J_0^+(N)$ with an abelian subvariety of $J_0(N)$. The composite map $\pi^* \circ \pi_* : J_0(N) \to J_0(N)$ is easily seen to be the map $1 + w$, so that $J_0^+(N)$ is naturally identified with the subvariety $J_+ := (1 + w) J_0(N)$ of $J_0(N)$ (see [18, II, Sect. 10] for another discussion of this).

We also note that $J_- := (1 - w) J_0(N)$ is naturally identified with the kernel of multiplication by $1+w$ on $J_0(N)$. Indeed, J_- is certainly contained in this kernel; in fact, for dimension reasons J_- is the connected component of the identity in this kernel. But $\ker(1 + w)$ is connected (this is equivalent to the fact that the map π^* is injective).

For future reference, we define the Hecke algebra $\mathbf{T}^+$ to be the image of $\mathbf{T}$ in the endomorphism ring of $J_0^+(N)$ (thought of as the subvariety $(1 + w) J_0(N)$ of $J_0(N)$).

Recall that C is the cuspidal subgroup of $J_0(N)$, which is the cyclic subgroup generated by the point $c := i_\infty(0) \in J_0(N)$, i.e., by the divisor $(0) - (\infty)$. Since $w(c) = -c$, $1 + w$ (and hence π_*) annihilates C.

Furthermore, let $\mathfrak{I}$ be the Eisenstein ideal of $\mathbf{T}$ (see [18, II, Sect. 9]). It is the ideal generated by $p+1-T_p$ for p not dividing N and by $1+w$. The kernel $J_0(N)[\mathfrak{I}]$ of the Eisenstein ideal is a finite Galois module containing C. It, too, is annihilated by π_*, since $1+w \in \mathfrak{I}$. It follows that if $\mathfrak{m}$ is any maximal ideal of $\mathbf{T}$ containing $\mathfrak{I}$ (i.e., $\mathfrak{m}$ is Eisenstein), $\pi_*(J_0(N)[\mathfrak{m}]) = 0$.

A stronger way of expressing the fact that $X_0^+(N)$ is "non-Eisenstein" is to say that $J_0^+(N)[\mathfrak{m}]=0$ whenever $\mathfrak{m}$ is an Eisenstein prime. When the residue characteristic p of $\mathfrak{m}$ is different from 2, this is clear, since w acts as $+1$ on $J_0^+(N)$ and as -1 on $J_0(N)[\mathfrak{m}]$. When $p = 2$ this is more subtle and is established in the proof of [18, II, Proposition 17.10].

Along similar lines, we have the following proposition.

Proposition 2.6. *Let p be an odd prime. The Jordan–Hölder factors (as a module for $\mathbf{T}^+[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$) of $J_0^+(N)[p]$ are all two-dimensional, and are isomorphic to the representations $J_0(N)[\mathfrak{m}]$ for $\mathfrak{m}$ a (non-Eisenstein) maximal ideal of $\mathbf{T}$ containing $1 - w$.*

Proof. Let V be such a Jordan–Hölder factor – its annihilator $\mathfrak{m}'$ is a maximal ideal of $\mathbf{T}^+$ of characteristic p. Clearly $1 - w \in \mathfrak{m}'$, and since $p \neq 2$, $1 + w \notin \mathfrak{m}'$. The inverse image of $\mathfrak{m}'$ in $\mathbf{T}$ is a maximal ideal $\mathfrak{m}$ containing $1 - w$ but not $1 + w$. We claim that $J_0(N)[\mathfrak{m}] = J_0^+(N)[\mathfrak{m}] = V$. Indeed, V is a subquotient of $J_0^+(N)[\mathfrak{m}]$, and hence of $J_0(N)[\mathfrak{m}]$, and it is stable under the action of $\mathbf{T}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$. Since $\mathfrak{m}$ is not Eisenstein, $J_0(N)[\mathfrak{m}]$ is irreducible and two-dimensional by [18, II, Proposition 14.2]. We must then have $V = J_0^+(N)[\mathfrak{m}] = J_0(N)[\mathfrak{m}]$ as claimed. □

For each maximal ideal $\mathfrak{m}$ of $\mathbf{T}^+$, we can form the $\mathfrak{m}$-divisible group $J_0^+(N)_\mathfrak{m} := \cup J_0^+(N)[\mathfrak{m}^i]$. If the residue characteristic p of $\mathfrak{m}$ is different from 2, then the $\mathfrak{m}$-adic Tate module $\mathrm{Hom}(\mathbb{Q}_p/\mathbb{Z}_p,\ J_0(N)^+_\mathfrak{m})$ is free of rank 2 over $\mathbf{T}^+_\mathfrak{m}$. This follows on replacing $X_0(N)$ and $J_0(N)$ by $X_0^+(N)$ and $J_0^+(N)$ in the proof of [18, II, Lemma 15.1].

We also have the following result about $J_0^+(N)$ (compare with Theorem 2.5).

Theorem 2.7. *Let N and $l \neq N$ be prime numbers. Suppose $P \in J_0^+(N)^{\mathrm{tor}}$ is such that its l-primary component P_l is nonzero. Assume either that N does not divide the order of P or that N is unramified in $\mathbf{T}^+$. Then there is an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order l in $J_0^+(N)$.*

Proof. If N is unramified in $\mathbf{T}$, this follows immediately from Theorem 2.5 from the fact that $J_0^+(N)$ can be thought of as a subvariety of $J_0(N)$ whose intersection with $J_0(N)[\mathfrak{I}]$ is trivial. Otherwise, we note that Ribet's proof of Theorem 2.5 follows *mutatis mutandis* for $J_0^+(N)$. In fact, Theorem 2.7 is actually easier to prove than Theorem 2.5: in view of Theorem 3.12 of

this paper, the case in Ribet's proof in which P_l is unramified at N does not occur. □

3. Proof of the Coleman–Kaskel–Ribet conjecture

In this section we prove the Coleman–Kaskel–Ribet conjecture. In fact, we give two proofs of this conjecture, with the hope that the ideas used in both proofs will be useful in other contexts. In the following section we apply similar arguments to determine torsion points on $X_0^+(N)$ in the cuspidal embedding, and also to study arbitrary torsion packets on $X_0(N)$ and $X_0^+(N)$.

The key idea in the first proof of the CKR conjecture is to use π to project torsion points on $X_0(N)$ to torsion points on $X_0^+(N)$, and then to use the fact that $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts in a particularly simple way on torsion points of $J_0^+(N)$.

Before beginning the proof, we collect here some facts which we will need. The first result concerns maps of low degree from $X_0(N)$ and $X_0^+(N)$ to $\mathbf{P}^1$.

