Torsion Points on Modular Curves

by

Matthew Howard Baker

B.S. (University of Maryland at College Park) 1994

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mathematics

in the

GRADUATE DIVISION
of the
UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Robert F. Coleman, Chair
Professor Kenneth A. Ribet
Professor Terence P. Speed

1999

Abstract

Torsion Points on Modular Curves

by

Matthew Howard Baker
Doctor of Philosophy in Mathematics

University of California at Berkeley

Professor Robert F. Coleman, Chair

Let $N \geq 23$ be a prime number. In this thesis, we prove a conjecture of Coleman, Kaskel, and Ribet about the $\bar{\mathbb{Q}}$-valued points of the modular curve $X_0(N)$ which map to torsion points on $J_0(N)$ via the "usual" embedding. We also have results about the set of "torsion packets" on $X_0(N)$, and on torsion points on other modular curves. In addition, we study some related questions. For example, we determine all of the modular curves $X_0(N)$ which are trigonal, i.e., admit a degree three map to $\mathbf{P}^1$. We do this by generalizing a method Ogg used to determine the hyperelliptic curves $X_0(N)$. We also discuss more general questions about the gonality of modular curves, and discuss how these questions are related to the conjecture of Coleman, Kaskel, and Ribet. We also discuss Cartier points on curves, a concept motivated by Coleman's paper "Ramified Torsion Points on Curves". After proving some general results about Cartier points, we give some examples which yield information about torsion points on certain modular curves.

Professor Robert F. Coleman
Dissertation Committee Chair

To my parents, Michael and Fay Baker,

my brother, Darrin,

and my (soon-to-be) wife Kate.

Contents

1 Introduction **1**

2 Torsion Points on $X_0(N)$ **6**
2.1 The Coleman–Kaskel–Ribet Conjecture 6
2.2 Proof of the Coleman–Kaskel–Ribet Conjecture 10
2.3 Generalizations 18

3 Gonality of Modular Curves **23**
3.1 A Generalization of Ogg's Method 23
3.2 Trigonal $X_0(N)$ 30
3.3 Trigonal $X_0^+(N)$ 32

4 Cartier Points on Curves **38**
4.1 Introduction 38
4.2 Duality, Linear Systems, and Ekedahl's Theorem 41
4.3 Bounds 48
4.4 Examples 51

5 Additional Techniques **58**
5.1 Purpose 58
5.2 Useful Facts 58
5.3 The Curve $X_0(29)$ 59
5.4 The Curve $X_0(31)$ 62
5.5 A Computational Method 63
5.6 The Curves $X_0(43)$ and $X_0(61)$ 65
5.7 The Curve $X_0(53)$ 66

6 Appendix **70**

Bibliography **77**

Acknowledgements

There are many people who provided valuable assistance with this work, and I hope that their contributions are adequately recognized here.

First of all, I would like to thank Hendrik Lenstra, Arthur Ogus, Andrew Ogg, and Bjorn Poonen, with whom I have engaged in many interesting conversations during my graduate career. Time and again they were able to clearly communicate mathematical ideas to me. For example, A. Ogus showed me the proofs of Lemmas 3.5 and 4.23, and B. Poonen helped out greatly with the ideas in Section 5.5 and the appendix.

Joe Harris, Robin Hartshorne, and Felipe Voloch also made useful remarks about various parts of this thesis. In particular, J. Harris first told me about Proposition 4.9, which I needed to generalize Ekedahl's theorem, and F. Voloch made a nice improvement to the proof of Proposition 4.18. I also am indebted to Voloch for suggesting I look at the paper [27], which helped inspire the main results of Chapter 3.

There are many fellow graduate students I should thank, too many to name them all here. I would at least like to acknowledge here the many valuable ideas I learned in conversations with William Stein and János Csirik.

I would also like to thank Kevin Buzzard and Loïc Merel for sharing many mathematical ideas with me, and for the encouragement they gave me as a young, impressionable graduate student.

I would like to specially acknowledge at this point the crucial contributions of Ken Ribet to Chapter 2 of this work. The second proof I give of the CKR conjecture is based on emails from and discussions with him. Ribet also came up with 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 2.10, and explained how Lemmas 2.18 and 5.1 could be used to simplify some of my arguments. In general, I am grateful for the many useful remarks Ribet made concerning modular curves and Galois representations.

Last but not least, I wish to thank Robert Coleman for suggesting that I work on the Coleman-Kaskel-Ribet conjecture, and for his continuous help and encouragement. Many of the results in Chapter 5 also emerged from his suggestions.

Commutative diagrams in this thesis were designed using Paul Taylor's Commu-

tative Diagrams in TeX package. At various points during this research I made use of the software packages GAP, MAPLE, and PARI–GP. Typesetting for this thesis was done in LaTeX.

This research was supported by an NDSEG Fellowship and an Alfred P. Sloan Dissertation Fellowship.

Chapter 1

Introduction

The contents of this dissertation, while somewhat diverse, evolved from a common source: a conjecture of Coleman, Kaskel, and Ribet (see [12]) concerning the way in which the modular curve $X_0(N)$, with N prime, intersects the torsion subgroup of its Jacobian.

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 P_0 on X and define the map $i : X \hookrightarrow J$ by sending P to the divisor class $[(P) - (P_0)]$. (Recall that if L is an extension of K, the L-valued points of J correspond to linear equivalence classes of degree zero divisors on J defined over L.)

Now define the set T to be $\{P \in X(\bar{K}) \mid i(X) \in J^{\mathrm{tor}}\}$. In other words, T is the set of $\bar{K}$-valued points of X which map to torsion points on J.

If $g = 1$ (i.e., X is an elliptic curve), then i is an isomorphism, and so of course T is infinite. But if $g \geq 2$, the situation is entirely different. In fact, it is a theorem when $g \geq 2$ that T is a finite set of points. This theorem, proved by M. Raynaud in 1983, is known as the Manin–Mumford conjecture. Many different proofs of this conjecture followed, including a 1987 proof by R. Coleman based on p-adic integration.