Theorem 3.1. *Let $N \geq 23$ be a prime number.*

1. *$X_0(N)_{\mathbb{C}}$ is hyperelliptic (admits a degree 2 map to $\mathbf{P}^1_{\mathbb{C}}$) iff $N \in \{23, 29, 31, 37, 41, 47, 59, 71\}$.*
2. *$X_0^+(N)_{\mathbb{C}}$ is hyperelliptic iff $g_0^+(N) = 2$ iff $N \in \{67, 73, 103, 107, 167, 191\}$.*
3. *$X_0(N)_{\mathbb{C}}$ is trigonal (admits a degree 3 map to $\mathbf{P}^1_{\mathbb{C}}$) iff $N \in \{23, 29, 31, 37, 43, 53, 61\}$.*
4. *If $X_0^+(N)_{\mathbb{C}}$ is trigonal, then $N \leq 311$.*
5. *If $X_0(N)_{\mathbb{C}}$ admits a map of degree at most 4 to $\mathbf{P}^1_{\mathbb{C}}$ then $N \leq 191$.*
6. *If $X_0^+(N)_{\mathbb{C}}$ admits a map of degree at most 4 to $\mathbf{P}^1_{\mathbb{C}}$ then $N \leq 479$.*
7. *If $X_0(N)_{\mathbb{C}}$ admits a map of degree at most 6 to $\mathbf{P}^1_{\mathbb{C}}$ then $N \leq 311$.*
8. *If $X_0^+(N)_{\mathbb{C}}$ admits a map of degree at most 6 to $\mathbf{P}^1_{\mathbb{C}}$ then $N \leq 911$.*

Proof. Part (1) follows from the main result of [21], and part (2) from the main result of [11]. The rest of the assertions are proved in Chap. 3 of [1]. We note that similar assertions have recently been proved by Hasegawa–Shimura (see [12], [13]) and by Nguyen–Saito (see [20]). □

The following theorem is proved in Chap. 5 of [1] by a potpourri of techniques. For the reader's benefit, we remark that it is also a consequence of the second proof we give of the Coleman–Kaskel–Ribet conjecture, which unlike our first proof does not treat the trigonal modular curves $X_0(N)$ as exceptional cases.

Theorem 3.2. *The CKR conjecture is true for the trigonal modular curves $X_0(N)$, i.e., the curves $X_0(N)$ with $N \in \{23, 29, 31, 37, 43, 53, 61\}$.*

From now on, N will always denote a prime number such that $g_0(N)$ (the genus of $X_0(N)$) is at least 1. Often in our applications $g_0(N)$ will in fact be at least 2.

If p is a prime, we denote by $T_p(J_0(N))$ the p-adic Tate module of $J_0(N)$. If $p > 2$, the results of [18, II, Sects. 14–15] show that $T_p(J_0(N))$ is free of rank 2 over $\mathbf{T}_p := \mathbf{T} \otimes \mathbb{Z}_p$, where $\mathbf{T}$ is the Hecke algebra associated to $J_0(N)$. Choosing a basis for $T_p(J_0(N))$ over $\mathbf{T}_p$, we can view the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $T_p(J_0(N))$ as providing a continous representation

$$\rho = \prod \rho_{\mathfrak{m}_i} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}_p) \cong \prod \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}_i}),$$

where the $\mathfrak{m}_i$ are the maximal ideals of $\mathbf{T}$ lying over p. We also have a continuous representation

$$\overline{\rho} = \prod \overline{\rho}_{\mathfrak{m}_i} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \prod \mathrm{GL}(2, \mathbf{T}/\mathfrak{m}_i) \cong \mathrm{GL}(2, \prod \mathbf{T}/\mathfrak{m}_i).$$

For $J_0^+(N)$, we obtain analogous representations

$$\rho^+ = \prod \rho^+_{\mathfrak{m}_i} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}_p^+) \cong \prod \mathrm{GL}(2, \mathbf{T}^+_{\mathfrak{m}_i}),$$

and

$$\overline{\rho}^+ = \prod \overline{\rho}^+_{\mathfrak{m}_i} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \prod \mathrm{GL}(2, \mathbf{T}^+/\mathfrak{m}_i) \cong \mathrm{GL}(2, \prod \mathbf{T}^+/\mathfrak{m}_i).$$

Let $\mathbf{F}$ be the product of finite fields $\prod \mathbf{T}/\mathfrak{m}_i$, and let $\mathbf{F}^+$ be $\prod \mathbf{T}^+/\mathfrak{m}_i$.

By an inertia group at p, we mean the inertia subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ at some prime lying over p. By a wild inertia group at p, we mean the p-Sylow subgroup of an inertia group at p of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$; see [30, Sect. 1.4] for a discussion of p-Sylow subgroups of profinite groups. If $p \neq N$, let X_p be the normal closure inside $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ of an inertia group at p, and if $p = N$ let X_p be the normal closure of a wild inertia group at N.

Also, if $p \neq N$, then set $\Gamma_p = \mathbb{Z}_p^*$, and set $\Gamma_N = 1 + N\mathbb{Z}_N \subseteq \mathbb{Z}_N^*$. Let $\overline{\Gamma}_p$ be the reduction of Γ mod p; i.e., $\overline{\Gamma}_p$ equals $(\mathbb{Z}/p\mathbb{Z})^*$ when $p \neq N$ and 1 when $p = N$.

We have the following results concerning the images $\rho(X_p)$ and $\overline{\rho}(X_p)$:

Theorem 3.3. *Suppose $p \geq 5$.*

1. *Assume p does not divide $N - 1$. Then the image $\overline{\rho}(X_p)$ of X_p is the group of matrices in $\mathrm{GL}(2, \mathbf{F})$ having determinant in $\overline{\Gamma}_p$ (embedded diagonally in $\mathbf{F}$).*
2. *Assume, in addition to $p \nmid N - 1$, that p is unramified in $\mathbf{T}$. Then $\rho(X_p)$ is the group of matrices in $\mathrm{GL}(2, \mathbf{T}_p)$ having determinant in Γ_p.*
3. *The image $\overline{\rho}^+(X_p)$ is the group of matrices in $\mathrm{GL}(2, \mathbf{F}^+)$ having determinant in $\overline{\Gamma}_p$.*
4. *Assume that p is unramified in $\mathbf{T}$. Then $\rho^+(X_p)$ is the group of matrices in $\mathrm{GL}(2, \mathbf{T}_p^+)$ having determinant in Γ_p.*

Proof. When $p \neq N$, part (1) follows by combining Proposition 6.3 and Theorem 3.4 of [27]. When $p = N$, the assertion of part (1) is established during the proof of [26, Proposition 6.4]. Part (2) is proved in the remark following [27, Theorem 6.4] when $p \neq N$, and in the proof of [26, Proposition 6.4] when $p = N$.

The assertions in (3) and (4) are proved similarly. We provide the reader with the following guide for translating the required results for $J_0(N)$ to results for $J_0^+(N)$.

To prove part (3) when $p \neq N$, we need to check that the hypotheses of Proposition 6.3 and Theorem 3.4 of [27] are satisfied by $\overline{\rho}^+$. It suffices to note that all $\overline{\rho}^+_{\mathfrak{m}_i}$ are irreducible (since there are no Eisenstein primes in $\mathbf{T}^+$), and that $\mathbf{F}^+ = \prod \mathbf{T}^+/\mathfrak{m}_i$ is generated by the traces of elements $\overline{\rho}^+(\sigma)$ for $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ (since the analogous statement is true for $\mathbf{T}$). When $p = N$, part (3) also follows from Proposition 6.3 and Theorem 3.4 of [27] (with the hypotheses verified as above), together with the argument in the proof of Proposition 6.4 of [26]. Finally, part (4) follows from part (3) together with Proposition 4.2 of [27]. □

Remark 3.4. It is a consequence of the results of [16] that the assertions of parts (2) and (4) of Theorem 3.3 are true even without the hypothesis that p is unramified in $\mathbf{T}$ or $\mathbf{T}^+$, respectively. We will not use the results of [16] in any essential way in this paper, though appealing to them makes some of our arguments shorter.

The proof of the following lifting lemma is reminiscent of the techniques of [28, IV-23, Lemma 3].

Lemma 3.5. *Let p be an odd prime, let A be a commutative ring with identity, and let R be a nilpotent ideal in A containing p. If H is a subgroup of $\mathrm{GL}(n, A)$ whose image $\overline{H}$ in $\mathrm{GL}(n, A/R)$ contains the homothety -1, then H contains -1.*

Proof. We are given that $\overline{H}$ contains -1. This means that $h \equiv -1 \bmod R$, i.e., there exist $h \in H$ and $r \in M(n, R)$ such that $h = -1 + r$. Since -1 and r commute, one can use the binomial theorem to see that $h^p \equiv -1 \bmod R^2$, and more generally $h^{p^i} \equiv -1 \bmod R^{i+1}$. Since $R^i = 0$ for i sufficiently large, it follows that the group H contains -1. □

Though we originally conceived of the next result as an application of the results of [16], it follows from the much easier Theorem 3.3(1), together with the elementary Lemma 3.5.

Proposition 3.6. *Let $p \geq 5$ be prime, and suppose that p does not divide $N-1$. Then the image $\rho(X_p) \subseteq \mathrm{GL}(2, \mathbf{T} \otimes \mathbb{Z}_p)$ contains the homothety -1.*

Proof. Let $T = \mathbf{T} \otimes \mathbb{Z}_p$, and for $n \geq 1$, let A be the Artinian ring $T/p^n T$. If H denotes the image of $\rho(X_p)$ in $\mathrm{GL}(2, A)$, then it suffices to prove that $-1 \in H$ for all n. Let R be the radical of A, i.e., the set of nilpotent elements

in A. Then $p \in R$, and A/R is isomorphic to $\mathbf{F} := \prod T/\mathfrak{m}_i$, a product of finite fields of characteristic p. By Lemma 3.5, it suffices to prove that the image $\overline{H}$ of H inside $\mathrm{GL}(2, A/R)$ contains -1. In fact, $\overline{H}$ contains all of $\mathrm{SL}(2, A/R)$, as follows from Theorem 3.3(1). □

Similarly, we have the following:

Proposition 3.7. *Let $p \geq 5$ and N be primes such that $g_0^+(N) > 0$. (We allow the case where p divides $N-1$). Then the image $\rho(X_p) \subseteq \mathrm{GL}(2, \mathbf{T}^+ \otimes \mathbb{Z}_p)$ contains the homothety -1.*

Proof. This is proved in the same way as the previous Proposition, replacing the reference to Theorem 3.3(1) by a reference to Theorem 3.3(3). □

Corollary 3.8. *If $p \geq 5$ and N are prime numbers with $g_0^+(N) > 0$, then there exists a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which acts as -1 on all torsion points of $J_0^+(N)$ of p-power order and as $+1$ on torsion points of order prime to p.*

Proof. For $p \neq N$, this follows from Proposition 3.7, together with the criterion of Néron–Ogg–Shafarevich. At N this follows from the fact that elements of inertia groups at N act unipotently on prime-to-N torsion, so that the image $\rho_l(I_N)$ of any wild inertia group at N under an l-adic representation with $l \neq N$ is both pro-l and pro-N, hence trivial. □

We also need the following easy lemma.

Lemma 3.9. *Let X be a curve of genus at least 2 mapping to its Jacobian via $\phi : P \mapsto [(P) - (P_0)]$ for some fixed $P_0 \in X$. If there exists a point $P \neq P_0$ on X such that $-\phi(P)$ is in the image of X, then X is hyperelliptic and P_0 is a hyperelliptic branch point.*

Proof. We are given that $(P_0) - (P)$ is linearly equivalent to $(Q) - (P_0)$ for some point Q. Therefore there is a rational function on X with divisor equal to $(P) + (Q) - 2(P_0)$, and since $P \neq P_0$ this forces X to be hyperelliptic. □

We now give our first proof of the Coleman–Kaskel–Ribet conjecture.

Theorem 3.10. *Conjecture 1.2 is true for all N.*

Proof. We first prove the conjecture under the hypothesis (*), which says that N does not divide the discriminant of the Hecke algebra $\mathbf{T}$.

Suppose we have a point $Q \in X_0(N)(\overline{\mathbb{Q}})$ such that $i_\infty(Q)$ is a torsion point of $J_0(N)$. Write

$$P := i_\infty(Q) = P_2 + P_3 + P_N + \sum_{l \neq 2,3,N} P_l,$$

where P_l has l-power order in $J_0(N)$ for all primes l.

If all $P_l \in C$, then $P \in C$, which by Proposition 2.1 implies that $Q \in \{0, \infty\}$. If $P_2 \notin C$, then by Theorem 2.5 [since we are assuming (*)]

there exists a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order 2 in $J_0(N)$. Since $\sigma P - P = [(\sigma Q) - (Q)]$, it follows that $X_0(N)$ is hyperelliptic and Q is a hyperelliptic branch point. This possibility is already accounted for in the statement of the Coleman–Kaskel–Ribet conjecture. (Recall from Proposition 1.1 that the hyperelliptic branch points on $X_0(37)$ are not torsion points.)

If $P_3 \notin C$, then there exists a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order 3. This means that $X_0(N)$ is trigonal, i.e., admits a degree three morphism to $\mathbf{P}^1$. According to Theorem 3.1(3), this implies that $N \in \{23, 29, 31, 37, 43, 53, 61\}$. But Theorem 3.2 asserts that the CKR conjecture is true for these values of N.

So assume, then, that $P_2, P_3 \in C$. Let $Q^+ = \pi(Q) \in X_0^+(N)$. Since the group C is annihilated by π_*, we see that

$$P^+ = i_\infty(Q^+) = \sum_{l \neq 2,3,N} P_l^+ + P_N^+,$$

where $P_l^+ = \pi_*(P_l)$. By Corollary 3.8, we can find an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P^+ = -P^+$, so that $-i_\infty(Q^+)$ lies on the image of $X_0^+(N)$. By Lemma 3.9, this implies that $X_0^+(N)$ is sub-hyperelliptic, i.e., has genus 0 or 1 or is hyperelliptic. When $g^+ \geq 2$, it also implies that infinity is a hyperelliptic branch point on $X_0^+(N)$, which it is not (see the proof of Lemma 4.3 below).

So we can assume that $g^+ \leq 1$. One can then explicitly check that the CKR conjecture is true whenever $g^+ \leq 1$ (the largest N for which $X_0^+(N)$ has genus 0 or 1 is $N = 131$). For according to the tables in [32], for each N such that $g^+ \leq 1$ there are no primes p between 5 and $2g$ (g=genus of $X_0(N)$) which divide the discriminant of the Hecke algebra $\mathbf{T}$ and are simultaneously non-ordinary. By Theorem 2.2, this shows that $P = P_C + P_N$ with $P_C \in C$. We then have $P_N = 0$ by [6, Theorem 3.15], which shows that $P \in C$ and hence Q is a cusp.

This proves the CKR conjecture for all N satisfying hypothesis (*). We recall from Proposition 2.4 that this hypothesis is satisfied for all primes $N < 5000$ except for $N = 389$, and that 389 does not divide the discriminant of $\mathbf{T}_0^+(389)$.

Projecting a potential torsion point Q on $X_0(389)$ right away to $X_0^+(389)$, we see from the above arguments that Q is a cusp unless $X_0^+(389)$ (which has genus 11) admits a map of degree 2 or 3 to $\mathbf{P}^1$. But according to Theorem 3.1(2,4), this is not the case.

The key thing to notice in general when (*) is not necessarily satisfied is that we can still apply Theorem 2.7 to a torsion point when the order of that point is not divisible by N.

Take a torsion point

$$P = i_\infty(Q) = P_2 + P_3 + P_N + \sum_{l \neq 2,3,N} P_l$$

on the image of $X_0(N)$ as before. We now project right away to $X_0^+(N)$, so that we have $P^+ = i_\infty(Q^+) = P_2^+ + P_3^+ + P_N^+ + R^+$, where

$$R^+ = \sum_{l \neq 2,3,N} P_l^+$$

with the various P_l^+ defined in the obvious way. By Corollary 3.8, there is a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P^+ = P_2^+ + P_3^+ - P_N^+ - R$. So $P' := P^+ + \sigma P^+$ is equal to $2P_2^+ + 2P_3^+$.