There is a striking analogy between the Manin–Mumford conjecture, on the one hand, and the Mordell conjecture on the other. This analogy runs much deeper than the fact that both are theorems which are called conjectures. For example, in [36], Lang conjectured that $i(X) \cap \Gamma'$ is finite whenever 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 [48] 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 (which we will refer to as "Explicit Mordell") for a "random" curve X is an extremely hard problem, to put it mildly. 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. Bjorn Poonen has pointed out, however, that this problem is, at least in principle, effective. In this setting as well, the appropriate test cases seem to be curves which either have small genus (see [6] for some examples when $g = 2$) or unusually rich arithmetic structure. For an example of the latter, see [14], in which the authors determine T when X is a Fermat curve embedded in J using a "cusp".

In their joint paper [12], 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 $X_0(N)$.)

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. $X_0(N)$ is an algebraic curve 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" ([39]).

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 [45] that the converse is almost true as well: $X_0(N)$ is hyperelliptic if and only if $g^+ = 0$ or $N = 37$. $X_0(37)$ 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 $i_\infty(Q) \in J_0(N)$ 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 [12, 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 [12] 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 Chapter 2 of this thesis, we prove Conjecture 1.2 in full generality. In Section 2.1, we summarize the work previously done on this problem, in particular the results of [12] and [51]. We also discuss a few technical results needed later on. In Section 2.2, we give two proofs of the CKR conjecture. The first of these arguments relies on some of the material

found in later chapters of the present work. For example, it uses our classification of the trigonal modular curves $X_0(N)$ found in Chapter 3. It also uses some of the results in Chapters 4 and 5. The second proof, which came into being shortly after the first, is due mainly to Ribet and seems more ripe for generalization than the first proof. In particular, the trigonal modular curves no longer play an exceptional role in this approach.

We have recently learned that A. Tamagawa has independently proved Conjecture 1.2 by methods similar to Ribet's.

We give proofs in Section 2.3 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 (see Section 2.3 or the appendix) on $X_0(N)$ and $X_0^+(N)$ when N is sufficiently large.

More precisely, if X denotes either $X_0(N)$ or $X_0^+(N)$, we show that the torsion packets on X are just the "expected ones" unless X admits a (non-constant) morphism of small degree to the projective line $\mathbf{P}^1$. A theme which runs through Section 2.2 is that to understand torsion points on X, it is useful to understand the gonality of X, i.e., the smallest degree of a morphism from X to $\mathbf{P}^1$. So in Chapter 3, we provide an extensive discussion of the gonality of modular curves. For example, we establish an anologue of Ogg's classification of the hyperelliptic modular curves $X_0(N)$ by determining all N such that $X_0(N)$ is trigonal, i.e., admits a degree three map to $\mathbf{P}^1$. As we mentioned above, this is used in our first proof of the Coleman–Kaskel–Ribet conjecture.

We remark that recently other authors (Nguyen–Saito, Hasegawa–Shimura) have independently obtained results similar to those in Chapter 3 – see [43], [30], and [31].

In Chapter 4 we tackle a rather different topic, namely Cartier points on curves. Without defining at the moment what these are, we simply note that their definition was motivated by work of Coleman (see [10]) concerning ramified torsion points on curves. Some of our early work on the Coleman–Kaskel–Ribet conjecture for small N hinged on this connection between Cartier points and ramified torsion points, and we give some concrete examples of this connection in Section 4.4 and Chapter 5. We also prove a theorem about Cartier points which generalizes and gives a new proof of a theorem of Ekedahl on superspecial curves in characteristic p.

Finally, in Section 5 we present an assortment of techniques for dealing with torsion

points ramified at the special prime number 3. This material is largely based on suggestions of Robert Coleman. We hope that the techniques used in this section can also be used to study torsion points on other curves in other scenarios.

Chapter 2

Torsion Points on $X_0(N)$

2.1 The Coleman–Kaskel–Ribet Conjecture

We begin with a summary of [12], the paper in which the problem of determining torsion points on $X_0(N)$ was first posed.

The basic approach in [12] 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 [39, Theorem (1)] is the full group of rational torsion points on $J_0(N)$.

The following proposition follows from the main result of [40]; see [12, 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 [39, Proposition 9.5]). The main general result in [12] is the following theorem ([12, 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 [12] and [10]) plus the following theorem [12, Theorem 2.2] proved using techniques of Mazur (see [39]):

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

Ken Ribet's papers [51],[52] 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 [60]) 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 from [51] 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. 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 $J_0(N)$.*

Remark. In the proof of this theorem given in [51, Theorem 1.6], note that if $P_2 \in J_0(N)[\mathfrak{I}]$ (where $\mathfrak{I}$ is the kernel of the Eisenstein ideal), the hypothesis that $P_2 \notin C$ implies that the order 2^a of P_2 is in fact divisible by 4, since $C[2] = \Sigma[2]$. So the extension $\mathbb{Q}(\mu_{2^a})/\mathbb{Q}(\mu_{2^{a-1}})$ really is ramified at 2.

As we remarked in the introduction, it is useful to work with 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.

Recall that $X_0^+(N)$ is the quotient of $X_0(N)$ by the Atkin–Lehner involution $w = w_N$. There is a unique cusp at infinity on $X_0^+(N)$, which we denote by ∞^+, or simply ∞ if no confusion is likely to arise. 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

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

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 [39, II.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 [39]). 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 [39, II, Proposition 17.10].

Along similar lines, we have the following proposition.

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

PROOF. Let V be such a Jordan-Holder 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 [39, 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 [39, 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, one simply notes that Ribet's proof of Theorem 2.5 follows *mutis mutandis* for $J_0^+(N)$. In fact, Theorem 2.7 is actually easier to prove than Theorem 2.5, because the complications arising from Eisenstein primes in the proof of the latter result do not occur here. ■

2.2 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 a few facts which we will need.

Theorem 2.8. *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 [45], and part (2) from the main result of [29]. Assertion (3) follows from Theorem 3.12, and (4) follows from Theorem 3.14. The rest of the theorem is proved in Example 3.11 of Chapter 3. ■

The following theorem is proved in Chapter 5 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 2.9. *The CKR conjecture is true for the trigonal modular curves $X_0(N)$, i.e., the curves $X_0(N)$ with $N \in \{23, 29, 31, 37, 43, 53, 61\}$.*

The proof of the following lifting lemma is reminiscent of the techniques of [53, IV-23 Lemma 3].

Lemma 2.10. *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. (Ribet) 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 [37], it follows from the much more elementary Lemma 2.10.

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 a p-Sylow subgroup of an inertia group at p of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$; see [57, Section 1.4] for a discussion of p-Sylow subgroups of profinite groups. Also, we denote the genus of $X_0(N)$ by $g_0(N)$.

Proposition 2.11. *Let $p \geq 5$ and N be primes with $g_0(N) > 0$, and suppose that p does not divide $N - 1$. Let $\mathbf{T} = \mathbf{T}_0(N)$ be the Hecke algebra associated to $J_0(N)$, and let $\rho = \prod \rho_i : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T} \otimes \mathbb{Z}_p) = \prod \mathrm{GL}(2, \mathbf{T}_{\mathfrak{m}_i})$ be the semistable representation coming from the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the p-adic Tate module of $J_0(N)$. If $p \neq N$, let X be the normal closure of an inertia group at p, and if $p = N$ let X be the normal closure of a wild inertia group at N. Then the image $\rho(X) \subseteq \mathrm{GL}(2, \mathbf{T} \otimes \mathbb{Z}_p)$ 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 $\rho(X)$ 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 2.10, 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 $\mathrm{SL}(2, A/R)$. This follows from [52, Theorem 3.4, Proposition 6.3] when $p \neq N$, and from [51, Proposition 6.4] when $p = N$. ■

Similarly, we have the following:

Proposition 2.12. *Let $p \geq 5$ and N be primes such that $g_0^+(N) > 0$. (We allow the case where p divides $N-1$). Let $\mathbf{T}^+ = \mathbf{T}_0^+(N)$ be the Hecke algebra associated to $J_0^+(N)$, and let $\rho = \prod \rho_i : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(2, \mathbf{T}^+ \otimes \mathbb{Z}_p) = \prod \mathrm{GL}(2, \mathbf{T}^+_{\mathfrak{m}_i})$ be the semistable representation coming from the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the p-adic Tate module of $J_0^+(N)$. If $p \neq N$, let X be the normal closure of an inertia group at p, and if $p = N$ let X be the normal closure of a wild inertia group at N. Then the image $\rho(X) \subseteq \mathrm{GL}(2, \mathbf{T}^+ \otimes \mathbb{Z}_p)$ contains the homothety -1.*

Proof. The proof is essentially the same as the proof of the previous Proposition. It suffices to note that in this case all ρ_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 $\rho(\sigma)$ for $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ (since the analogous statement is true for $\mathbf{T}$), thus allowing us to invoke the results of [52]. ■

Corollary 2.13. *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 2.12, 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 $\rho_l(I)$ 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 2.14. *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 2.15. *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 $Q \in X_0(N)(\bar{\mathbb{Q}})$ 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 2.8(3), this implies that $N \in \{23, 29, 31, 37, 43, 53, 61\}$. But Theorem 2.9 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 2.13, 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 2.14, 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 2.22 below).

So we can assume that $g^+ \leq 1$. One can then explicitly check using Theorem 2.2 that the CKR conjecture is true whenever $g^+ \leq 1$ (the largest N for which $X_0^+(N)$ has genus 0 or 1 is $N = 131$). Indeed, it suffices to show for these values of N that 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. That this is the case follows from the tables in [60].

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 2.8(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 2.13, 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 2.8(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 2.14, $X_0^+(N)$ is sub-hyperelliptic. But we have already dispensed of this case, so our first proof of the Coleman–Kaskel–Ribet conjecture is complete. ■

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

We recall that $J_0(N)[\mathfrak{I}]$, the kernel of the Eisenstein ideal, has order n^2, where $n = \mathrm{Num}\frac{N-1}{12}$. 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 $C + \Sigma$. When n is even, $C + \Sigma$ 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 2.16. *$J_0(N)[\mathfrak{I}]$ 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 [51, 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, for the same reason we mention in the proof of Theorem 2.7. ■

Lemma 2.17. *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 alternate argument. Let $P = i_\infty(Q) \in J_0(N)^{tor}$. 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 [47]). Otherwise Q is not rational, so there is some $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ 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 $i_\infty(Q), i_\infty(Q')$ 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 contradicts the fact that the order of P is even. ■

Lemma 2.18. *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 2.19. *Let $N \geq 23$ be prime. The only points $Q \in X_0(N)(\bar{\mathbb{Q}})$ 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 $n = \mathrm{Num}\frac{N-1}{12}$) 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 [60], 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 [10]), 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 $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ 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 2.17.

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 [10].

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 2.18 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 2.17. ■

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

PROOF. Suppose $Q \in X_0(N)(\bar{\mathbb{Q}})$ 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 2.19. So we can assume that $P \notin J_0(N)[\mathfrak{I}]$, which by Theorem 2.16 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. Hence $X_0(N)$ is hyperelliptic and σQ (and hence Q, since the hyperelliptic involution

is defined over $\mathbb{Q}$) 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 semistable reduction theorem (see [24], and also [51, (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 $(\sigma - 1)^2 P = 0$. We can find such a σ in X, the normal closure of a wild inertia group at N in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Elements of X fix prime-to-N torsion, and moreover it follows from [37] (or from [52, Proposition 6.4] when N satisfies hypothesis (*)) that the image of X in $\mathrm{Aut}(Ta_N(J_0(N))) \cong \mathrm{GL}(2, \mathbf{T} \otimes \mathbb{Z}_N)$ contains $\mathrm{SL}(2, \mathbf{T} \otimes \mathbb{Z}_N)$. In particular, there exist $\sigma_1, \sigma_2 \in X$ 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 also satisfies $(\sigma - 1)^2 P_N = 0$, so we're done. ■

We remark that we can modify this proof so that only [52, Proposition 6.4], and not [37], is needed. According to Corollary 2.13, there exists $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which acts trivially on prime-to-N torsion points of $J_0(N)$ and as -1 on N-power torsion points. Therefore $P' = (1 + \tau)P$ has order prime to N, and we find that there exists σ such that $\sigma P' \neq P'$ but $(\sigma - 1)^2 P' = 0$ using the easier order prime-to-N case of Theorem 2.5. It is not hard to see that this implies that $X_0(N)$ admits a morphism to $\mathbf{P}^1$ of degree at most 4. If such a morphism exists then $N \leq 191$ by Theorem 2.8(5). But we know from Proposition 2.4 that the smallest (and only known) prime for which (*) is not satisfied is 389. In particular, for prime numbers N at most 191 this condition *is* satisfied. So we are done by the argument above, which assumes (*).