Now this torsion point on $J_0^+(N)$ has order prime to N, and so Theorem 2.7 applies to it.

If $2P_2^+ \neq 0$, we find that there exists a $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $P'' := \tau P' - P'$ has order 2 in $J_0^+(N)$. We have $P'' = [(\tau Q^+) + (\tau\sigma Q^+) - (Q^+) - (\sigma Q^+)]$, so the divisor $2(\tau Q^+) + 2(\tau\sigma Q^+) - 2(Q^+) - 2(\sigma Q^+)$ is principal. This divisor is either identically zero, or else $X_0^+(N)$ admits a map to $\mathbf{P}^1$ of degree at most four. The first case is impossible, because it implies that either $Q^+ = \tau Q^+$ and $\sigma Q^+ = \tau\sigma Q^+$, or $Q^+ = \tau\sigma Q^+$ and $\sigma Q^+ = \tau Q^+$, but either way we would have $P'' = 0$ whereas we assumed P'' had order 2. So $X_0^+(N)$ admits a map of degree at most four to $\mathbf{P}^1$.

Similarly, if $2P_3^+ \neq 0$, we find that $X_0^+(N)$ admits a map of degree at most six to $\mathbf{P}^1$. This implies that $N \leq 911$ by Theorem 3.1(8). Since we have already established the CKR conjecture for primes less than 5000, we reduce to the case where $P' = 0$. But then $P^+ = -\sigma P^+$, hence $-i_\infty(Q^+)$ is in the image of $X_0^+(N)$ and by Lemma 3.9, $X_0^+(N)$ is sub-hyperelliptic. But we have already dispensed of this case, so our first proof of the Coleman–Kaskel–Ribet conjecture is complete. □

Remark 3.11. Since the proof of [6, Theorem 3.15] is rather complicated, the reader may prefer the following argument to see that $P_N = 0$ when $g^+ \leq 1$. According to [27, Theorem 3.2] (or the proof of [26, Proposition 6.4]), if $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}_N)$ is the representation giving the action of Galois on the Tate module $T_N(J_0(N))$ and $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}/N\mathbf{T})$ is its reduction mod N, then the image of $\overline{\rho}$ is precisely $\{M \in \mathrm{GL}(2, \mathbf{T}/N\mathbf{T}) | \det(M) \in (\mathbb{Z}/N\mathbb{Z})^*\}$. So if $P_N \neq 0$, then it is easy to see that P_N (and hence P) has at least $N^2 - 1$ Galois conjugates (see [1, Lemma 5.1] for the argument). On the other hand, since N is unramified in $\mathbf{T}$, we know from [26, Proposition 6.4] that the image of ρ equals $\{M \in \mathrm{GL}(2, \mathbf{T}_N) | \det(M) \in \mathbb{Z}_N^*\}$. In particular, for each integer d such that $(d, N) = 1$, there is an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that σ acts on P_N as the homothety d. We take $d = (1-n)$; in this case, since P_C is killed by n and σ acts trivially on P_C, it follows that σ acts as the homothety $1-n$ on P itself. An intersection theory argument (see [6, Proposition 4.2]) shows that P can have at most $gd^2 = g(n-1)^2$ Galois conjugates, where g is the genus of $X_0(N)$. Therefore if $P_N \neq 0$, the inequality $N^2 - 1 \leq g(n-1)^2$ holds. Using the easy estimates $g \leq n-1$ and $n < \frac{N}{12}$, we find that $N^2 - 1 < \frac{N^3}{12^3} - 1$, and

so $N \geq 1728$. Since we are assuming that $N \leq 131$, it follows that $P_N = 0$ as claimed.

Before discussing the second proof, we consider the special case $i_\infty(X_0(N)) \cap J_0(N)[\mathfrak{I}]$.

We begin with a review of some well-known facts about $J_0(N)[\mathfrak{I}]$, the kernel of the Eisenstein ideal; for proofs see [18, Chap. II] and [26, Sect. 3]. As a group, $J_0(N)[\mathfrak{I}]$ has order n^2, where $n = \mathrm{Num}\frac{N-1}{12}$, and as a Hecke module, $J_0(N)[\mathfrak{I}]$ is free of rank 2 over $\mathbf{T}/\mathfrak{I} \cong \mathbb{Z}/n\mathbb{Z}$. Also, $J_0(N)[\mathfrak{I}]$ contains both the cuspidal subgroup C and the Shimura subgroup Σ. When n is odd, $J_0(N)[\mathfrak{I}]$ is in fact equal to the direct sum of C and Σ. When n is even, however, the sum $C + \Sigma$ is no longer direct and has index 2 in $J_0(N)[\mathfrak{I}]$. The Galois action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on C is trivial and on Σ is given by the cyclotomic character $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \twoheadrightarrow \mathrm{Gal}(\mathbb{Q}(\mu_n)/\mathbb{Q}) \xrightarrow{\sim} (\mathbb{Z}/n\mathbb{Z})^*$.

We have the following very useful result:

Theorem 3.12. *$J_0(N)[\mathfrak{I}]$ is exactly the set of torsion points of $J_0(N)$ which are unramified at N. On $J_0^+(N)$, there are no nonzero torsion points unramified at N.*

Proof. The first statement is proved in [26, Proposition 3.1, Proposition 3.3]. The second statement follows from the same proof; it is in fact easier to prove than the first statement, so for the reader's convenience we give a proof here. Suppose that P is a nonzero torsion point in $J_0^+(N)(\overline{\mathbb{Q}})$, and let M be the $\mathbf{T}^+[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-submodule of $J_0^+(N)(\overline{\mathbb{Q}})$ generated by P. Let V be a Jordan–Hölder constituent of M, and let $\mathfrak{m}$ be its annihilator, which is a maximal ideal in $\mathbf{T}^+$. By Proposition 2.6, V is isomorphic to $J_0(N)[\mathfrak{m}]$ for some non-Eisenstein maximal ideal $\mathfrak{m}$ of $\mathbf{T}$ having characteristic p. Furthermore, [26, Proposition 2.2] and the discussion preceding it show that all such $J_0(N)[\mathfrak{m}]$ are ramified at N (since if $p \neq N$, then this is equivalent to the statement that $J_0(N)[\mathfrak{m}]$ is not finite at p, and if $p = N$, the determinant of $J_0(N)[\mathfrak{m}]$ is the mod N cyclotomic character, which is ramified at N). It follows that P is ramified at N. □

Lemma 3.13. *Suppose $X_0(N)$ is hyperelliptic and that Q is a hyperelliptic branch point. Then $i_\infty(Q) = [(Q) - (\infty)] \notin J_0(N)[\mathfrak{I}]$.*

Proof. When $N = 37$, the hyperelliptic branch points do not map to torsion points of $J_0(N)$ at all by Proposition 1.1. So we can assume that $N \neq 37$. In this case, the hyperelliptic involution coincides with the Atkin–Lehner involution w.

One way to conclude is to note that a hyperelliptic branch point corresponds to an N-isogeny $E \to E$, where E is an elliptic curve with complex multiplication by an order in the ring of integers of $\mathbb{Q}(\sqrt{-N})$. The field of definition of this point contains $\mathbb{Q}(\sqrt{-N})$, which is ramified at N. Hence $i_\infty(Q)$ cannot be in $J_0(N)[\mathfrak{I}]$, which is unramified at N.