2.3 Generalizations

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

Let X/k be a curve. Following [10], we define an equivalence relation on $X(\bar{k})$ by saying that P is equivalent to Q if some multiple of the divisor $(P)-(Q)$ is principal. The set of all points in a single equivalence class is called a *torsion packet* on X.

For example, we showed in the previous section that when $g_0^+(N) > 0$, the torsion packet on $X_0(N)$ containing ∞ is precisely the set of cusps. More generally we have the following.

Proposition 2.20. *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 [38] and Kubert–Lang [34] 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 X $^{i_\infty}$ J

$X_0(N)^{i_\infty}$ $J_0(N)$

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 2.20 with a theorem of Mazur proved in [59, Theorem 0.4].

Corollary 2.21. *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 2.22. *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 2.8(2), $X_0^+(N)$ is hyperelliptic exactly when it has genus 2. Hasegawa [32] found that 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. (Specifically, the image of ∞ is a Heegner point of class number one).

In other words, we have $h\infty \neq \infty$. This can also be seen by looking at q-expansions of weight-two cusp forms for $\Gamma_0^+(N)$, since $h\infty = \infty$ iff ∞ is a Weierstrass point on $X_0^+(N)$ iff there is a form $f = a_1q + a_2q^2 + a_3q^3 + \ldots$ 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 [60]. (For somewhat larger prime values of N, however, it seems that ∞ usually *is* a Weierstrass point on $X_0^+(N)$. See [19] for a discussion of this.)

In any case, it is a theorem of Mazur [39, 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 2.23. *When $g_0^+(N) \geq 2$, ∞ is the only point $Q \in X_0^+(N)(\bar{\mathbb{Q}})$ such that $i_\infty(Q) \in J_0^+(N)^{tor}$. 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 2.16 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 $J_0(N)$, it follows from our second proof of the CKR conjecture that there exists a σ in an inertia group for 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 2.22. ■

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

Theorem 2.24. *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 2.8(6), the first assertion implies the second. So we assume that $P = [(Q_1) - (Q_2)] \in J_0^+(N)^{tor}$ 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 2.25. *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 2.8(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 2.24 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 2.19, 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 2.18 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 and Q_1 is a hyperelliptic branch point, or $Q_1 = \sigma Q_1$ for all σ, i.e., Q_1 is defined over the rational numbers. By the main result of [40], 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 2.26. *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 $Q \in X_0(N)(\bar{\mathbb{Q}})$. 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 [47] 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 2.25. ■

We also mention the following result, whose proof is nearly the same as the proofs of Theorems 2.24 and 2.25.

Theorem 2.27. *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 $X_0(N)^{(d)}(\bar{\mathbb{Q}})$ (resp. $X_0^+(N)^{(d)}(\bar{\mathbb{Q}})$) 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 $\mathbf{P}^1$.*

Chapter 3

Gonality of Modular Curves

3.1 A Generalization of Ogg's Method

A curve X/k is *hyperelliptic over* k if it has genus $g \geq 2$ and admits a degree two map to $\mathbf{P}^1$ defined over k. It is hyperelliptic if it is hyperelliptic over $\bar{k}$. Similarly, we will say that X/k is *trigonal* if it is trigonal over $\bar{k}$, which means that it has genus at least two and admits a degree three map to $\mathbf{P}^1$ defined over $\bar{k}$.

In general, we say that the *k-gonality* of X is the degree of the smallest morphism from X to $\mathbf{P}^1$ defined over k. The gonality of X (with no modifier) will mean its $\bar{k}$-gonality. With our terminology, a curve X of genus $g \geq 2$ is hyperelliptic iff its gonality is 2. It is well-known (and is easily deduced from Lemma 3.1) that X is trigonal iff $g = 2$ or the gonality of X is 3.

In this section, unlike the previous ones, N is allowed to be a composite integer unless otherwise specified.

In [45], A. Ogg proved that the modular curve $X_0(N)$ (for N a positive integer) is hyperelliptic exactly when $N = 22, 23, 26, 28, 29, 30, 31, 33, 35, 37, 39, 41, 46, 47, 50, 59, 71$.

His proof works by reducing modulo p and using the fact that if $X_0(N)$ is hyperelliptic, then it is hyperelliptic over $\mathbb{Q}$ and the reduction mod p is hyperelliptic over $\mathbf{F}_p$ for $p \nmid N$. But then the number of points in $X_0(N)(\mathbf{F}_q)$ for q a power of p is bounded by twice the number of points of $P^1(\mathbf{F}_q)$, or $2(q+1)$. However, when N is large $X_0(N)(\mathbf{F}_{p^2})$ has many points, coming from supersingular elliptic curves in characteristic p. So one gets a bound on the values of N such that $X_0(N)$ can be hyperelliptic, and studying Atkin–Lehner

involutions and Weierstrass points gives the exact list mentioned above.

The main obstacle to a direct generalization of Ogg's method to degree d maps is that whereas a curve defined over k which is hyperelliptic and possesses a k-rational point is automatically hyperelliptic over k, the analogous fact need not be true for d-gonal curves if $d > 2$. But under certain hypotheses it *will* be true when the genus is sufficiently large with respect to d.