Here is an alternative argument. Let $P = i_\infty(Q) \in J_0(N)^{\text{tor}}$. If $P \in C$ then we are done, since by Proposition 2.1, $i_\infty^{-1}(C)$ consists only of the cusps, which are not hyperelliptic branch points. (In fact, the cusps on $X_0(N)$ are never Weierstrass points, see [22]). Otherwise Q is not rational, so there is some $\sigma \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $Q' = \sigma Q \neq Q$ is another hyperelliptic branch point. Then the divisor $(Q)-(Q')$ has order 2 in $J_0(N)$. On the other hand, $i_\infty(Q) = [(Q)-(Q')]+i_\infty(Q')$, so at least one of $P = i_\infty(Q)$, $P^\sigma = i_\infty(Q')$ has even order; since these points are conjugate, both have even order. Since w is the hyperelliptic involution on $X_0(N)$, which acts on $J_0(N)$ as -1, we have $[(Q)-(\infty)] = [(w\infty)-(wQ)] = [(0)-(Q)]$ as elements of $J_0(N)$. Adding $[(Q)-(\infty)]$ to both sides, we get $2P = [(0)-(\infty)]$, which is a generator of the cyclic group C of order n. If $P \in J_0(N)[\mathfrak{I}]$, then the order of P divides n. But we have just shown that $2P$ has order n, which is a contradiction: since the order of P is even, the order of P is twice the order of $2P$. □

Lemma 3.14. *If m is a positive integer not dividing 6, then there exist elements $a, b \in (\mathbb{Z}/m\mathbb{Z})^*$ with $a+b=2$ and $a \neq 1, b \neq 1$.*

Proof. By the Chinese remainder theorem, the result is true for m if it is true for at least one prime power p^t exactly dividing m. By assumption we can choose such a $p^t > 3$. If $p \neq 3$, then -1 and 3 satisfy the requirements of the lemma. Otherwise, if $p = 3$, we can take a and b to be -2 and 4. □

Proposition 3.15. *Let $N \geq 23$ be prime. The only points $Q \in X_0(N)(\overline{\mathbb{Q}})$ such that $P = [(Q)-(\infty)]$ lies in $J_0(N)[\mathfrak{I}]$ are 0 and ∞.*

Proof. We provide two proofs of this result. First, we note that if $P \in J_0(N)[\mathfrak{I}]$, then under the projection π_*, P is sent to zero. Therefore, when the genus of $X_0^+(N)$ is positive (so that the map $X_0^+(N) \to J_0^+(N)$ is an embedding), we have $Q = 0$ or $Q = \infty$ as desired. The genus of $X_0^+(N)$ is zero exactly when $X_0(N)$ is hyperelliptic and $N \neq 37$, i.e., when $N = 23, 29, 31, 41, 47, 59,$ or 71.

Suppose, then, that $P \in J_0(N)[\mathfrak{I}]$ (so its order divides $n = \text{Num}\frac{N-1}{12}$) and that $g_0^+(N) = 0$. Let g be the genus of $X_0(N)$. For each prime N in the above list, one can check, using [32], that n is prime to 3, and that there are no primes between 5 and $2g$ which are simultaneously non-ordinary and ramified in the Hecke algebra. By Theorem 2.2 (which is based on p-adic integration techniques of [5]), it follows that $P = P_2 + P_C$, with P_2 of 2-power order and $P_C \in C$.

If $P_2 \notin C$, then by Theorem 2.5 there exists a σ in an inertia group for 2 in $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order 2. This means that the divisor $2(\sigma Q) - 2(Q)$ is principal, so Q is a hyperelliptic branch point. But the hyperelliptic branch points of $X_0(N)$ do not map to $J_0(N)[\mathfrak{I}]$ by Lemma 3.13.

So we must have $P_2 \in C$ and therefore $P \in C$. But as we know from Proposition 2.1, the set of points on $X_0(N)$ mapping to C is always equal to $\{0, \infty\}$. This proves the result.

Here is another proof, which does not rely on any facts about ramified torsion points on curves derived from [5].

We again may assume, after projecting to $X_0^+(N)$, that $X_0(N)$ is hyperelliptic with $N \neq 37$. And proceeding as above we see that $P_2 \in C$, or else $X_0(N)$ would be hyperelliptic and Q would be a hyperelliptic branch point of $X_0(N)$, which is impossible. It follows that $P \in C + \Sigma$.

Since $i_\infty^{-1}(C) = \{0, \infty\}$, we may assume that $P_2 \in C$ but $P \notin C$, and therefore we may write $P = P_C + P_\Sigma$, where $P_C \in C$ and $P_\Sigma \in \Sigma$ is nonzero and of odd order m.

In fact, we may assume that $m > 3$, because n is prime to 3 for all N such that $g^+ = 0$, a fact we have already noticed above.

Since C has a trivial Galois action, it is easy to see that $(\sigma - 1)P = (\sigma - 1)P_\Sigma$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Also, since $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on Σ via the mod n cyclotomic character, it follows that for any $\mu \in (\mathbb{Z}/m\mathbb{Z})^*$ we can find $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P_\Sigma = \mu P_\Sigma$.

We conclude from Lemma 3.14 that there exist $\sigma, \tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $(\sigma - 1)P_\Sigma + (\tau - 1)P_\Sigma = 0$ but $(\sigma - 1)P_\Sigma$ and $(\tau - 1)P_\Sigma$ are nonzero. It follows that $(\sigma Q) + (\tau Q) - 2(Q)$ is a nonzero principal divisor on $X_0(N)$, and hence that Q is a hyperelliptic branch point. But this contradicts Lemma 3.13. □

With this proposition in hand, we give the second proof of the Coleman–Kaskel–Ribet conjecture.

Proof. Suppose $Q \in X_0(N)(\overline{\mathbb{Q}})$ maps to a torsion point P of $J_0(N)$. If $P \in J_0(N)[\mathfrak{I}]$, then $Q \in \{0, \infty\}$ by Proposition 3.15. So we can assume that $P \notin J_0(N)[\mathfrak{I}]$, which by Theorem 3.12 implies that P is ramified at N. We claim that there is an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P \neq P$ but $(\sigma - 1)^2 P = 0$. Given this, it is straightforward to conclude: in terms of divisors this means that $(\sigma^2 Q) + (Q) - 2(\sigma Q)$ is linearly equivalent to zero. Therefore $X_0(N)$ is hyperelliptic and σQ (and hence Q) is a hyperelliptic branch point.

To prove the claim, we first assume that N is prime to the order of P. In this case, we use the fact that P is ramified at N to find an inertia group I at N and an element $\sigma \in I$ such that $\sigma P \neq P$. By Grothendieck's Galois criterion for semistable reduction (see [10, Proposition 3.5], and also [26, (2.4)]), $(\sigma - 1)^2 P = 0$ as desired.

If N divides the order of P, write $P = P_N + P^N$ with P_N of N-power order and P^N of order prime to N. We claim that there exists a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which fixes P^N but not P_N such that $(\sigma - 1)^2 P_N = 0$. It then follows that $\sigma P \neq P$ and $(\sigma - 1)^2 P = 0$. We can find such a σ in X_N, the normal closure of a wild inertia group at N in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Elements of X_N fix prime-to-N torsion, and moreover it follows from [16] (or from Theorem 3.3(2) when N satisfies hypothesis (*)) that the image of X_N in $\mathrm{Aut}(T_N(J_0(N))) \cong \mathrm{GL}(2, \mathbf{T} \otimes \mathbb{Z}_N)$ contains $\mathrm{SL}(2, \mathbf{T} \otimes \mathbb{Z}_N)$. In particular,

there exist $\sigma_1, \sigma_2 \in X_N$ acting on an N-adic Tate module of $J_0(N)$ as

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

It is easy to see that since $P_N \neq 0$, one of σ_1, σ_2 must act nontrivially on P_N. This element σ of X_N also satisfies $(\sigma - 1)^2 P_N = 0$, so we're done. □

Remark 3.16. We can modify this proof so that results of [16], and even those of Theorem 3.3, are not needed. As we just saw, it is enough (by Grothendieck's Galois criterion for semistable reduction) to prove that there is an element σ in some inertia group for N such that $\sigma P_N \neq P_N$ but $(\sigma - 1)^2 P_N = 0$.