In [1], Abramovich proves (using methods of Yau) that if $X_0(N)$ admits a morphism of degree d to $\mathbf{P}^1$ over the complex numbers, then $N \leq \frac{800d}{7}$. For most d this lower bound for the gonality of $X_0(N)$ is much better than we can obtain with our techniques, but for small d (such as $d = 3$), our methods give better bounds. We also note that our methods give information about the gonality of $X_0(N)_{\mathbf{F}_p}$, the reduction modulo p of $X_0(N)$.

In subsequent sections we give some concrete applications of our techniques. For example, we determine all trigonal modular curves $X_0(N)$, and we also determine when $X_0^+(N)$ is trigonal in the case where N is a prime number.

We will have use for the following result, which we refer to as the Correspondence Lemma, throughout this section. If $f_1 : C \to C_1$ and $f_2 : C \to C_2$ are nonconstant morphisms between curves, we say that f_1 and f_2 are *independent* if C is birational to its image C' under the induced map to $C_1 \times C_2$.

C

$(4,2)^{f_2}(2,4)_{f_1}$

$C'C_2$

C_1

Lemma 3.1 (Correspondence Lemma). *Let C_1, C_2, C be curves of genera g_1, g_2, and g, respectively, over the field k. Let $f_i : C \to C_i$ be morphisms of degree d_i, $i = 1, 2$. Assume that f_1 and f_2 are independent. Then*

$$g \leq (d_1 - 1)(d_2 - 1) + d_1 g_1 + d_2 g_2.$$

Remark. The hypothesis of the lemma is satisfied, for example, if $\gcd(d_1, d_2) = 1$.

PROOF. This is a classical result of Castelnuovo and Severi, and can be found in many places. For example, it appears as Exercise VIII.C-1 in [4]. ■

We also give an easy lemma which allows us to recognize when maps are independent.

Lemma 3.2. *The morphisms $f_1 : C \to C_1$ and $f_2 : C \to C_2$ fail to be independent if and only if there exists a complete nonsingular curve D and morphisms $h : C \to D$, $h_1 : D \to C_1$, and $h_2 : D \to C_2$ such that $f_1 = h_1 \circ h$, $f_2 = h_2 \circ h$, and h has degree greater than 1.*

PROOF. Suppose the map $f : C \to C' \subset C_1 \times C_2$ has degree $d > 1$. Then the induced map $h : C \to D$ to the normalization D of C' also has degree $d > 1$, and the maps f_1 and f_2 factor as required. Conversely, suppose there exists D as in the statement of the lemma. Then the map from C to its image C' factors through the map from C to D, so does not have degree 1. ■

Using the Correspondence Lemma, we obtain a criterion for the existence of morphisms of a given degree defined over a given base field.

Proposition 3.3. *Suppose $X, Y/k$ are curves of genera g, g' over the perfect field k. Suppose $X(k)$ is nonempty. Let f be a degree d morphism from X to Y defined over $\bar{k}$, and suppose that f does not factor through any intermediate curve. If $g > (d-1)^2 + 2dg'$, then there exists a degree d morphism defined over k from X to a curve Y' isomorphic over $\bar{k}$ to Y.*

PROOF. (Compare with [27]) For each $\sigma \in G_k := \mathrm{Gal}(\bar{k}/k)$, we get a degree d morphism $f^\sigma : X \to Y$. Since f does not factor through any intermediate curves, the degree of the map from X to its image under $f \times f^\sigma$ is either 1 or d. But the Correspondence Lemma shows that it cannot be 1, since $g > (d-1)^2 + 2dg'$. So f and f^σ must differ by an automorphism of Y; in other words, for each $\sigma \in G_k$, there exists $\alpha_\sigma \in \mathrm{Aut}(Y)$ such that $f^\sigma = \alpha_\sigma \circ f$. The collection $\{\alpha_\sigma\}$ forms a 1-cocycle in $H^1(G_k, \mathrm{Aut}(Y))$, and by [58, X.2.2] corresponds to a twist of Y. In other words, there exists a curve Y'/k and an isomorphism $\lambda : Y' \to Y$ defined over $\bar{k}$ such that $\alpha_\sigma = \lambda_\sigma \circ \lambda^{-1}$ for all $\sigma \in G_k$. The morphism we seek from X to Y' is then $\lambda^{-1} \circ f$. ■

Here is the simple case of this proposition which we will actually need.

Corollary 3.4. *Let X be as in the proposition. Let l be a prime number, and suppose there exists a degree l morphism from X to $\mathbf{P}^1$ defined over $\bar{k}$. If $g > (l-1)^2$ then there exists a degree l morphism from X to $\mathbf{P}^1$ defined over k.*

Before we applying this result, we have a lemma which is closely related to the semicontinuity theorem [28, III, 12.8]:

Lemma 3.5. *Let R be a discrete valuation ring with field of fractions K, residue field k, and uniformizing parameter π. Let X be a one-dimensional regular scheme proper and flat over* $\mathrm{Spec}(R)$*, with generic fiber X_K (a nonsingular curve over K) and special fiber X_k (a possibly singular curve over k). Then any line bundle $\mathcal{L}_K$ of degree d on X_K extends to a line bundle $\mathcal{L}$ on X, and the pullback $\mathcal{L}_k$ to X_k is a degree d line bundle with* $\dim_K H^0(X_K, \mathcal{L}) \leq dim_k H^0(X_k, \mathcal{L}_k)$.

Proof. With our hypotheses, it follows from [28, II, 6.11] that there is a natural isomorphism between the divisor class group of X and the group $\mathrm{Pic}(X)$. The surjectivity of the natural map $\mathrm{Pic}(X) \to \mathrm{Pic}(X_K)$ then follows from [28, II, 6.5(a)]. (If X/R is smooth then this map is also injective). Let $\mathcal{L}$ be a line bundle on X extending the degree d line bundle $\mathcal{L}_K$ on X_K. If we define the degree of a line bundle $\mathcal{L}$ to be $\chi(\mathcal{L}) - \chi(\mathcal{O}_X)$, then this degree is constant on fibers because X/R is flat; by Riemann–Roch this degree must then be d.

We have $\dim_K H^0(X_K, \mathcal{L}_K) = \dim_K H^0(X, \mathcal{L}) \otimes_R K$ because cohomology commutes with the flat base extension $R \to K$ by [28, III, 9.3].

The pullback (restriction) $\mathcal{L}_k$ of $\mathcal{L}$ is a line bundle of degree d on X_k, and it remains to show that $\dim_K H^0(X, \mathcal{L}) \otimes_R K \leq \dim_k H^0(X_k, \mathcal{L}_k)$. This is a special case of the semicontinuity theorem, but for the reader's convenience we give a self-contained and more elementary proof.

First of all, it is not hard to see that the "multiplication by π" map $j : \mathcal{L} \to \mathcal{L}$ induces a short exact sequence of sheaves on X

$$0 \to \mathcal{L} \to \mathcal{L} \to i_*\mathcal{L}_k \to 0$$

where $i : X_k \hookrightarrow X$ is the natural inclusion. For example, the map j is injective because $\mathcal{L}$ is locally free over $\mathcal{O}_X$ and hence R-torsion free. We then get a long exact sequence of cohomology

$$0 \to H^0(X, \mathcal{L}) \otimes_R K \to H^0(X_k, \mathcal{L}_k) \to H^1(X, \mathcal{L})_\pi \to 0.$$

In particular, the injection on the left yields the desired inequality. ∎

Proposition 3.6. *Let N be a positive integer, let $\nu(N)$ be the number of distinct primes dividing N, let $\mu_N = \prod_{p|N}(1+\frac{1}{p})$, and let $g_0(N)$ be the genus of $X_0(N)$. Let l, p be prime numbers with p not dividing N. If $g_0(N) > (l-1)^2$ and $2^{\nu(N)} + \frac{p-1}{12} \cdot \mu_N > l(p^2+1)$, then there does not exist (over $\bar{\mathbb{Q}}$) a degree l morphism from $X_0(N)$ to $\mathbf{P}^1$.*

PROOF. Suppose, to the contrary, that there exists a degree l morphism $X_0(N) \to \mathbf{P}^1$. The hypothesis $g_0(N) > (l-1)^2$ means that there exists such a morphism defined over $\mathbb{Q}$. By Lemma 3.5, there exists a degree l morphism to $\mathbf{P}^1$ from the reduction of $X_0(N)$ mod p defined over $\mathbf{F}_p$. In particular the number of $\mathbf{F}_{p^2}$-valued points of $X_0(N)_{\mathbf{F}_p}$ is at most $l(p^2+1)$. On the other hand, Ogg's proof in [45], combined with the mass formula from [58, Ex. 5.9], shows that $\#X_0(N)(\mathbf{F}_{p^2}) \geq 2^{\nu(N)} + \frac{p-1}{12} \cdot \mu(N)$ (see [62, Remark 4.1.55]). This is a contradiction. ■

We note that the Correspondence Lemma also gives the following result about the gonality of $X_0(N)$. For N prime, define the function $h_0(N)$ to be $\epsilon(N) \cdot h(N)$, where $h(N)$ is the class number of $\mathbb{Q}(\sqrt{-N})$ and $\epsilon(N)$ is $\frac{1}{2}$ if $N \equiv 1 \pmod 4$, 1 if $N \equiv 7 \pmod 8$, and 2 if $N \equiv 3 \pmod 8$. By [46, p. 20], $h_0(N)$ is equal to half the number of fixed points of the Atkin–Lehner involution acting on $X_0(N)$.

Proposition 3.7. *Let d be a positive integer and let N be a prime such that $h_0(N) > d$. If $X_0(N)$ admits a morphism f of degree d to $\mathbf{P}^1$, then f factors through the quotient $X_0^+(N)$ of $X_0(N)$ by its Atkin–Lehner involution. In particular, if d is odd then $X_0(N)$ does not admit a morphism to $\mathbf{P}^1$ of degree d.*

PROOF. Let g, g^+ be the genera of $X_0(N), X_0^+(N)$, respectively. Suppose $X_0(N)$ admits a morphism f of degree d to $\mathbf{P}^1$. If this map does not factor through the map π from $X_0(N) \to X_0^+(N)$, then the maps f and π are independent. The Correspondence Lemmma then tells us that $g \leq d - 1 + 2g^+$, i.e., $g - 2g^+ + 1 \leq d$. But $g - 2g^+ + 1 = h_0(N)$ by the Riemann-Hurwitz formula since $h_0(N)$ is precisely half the number of ramification points of π. ■

Remark. Since $h_0(N)$ grows roughly like $\sqrt{N}$, one can view this proposition as giving bounds similar to those of Proposition 3.6. In practice, a combination of these two propositions yields more information than either by itself.

Next, we show how to apply these methods in the case of certain degree d morphisms to $\mathbf{P}^1$ with d composite. For similicity, we restrict to the case where d is a product

of at most two primes, since this is the only case we need in Chapter 2.

First we have a couple of lemmas. The first is well-known.

Lemma 3.8. *If there is a map of degree at most d from X to $\mathbf{P}^1$ and X covers Y, then there is a map of degree at most d from Y to $\mathbf{P}^1$ as well.*

PROOF. One can see this in terms of function fields by taking norms – see [27] for a proof. ■

Lemma 3.9. *Let f be a morphism of degree $d = d_1 d_2$ (d_1, d_2 prime) from a curve X/k with $X(k)$ nonempty to $\mathbf{P}^1$. Then f satisfies one of the following two conditions:*

(I) $g(X) \leq (d-1)^2$

(II) There exists a curve Y of genus at most $\max((d_1-1)^2, (d_2-1)^2)$ defined over k and a k-rational morphism $h : X \to Y$.