Fix an algebraic closure $\overline{\mathbb{Q}}_N$ of $\mathbb{Q}_N$ and an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_N$. We view $D_N = \mathrm{Gal}(\overline{\mathbb{Q}}_N/\mathbb{Q}_N)$ as a decomposition group for N inside $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, and let $I_N = \mathrm{Gal}(\overline{\mathbb{Q}}_N/\mathbb{Q}_N^{\mathrm{unr}})$ be its inertia subgroup, We also set $I'_N = \mathrm{Gal}(\overline{\mathbb{Q}}_N/\mathbb{Q}_N^{\mathrm{unr}}(\mu_{N^\infty}))$, where μ_{N^∞} denotes the set of all N-power roots of unity in $\overline{\mathbb{Q}}_N$. Denote by $m = N^r$ the order of P_N. Finally, let M be the $\mathbf{T}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-submodule of $J_0(N)[m]$ generated by P_N. Since $J_0(N)$ has toric reduction at N, there is an exact sequence of D_N-modules (compatible with the Hecke action, but we don't need this)

$$0 \to M' \to M \to M'' \to 0,$$

where I_N acts trivially on M'' and as the mod m cyclotomic character on M'. (For a more detailed discussion of this exact sequence, see [26, (2.4)] and the references cited there). Suppose that some $\sigma \in I'_N$ acts nontrivially on P_N. Since σ acts trivially on both M' and M'', the above exact sequence shows that $(\sigma - 1)^2 P_N = 0$, and we are done.

So I'_N must act trivially on M. Therefore the action of I_N on M is abelian, since it factors through the abelian group I_N/I'_N. We would like to show that this is impossible. Toward this end, let V be a Jordan–Hölder constituent of M, regarded as a $\mathbf{T}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-module. It is enough to show that the action of I_N on V is non-abelian. Let $\mathfrak{m}$ be the annihilator in $\mathbf{T}$ of V, which is a maximal ideal of $\mathbf{T}$. Since the characteristic of $\mathfrak{m}$ is N, which does not divide $N - 1$, $\mathfrak{m}$ is not an Eisenstein prime of $\mathbf{T}$, and the results of [18, Chap. II] (see [26, Sect. 2] for a more succinct discussion) show that V is irreducible and isomorphic to the standard two-dimensional representation $\rho_{\mathfrak{m}} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}/\mathfrak{m})$ attached to $\mathfrak{m}$. Now according to [26, Proposition 2.2], $\rho_{\mathfrak{m}}$ is not finite at N (in the sense of Serre's article [29]). Hence the action of I_N on $\rho_{\mathfrak{m}}$ is *très ramifiée* (and in particular non-diagonalizable) and is given matricially in the form

$$\begin{pmatrix} \chi & * \\ 0 & 1 \end{pmatrix},$$

where χ is the cyclotomic character. (For a proof see, for example, [8, Proposition 8.2] and the last paragraph in the proof of [27, Proposition 5.1]). It is then easy to see that the action of I_N on V is non-abelian, a contradiction.

Remark 3.17. A nearly identical argument shows that if $P \in J_0^+(N)$ is ramified at N, then there exists $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P_N \neq P_N$ but $(\sigma - 1)^2 P_N = 0$.

4. Generalizations

We now present some generalizations and applications of the Coleman–Kaskel–Ribet conjecture.

Proposition 4.1. *Let X be a modular curve covering some $X_0(N)$ with $g_0^+(N) > 0$; for example, X could be $X_0(NM)$ or $X_1(NM)$ for any positive integer M. Then the set of cusps on X forms a complete torsion packet.*

Proof. It follows from the work of Manin–Drinfeld [17] and Kubert–Lang [14] that the cusps of X lie in a common torsion packet. Furthermore, the fiber of $X \to X_0(N)$ over a cusp of $X_0(N)$ consists entirely of cusps. Suppose, now, that $Q \in X(\mathbb{Q})$ and that some multiple of the divisor $(Q)-(\infty)$ on X is principal. Let J be the Jacobian of X, and let $i_\infty : X \hookrightarrow J$ (resp. $X_0(N) \hookrightarrow J_0(N)$) be the Albanese embedding associated to the base point ∞. There is a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{i_\infty} & J \\ \downarrow & & \downarrow \\ X_0(N) & \xrightarrow{i_\infty} & J_0(N) \end{array}$$

which shows that the image Q' of Q in $X_0(N)$ is a torsion point on $J_0(N)$ via the mapping i_∞. Since $g_0^+(N) > 0$, we know that Q' is a cusp. Therefore its preimage Q on X is also a cusp. □

The following corollary follows directly by combining Proposition 4.1 with a theorem of Mazur proved in [31, Theorem 0.4].

Corollary 4.2. *Let $X = X_0(NM)$ or $X_1(NM)$ with $g_0^+(N) > 0$, let J be the Jacobian of X, and let $i_\infty : X \to J$ be the embedding defined by $Q \mapsto [(Q) - (\infty)]$. Fix a noncuspidal point $x \in X$ whose associated elliptic curve does not have CM. Let $\mathbb{Z}T_p(x)$ be the $\mathbb{Z}$-linear span in J of the p-Hecke points associated to x, i.e., if $T_p(x) = \sum(y_j)$, then $\mathbb{Z}T_p(x)$ is the subgroup of J generated by the $p+1$ points $i_\infty(y_j)$. Then for all sufficiently large primes p, $\mathbb{Z}T_p(x)$ has maximal rank $p+1$.*

Our next generalization concerns torsion points on $X_0^+(N)$. Let N be a prime number. When $g_0^+(N) \geq 1$, let i_∞ be the embedding of $X_0^+(N)$ into $J_0^+(N)$ defined by $Q \mapsto [(Q) - (\infty)]$. We will need the following lemma:

Lemma 4.3. *If $X_0^+(N)$ is hyperelliptic and Q is a hyperelliptic branch point, then $i_\infty(Q)$ is not a torsion point on $J_0^+(N)$.*

Proof. The hyperelliptic involution h operates on $J_0^+(N)$ as -1, so that as elements of $J_0^+(N)$ we have

$$[(Q) - (\infty)] = [(h\infty) - (hQ)] = [(h\infty) - (Q)]$$

and adding $[(Q) - (\infty)]$ to both sides of this equation,

$$2[(Q) - (\infty)] = [(h\infty) - (\infty)].$$

Therefore $i_\infty(Q)$ is a torsion point if and only if $i_\infty(h\infty)$ is a torsion point.

By Theorem 3.1(2), $X_0^+(N)$ is hyperelliptic exactly when it has genus 2. For those N for which this is the case (namely $N = 67, 73, 103, 107, 167, 191$), the image of the cusp ∞ under the hyperelliptic involution h is a noncuspidal rational point; in other words, we have $h\infty \neq \infty$. This can be seen by looking at q-expansions of weight-two cusp forms for $\Gamma_0^+(N)$, since $h\infty = \infty$ if and only if ∞ is a Weierstrass point on $X_0^+(N)$, if and only if there is a form $f = a_1 q + a_2 q^2 + a_3 q^3 + \ldots$ in the two-dimensional space $S_2(\Gamma_0^+(N), \mathbb{Q})$ such that $a_1 = a_2 = 0$. The result follows from scrutinizing the tables of [32]. (For somewhat larger prime values of N, however, it seems that ∞ usually *is* a Weierstrass point on $X_0^+(N)$. See [9] for a discussion of this.)