PROOF. Suppose condition (I) does not hold. Then by the Correspondence Lemma, f and f^σ are dependent for all $\sigma \in \mathrm{Gal}(\overline{k}/k)$. One possibility is that for each $\sigma \in \mathrm{Gal}(\overline{k}/k)$ there exists $\beta_\sigma \in \mathrm{Aut}(\mathbf{P}^1)$ such that $f^\sigma = \beta_\sigma \circ f$. In this case, the proof of Lemma 3.3 shows that there exists a morphism f' of degree d defined over k from X to $\mathbf{P}^1$. In particular, condition (II) is satisfied. By Lemma 3.2 and the Correspondence Lemma, the other possibility is that there exist $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Y, $f_1 : X \to Y$ and $f_2 : Y \to \mathbf{P}^1$ so that $f = f_2 \circ f_1$, where (switching d_1 and d_2 if necessary) f_1 has degree d_1 and f_2 has degree d_2, and such that f and f^σ both factor through f_1 but do not differ by an automorphism of $\mathbf{P}^1$. Since d_2 is prime, f_2 and f_2^σ are independent, and therefore $g(Y) \leq (d_2-1)^2$. Furthermore, since (I) does not hold, we also have

$$(d_1-1)^2 + 2d_1 g(Y) \leq (d_1-1)^2 + 2d_1(d_2-1)^2 \leq (d_1-1)^2 + d_1^2(d_2-1)^2$$

$$< (d_1 - 1 + d_1(d_2-1))^2 = (d-1)^2 < g(X).$$

Combining this fact with the Correspondence Lemma (and using the assumption that d_1 is prime), we find that for all $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, there is an isomorphism $\alpha_\tau : Y \xrightarrow{\sim} Y^\tau$ such that $f_1^\tau = \alpha_\tau \circ f_1$. It follows as in the proof of Lemma 3.3 that f, Y can be chosen so they are defined over k. ■

Proposition 3.10. *Fix a positive integer $d = d_1 d_2$ with d_1 and d_2 prime. Let N be an odd number and let $X/\mathbb{Q}$ be either $X_0(N)$ or $X_0^+(N)$. Let $\alpha_4(N)$ denote the number of $\mathbf{F}_4$-rational points of the reduction of X mod 2, and let g be the genus of X. Assume that $g > (d-1)^2$ and*

$$\alpha_4(N) > \max(5d, d_1(5 + 4(d_2 - 1)^2), d_2(5 + 4(d_1 - 1)^2)).$$

Then there is no degree d morphism from X to $\mathbf{P}^1$.

PROOF. Suppose there is a degree d morphism from X to $\mathbf{P}^1$. By Lemma 3.5, there is such a morphism in characteristic 2 from $X_{\mathbf{F}_2} \to \mathbf{P}^1_{\mathbf{F}_2}$. The result now follows from Lemma 3.9 using the Weil bound $\#Y(\mathbf{F}_4) \leq 5 + 4g(Y)$. ■

Remark. In the proof of Proposition 3.6, we saw that if $X = X_0(N)$ then $\alpha_4(N) \geq 2^{\nu(N)} + \frac{N+1}{12}$. Similarly, if $X = X_0^+(N)$ then $\alpha_4(N) \geq 2^{\nu(N)-1} + \frac{N+1}{24}$. This follows from the fact that there is a degree 2 map from $X_0(N)_{\mathbf{F}_2} \to X_0^+(N)_{\mathbf{F}_2}$ defined over $\mathbf{F}_2$.

Example 3.11. *Maps of degree 4,5, or 6 from X to $\mathbf{P}^1$.*

For simplicity assume that $X = X_0(N)$ or $X_0^+(N)$ over $\mathbb{Q}$ with N prime. Suppose there is a degree 4 map f from X to $\mathbf{P}^1$ defined over $\bar{\mathbb{Q}}$. Then either $g(X) \leq (4-1)^2 = 9$ or $\alpha_4(N) \leq 4 \cdot 5 = 20$. Taken together, these conditions mean that $N < 215$ if $X_0(N)$ admits a degree 4 map to $\mathbf{P}^1$, or $N \leq 479$ if $X_0^+(N)$ does.

If we take $d = 5$, which is prime, we quickly find that either $g(X) \leq (5-1)^2 = 16$ or $\alpha_4(N) \leq 25$. This implies that $N < 275$ if $X_0(N)$ admits a degree 5 map to $\mathbf{P}^1$, and $N \leq 719$ if $X_0^+(N)$ does.

Finally, when $d = 6$ we have either $g(X) \leq (6-1)^2 = 25$ or $\alpha_4(N) \leq 42$. This implies that $N \leq 479$ if $X_0(N)$ admits a degree 6 map to $\mathbf{P}^1$, and $N \leq 911$ if $X_0^+(N)$ does.

We remark that it is often possible to improve these bounds. For example, for any given value of N we can compute $\alpha_4(N)$ and $\alpha_9(N)$ (the number of points in $X_0(N)(\mathbf{F}_9)$) exactly — see Section 3.2 for a method for doing this. This gives additional constraints. For the gonality of $X_0(N)$, we get additional constraints by utilizing 3.7 along with the results of Section 3.3. Implementing all of these improvements in PARI, we get the following improved bounds:

If $X_0(N)$ admits a degree 4 map to $\mathbf{P}^1$ then $N \leq 191$.

If $X_0(N)$ admits a degree 5 map to $\mathbf{P}^1$ then $N \leq 197$.

If $X_0(N)$ admits a degree 6 map to $\mathbf{P}^1$ then $N \leq 311$.

The first and last of these bounds are *sharp*. Since $X_0^+(191)$ has genus 2, it is hyperelliptic, so $X_0(N)$ admits a degree 4 map to $\mathbf{P}^1$. Similarly, $X_0^+(311)$ has genus 4 and therefore (as we note in the proof of Theorem 3.12) is trigonal.

3.2 Trigonal $X_0(N)$

We can now generalize Ogg's results to determine for which N the modular curve $X_0(N)$ is trigonal.

In the statement of the following theorem, recall that $g_0(N)$ denotes the genus of $X_0(N)$.

Theorem 3.12. *$X_0(N)$ is trigonal if and only if $g_0(N) = 2$ or $g_0(N)$ is 3 or 4 and $X_0(N)$ is not hyperelliptic.*

Remark. When N is prime, this says that $X_0(N)$ is trigonal if and only if

$$N \in \{23, 29, 31, 37, 43, 53, 61\}.$$

Proof. One direction is immediate: curves of genus 2 are trigonal by Riemann–Roch, and nonhyperelliptic curves of genus 3 or 4 are trigonal by [28, IV, 5.5.2]. Also, notice that a hyperelliptic curve of genus greater than 2 cannot be trigonal. This follows, for example, from the Correspondence Lemma.