It is a theorem of Mazur [18, III, Corollary 1.5] that the torsion subgroup of $J_0^+(N)(\mathbb{Q})$ is zero. So $[(h\infty) - (\infty)]$, and therefore $[(Q) - (\infty)]$, has infinite order. □

Theorem 4.4. *When $g_0^+(N) \geq 2$, ∞ is the only point $Q \in X_0^+(N)(\overline{\mathbb{Q}})$ such that $i_\infty(Q) \in J_0^+(N)^{\mathrm{tor}}$. In other words, the torsion packet on $X_0^+(N)$ containing the cusp ∞ is trivial.*

Proof. We emulate the second proof of the CKR conjecture. We know by Theorem 3.12 that on $J_0^+(N)$, every nonzero torsion point is ramified at N, and therefore if $Q \neq \infty$ maps to a torsion point P on $J_0^+(N)$, P is ramified at N. Thinking of $J_0^+(N)$ as a subvariety of $J_0(N)$ (or using the remark at the end of Sect. 3), it follows from our second proof of the CKR conjecture that there exists a σ in an inertia group at N such that $\sigma P - P$ is nontrivial and $(\sigma - 1)^2 P = 0$. Hence $X_0^+(N)$ is hyperelliptic and Q is a hyperelliptic branch point. But this is impossible by Lemma 4.3. □

Our techniques extend in a rather straightforward manner to arbitrary torsion packets on $X_0(N)$ and $X_0^+(N)$.

Theorem 4.5. *If $X_0^+(N)$ has a nontrivial torsion packet, then $X_0^+(N)$ admits a map of degree at most 4 to $\mathbf{P}^1$. In particular, if $N > 479$ then every torsion packet on $X_0^+(N)$ is trivial.*

Proof. By Theorem 3.1(6), the first assertion implies the second. So we assume that $P = [(Q_1) - (Q_2)] \in J_0^+(N)^{\mathrm{tor}}$ with $Q_1 \neq Q_2$ and hope to deduce that $X_0^+(N)$ admits a map of degree at most 4 to $\mathbf{P}^1$.

The proof proceeds like our previous arguments. Since P is a nonzero torsion point on $J_0^+(N)$ it is ramified at N, and reasoning as above we can find a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P \neq P$ and $(\sigma - 1)^2 P = 0$. Therefore $(\sigma^2 Q_1) + (Q_1) + 2(\sigma Q_2) - (\sigma^2 Q_2) - (Q_2) - 2(\sigma Q_1)$ is principal. This implies that there is a rational function on $X_0^+(N)$ of degree at most 4. For if not, we would have total cancellation in the above expression. But it is easy to see that this would contradict the fact that $\sigma P \neq P$. □

For $X_0(N)$, we have the following result.

Theorem 4.6. *If $X_0(N)$ has a nontrivial torsion packet other than the cuspidal packet $\{0, \infty\}$, then $X_0(N)$ admits a map of degree at most 6 to $\mathbf{P}^1$. In particular, if $N > 311$ then every noncuspidal torsion packet on $X_0(N)$ is trivial.*

Proof. It follows from Theorem 3.1(7) that the first assertion implies the second. So we assume that $P = [(Q_1) - (Q_2)] \in J_0(N)^{\mathrm{tor}}$ with $Q_1 \neq Q_2$ and hope to deduce that $X_0(N)$ admits a map of degree at most 6 to $\mathbf{P}^1$.

The proof proceeds like our previous arguments. If $P \notin J_0(N)[\mathfrak{I}]$ then it is ramified at N, and as in the proof of Theorem 4.5 there exists a rational function on $X_0(N)$ of degree at most 4.

It remains to consider the case where $P \in J_0(N)[\mathfrak{I}]$. We may assume that $P \in C + \Sigma$; otherwise, as in the proof of Proposition 3.15, there exists $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order 2, which implies that $X_0(N)$ admits a rational function of degree at most 4. Write $P = P_C + P_\Sigma$ with $P_C \in C$ and $P_\Sigma \in \Sigma$. Let m be the order of P_Σ. If 3 divides m then there exists a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $\sigma P - P$ has order 3. It follows that $X_0(N)$ admits a rational function of order at most 6. So we can assume that 3 does not divide m.

Notice that for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, $(\sigma - 1)P = (\sigma - 1)P_\Sigma$. If $m > 2$, then Lemma 3.14 implies that there exist $\sigma, \tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $(\sigma - 1)P + (\tau - 1)P = 0$ but $(\sigma - 1)P$ and $(\tau - 1)P$ are nonzero. This easily implies that $X_0(N)$ admits a rational function of degree at most 4.

So finally, without loss of generality we assume that m divides 2, i.e., that $P \in C$. Then $(\sigma - 1)P = 0$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, so either $X_0(N)$ is hyperelliptic or $Q_1 = \sigma Q_1$ for all σ, i.e., Q_1 is defined over the rational numbers. By the main result of [19], in the latter case Q_1 (and similarly Q_2) is a cusp, unless $N = 37, 43, 67$, or 163. In each of these cases, there is a single noncuspidal rational point on $X_0(N)$, which by uniqueness is fixed by the Atkin–Lehner involution w. But it is easy to see that $Q_1 = wQ_2$ whenever $g^+ > 0$ using the fact that w acts on $J_0(N)[\mathfrak{I}]$ as -1. So in fact $Q_1 = Q_2$, and hence $P = 0$, in each of these cases. □

Corollary 4.7. *For all prime numbers $N > 311$, there is no regular differential on $X_0(N)$ vanishing to order $2g-2$ at a single point, where g denotes the genus of $X_0(N)$.*

Proof. Suppose, on the contrary, that some differential ω has divisor $(2g-2)(Q)$ for some point $Q \in X_0(N)(\overline{\mathbb{Q}})$. This certainly implies that Q is a Weierstrass point on $X_0(N)$. If Q is defined over $\mathbb{Q}$, then results of Mazur show that Q must be a cusp, but according to [22] the cusps on $X_0(N)$ are not Weierstrass points. Therefore there is some $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ such that $(\sigma\omega) = (2g-2)(\sigma Q) \neq (2g-2)(Q)$. The ratio of ω and σw is a rational function on $X_0(N)$ with divisor $(2g-2)(Q) - (2g-2)(\sigma Q)$, so Q and σQ are in the same torsion packet on $X_0(N)$. This is impossible by Theorem 4.6.

□

We also mention the following result, whose proof is nearly the same as the proofs of Theorems 4.5 and 4.6.

Theorem 4.8. *Let $X_0(N)^{(d)}$ (resp. $X_0^+(N)^{(d)}$) map to $J_0(N)$ (resp. $J_0^+(N)$) by the map i sending $\sum Q_i$ to $[\sum(Q_i) - Q']$, where $Q' = \sum(Q_i')$ is a $\mathbb{Q}$-rational point. Then if $Q \neq Q'$ in $X_0(N)^{(d)}(\overline{\mathbb{Q}})$ (resp. $X_0^+(N)^{(d)}(\overline{\mathbb{Q}})$) maps to a torsion point via i, then $X_0(N)$ (resp. $X_0^+(N)$) admits a map of degree at most $3d$ (resp. $2d$) to $\mathbf{P}^1$.*

Finally, we show how Theorems 4.5 and 4.6 can be used to obtain lower bounds for certain Mordell–Weil ranks.

Given a positive integer m, let $\eta(m)$ be the smallest positive integer n such that $\mathrm{GL}(n, \mathbb{Z})$ has a finite subgroup of order divisible by m. For example, $\eta(1) = \eta(2) = 1$, and $\eta(3) = \eta(4) = 2$. The following lemma gives an explicit lower bound for $\eta(m)$:

Lemma 4.9. *For n a positive integer, let $\beta_n(2) = n + 2[\frac{n}{2}] + \sum_{i=1}^{\infty}[\frac{n}{2^i}]$, and for p an odd prime, let $\beta_n(p) = \sum_{i=0}^{\infty}[\frac{n}{p^i(p-1)}]$. Let $\beta_n = \prod p^{\beta_n(p)}$, where the product is taken over all primes p. Finally, if m is a positive integer, let $\delta(m)$ be the smallest positive integer n such that $m \mid \beta_n$. Then $\eta(m) \geq \delta(m)$.*

Proof. See [3, Chap. IV, Theorem 2.1] for a proof, which is based upon embedding $\mathrm{GL}(n, Z)$ into $\mathrm{GL}(n, \mathbb{Z}_p)$ for each prime p, and looking at the valuations of matrix coefficients. □

As a special case of the lemma, we have the inequality $\eta(p) \geq p-1$ whenever p is a prime number.

Proposition 4.10. *Let X be a curve of genus $g \geq 2$ defined over a number field K. Assume that X has a K-rational point P_0. Let L be a finite Galois extension of K, and suppose that $P_1, \ldots, P_m \in X(L)$ is a complete set of Galois conjugates, no two of which lie in a common torsion packet on X. Let J be the Jacobian of X. Then the Mordell–Weil rank of $J(L)$ is at least $\eta(m)$.*

Proof. We think of the points P_i as elements of $J(L)$ via the Albanese map sending a point $P \in X(\bar{K})$ to the class of the divisor $(P) - (P_0)$. Suppose that the Mordell–Weil rank of $J(L)$ is n. Let Λ be the rank n free $\mathbb{Z}$-module $J(L)/J(L)^{\mathrm{tor}}$. The Galois group G of L/K acts on Λ, giving rise to an injective homomorphism $H \hookrightarrow \mathrm{Aut}(\mathbb{Z}^n) \cong \mathrm{GL}(n, \mathbb{Z})$, where H is G modulo the kernel of the action. Let h be the order of H. Since no two of the P_i are in the same torsion packet, the points $P_1, \ldots, P_m$ are distinct elements of Λ, and by assumption $P_1, \ldots, P_m$ form a complete orbit under H. Therefore m divides h. Since H is a subgroup of $\mathrm{GL}(n, \mathbb{Z})$, it follows from the definition of our function η that $\eta(m) \leq n$. □

Combining Proposition 4.10 and Theorem 4.6, for example, we obtain:

Corollary 4.11. *Let $N > 311$ be a prime number, let L be a Galois number field, and let $P \in X_0(N)(L)$ be a noncuspidal point having exactly m Galois conjugates. Then the Mordell–Weil rank of $J_0(N)(L)$ is at least $\eta(m)$.*

References

[1] M. Baker, Torsion points on modular curves, Ph.D. Dissertation, University of California, Berkeley, 1999

[2] J. Boxall, D. Grant, Examples of torsion points on genus two curves, to appear in Trans. Amer. Math. Soc.

[3] J.W.S. Cassels, Local Fields, London Mathematical Society Student Texts, vol. 3, Cambridge University Press, 1986

[4] R.F. Coleman, Torsion points on curves and p-adic Abelian integrals, Annals of Mathematics **121** (1985), 111–168

[5] R.F. Coleman, Ramified torsion points on curves, Duke Math. J. **54** (1987), 615–640

[6] R.F. Coleman, B. Kaskel, K. Ribet, Torsion points on $X_0(N)$, Automorphic forms, automorphic representations, and arithmetic (Fort Worth, TX, 1996), 27–49, Proc. Sympos. Pure Math. **66**, Part 1, Amer. Math. Soc., Providence, RI, 1999

[7] R.F. Coleman, A. Tamagawa, P. Tzermias, The cuspidal torsion packet on the Fermat curve, J. Reine Angew. Math. **496** (1998), 73–81

[8] B. Edixhoven, The weight in Serre's conjectures on modular forms, Invent. math. **109** (1992), 563–594

[9] N. Elkies, Elliptic and modular curves over finite fields and related computational issues, Computational Perspectives on Number Theory (Chicago, IL, 1995), 21–76, AMS/IP Stud. Adv. Math. **7**, Amer. Math. Soc., Providence, RI, 1998

[10] A. Grothendieck, SGA7 I, Exposé IX, Lecture Notes in Mathematics, vol. 288, Springer-Verlag, Berlin, New York, 1972, 313–523

[11] Y. Hasegawa, K. Hashimoto, Hyperelliptic modular curves $X_0^*(N)$ with square-free levels, Acta Arith. **77** (1996), 179–193

[12] Y. Hasegawa, M. Shimura, Trigonal modular curves. Acta Arith. **88** (1999), 129–140

[13] Y. Hasegawa, M. Shimura, Trigonal modular curves $X_0^{+d}(N)$. Preprint

[14] D. Kubert, S. Lang, Modular units, Grundlehren der mathematischen Wissenschaften 244, Springer-Verlag, Berlin, New York, 1981

[15] S. Lang, Division points on curves, Ann. Mat. Pura Appl. **70** (1965), 229-234

[16] H.W. Lenstra, Jr., K. Ribet. In preparation

[17] J. Manin, Parabolic points and zeta functions of modular curves, (Russian), Izv. Akad. Nauk SSSR Ser. Math. **36** (1972), 19–66

[18] B. Mazur, Modular curves and the Eisenstein ideal, Publ. Math. IHES **47** (1977), 33–186

[19] B. Mazur, Rational isogenies of prime degree, Invent. math. **44** (1978), 129–162
[20] K.V. Nguyen, M. Saito, D-Gonality of modular curves and bounding torsions, Preprint. Available on the Algebraic Geometry Web as alg-geom/9603024
[21] A.P. Ogg, Hyperelliptic modular curves, Bull. Soc. Math. France **102** (1974), 449–462
[22] A.P. Ogg, On the Weierstrass points of $X_0(N)$, Illinois J. Math. **22** (1978), no. 1, 31–35
[23] B. Poonen, Mordell-Lang plus Bogomolov. Invent. math. **137** (1999), 413–425
[24] M. Raynaud, Courbes sur une variété Abélienne et points de torsion, Invent. math. **71** (1983), 207–233
[25] M. Raynaud, Sous-variétés d'une variété Abélienne et points de torsion, in Arithmetic and Geometry, Vol. I, Progr. Math. 35, Birkhauser, Boston, 1983, 327–352
[26] K. Ribet, Torsion points on $J_0(N)$ and Galois representations, to appear in the Proceedings of the CIME conference on the Arithmetic of Elliptic Curves, Lecture Notes in Math., Springer-Verlag, Berlin, New York
[27] K. Ribet, Images of semistable Galois representations, Olga Taussky-Todd: in memoriam. Pacific J. Math. **1997**, Special Issue, 277–297
[28] J–P. Serre, Abelian l-adic representations and elliptic curves, Research Notes in Mathematics, 7, A.K. Peters, Ltd., Wellesley, MA, 1998
[29] J–P. Serre, Sur les représentations modulaires de degré 2 de $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Duke Math. J. **54** (1987), 179–230
[30] J–P. Serre, Cohomologie Galoisienne, Lecture Notes in Mathematics, vol. 5, Springer–Verlag, Berlin, New York, 1994
[31] J. Silverman, Hecke points on modular curves, Duke Math. J. **60** (1990), 401–423
[32] W. Stein, Modular forms database. Available via the Number Theory Web: http://www.math.uga.edu/~ntheory/web.html