For the converse, we first treat the case where N is *prime*, since it is simpler than the general case but illustrates all the main ideas. Also, our application to the Coleman–Kaskel–Ribet conjecture concerns prime values of N.

It follows from Proposition 3.6 (taking $l = 3$ and $p = 2$) that for N prime, if $g_0(N) > 4$ and $X_0(N)$ is trigonal then $N \leq 151$. (The bound from [1] would only give $N < 343$.) In fact, we get something stronger, because for any specific N and $p \nmid N$ we can find the exact value of $\#X_0(N)(\mathbf{F}_{p^2})$. A quick way to do this is to first calculate the characteristic polynomial F of the Hecke operator T_p acting on $S_2(\Gamma_0(N))$ using modular symbols, and then to compute the numerator f of the zeta function of $X_0(N)$ over $\mathbf{F}_p$ using the identity $F(x + p/x) = x^{-g} f(x)$ proved in [46, p. 23].

Proposition 3.6 implies that if $X_0(N)$ is trigonal and $g_0(N) > 4$ then we have $\#X_0(N)(\mathbf{F}_4) \leq 15$ and $\#X_0(N)(\mathbf{F}_9) \leq 30$. Let T be the set of prime N such that $g_0(N) > 4$ and $X_0(N)$ is trigonal. Computing the quantities $\#X_0(N)(\mathbf{F}_4)$ and $\#X_0(N)(\mathbf{F}_9)$. exactly for $N \leq 151$, we find that T is contained in the set $\{67, 73, 79, 83, 89, 101, 103, 131\}$.

Now note that a curve which is both bielliptic (admits a degree 2 map to an elliptic curve) and trigonal has genus at most 4. This again follows from the Correspondence Lemma. Using the tables of [17] to identify when $X_0(N)$ is bielliptic (which only happens, it appears, when $g^+ = 1$), we see that $T \subseteq \{67, 73, 103\}$. Furthermore, $X_0(103)$ cannot be trigonal because it violates the condition $h_0(N) \leq 3$ given by Proposition 3.7. So we are left with the possibilities $N = 67, 73$, both of which have $g_0(N) = 5$. We use the criterion of [28, IV, 5.5.3], which says that a curve of genus 5 is non-trigonal if and only if it is a complete intersection (of 3 quadric hypersurfaces) in its canonical embedding. Using the tables found in [20], we see that $X_0(67), X_0(73)$ are both complete intersections in their canonical embeddings. So $T = \emptyset$ and we are done with the case where N is prime.

Now for the general case. In addition to the tools we have already used, we need two more. The first is the fact that by Lemma 3.8, if $X_0(MN)$ is trigonal and $g_0(N) \geq 2$ then $X_0(N)$ is either trigonal or hyperelliptic.

The second thing we need is a generalization of the criteria for trigonality given above for curves of genus 5. This is afforded by the following theorem (see [4, III.3]):

Theorem 3.13 (Petri's Theorem). *Let C be a curve of genus $g \geq 4$ canonically embedded in $\mathbf{P}^{g-1}$ by the homogeneous ideal I. C is neither trigonal nor a plane quintic if and only if I is generated by homogeneous quadratic relations. If C is trigonal or a plane quintic, then I is generated by homogeneous polynomials of degree 2 and 3.*

The reason this is particularly convenient is that using a computer one can work out generators for the canonical ideal I of $X_0(N)$. In particular (supposing that $X_0(N)$ is not hyperelliptic), we want to establish whether or not the vector space of cubic relations induced by the $\frac{(g-2)(g-3)}{2}$ quadratic ones spans the $(\frac{g(g+7)(g-4)}{6} + 5)$-dimensional vector space of all cubic relations. The dimension counts just given follow from Max Noether's theorem ([4, III.2]) that the natural maps $\mathrm{Sym}^n H^0(C, K) \to H^0(C, K^n)$ are surjective for $n \geq 1$, where K^n is the sheaf of n-fold differentials. By "induced cubic relations" we mean the image of $H^0(C, K) \otimes I_2$ inside the kernel of the map $\mathrm{Sym}^3 H^0(C, K) \to H^0(C, K^3)$,

where I_2 is the vector space of quadratic relations thought of as the kernel of the map $\mathrm{Sym}^2 H^0(C,K) \to H^0(C,K^2)$. In the case of the modular curves $X_0(N)$, we may think of elements of $\mathrm{Sym}^n H^0(C,K)$ as cusp forms of weight $k=2n$ for $\Gamma_0(N)$, and by [61] a cusp form of weight k is determined by the first $\mu_N k/12$ coefficients of its q-expansion, where $\mu_N = \prod_{p|N}(1+\frac{1}{p})$. So the question of determining whether or not a given curve is trigonal reduces to linear algebra over $\mathbb{Q}$, provided we can compute enough terms in the q-expansions for a basis of $S_2(\Gamma_0(N),\mathbb{Q})$. This can be done using, for example, modular symbols — see the tables in [60].

In theory, this gives a recipe for determining trigonality for any given value of N. In practice the linear algebra outlined above can become computationally difficult, so it is better to use other arguments when possible. Counting points over $\mathbf{F}_{p^2}$ for various primes $p \nmid N$, as in the case where N is prime, is a very helpful in this regard, because the computations are much faster.

Combining these techniques, we first compute the N for which $X_0(N)$ is trigonal and one of 2,3, or 5 does not divide N. This is done the same way as the case where N is prime, and we only have to compute the dimension of induced cubic relations a handful of times because counting points usually does the job. We are now done by combining Lemma 3.8 with the fact that by Petri's theorem, $X_0(N)$ is not trigonal for $N = 2\cdot 3^2\cdot 5, 3\cdot 5^2, 2^2\cdot 3\cdot 5$. This proves Theorem 3.12. ■

Remark. Combining Theorem 3.12 with Petri's Theorem actually provides a shortcut for finding equations for the canonical embedding of $X_0(N)$. In the case where $g_0(N) > 4$ and $g_0(N) \neq 6$, we see that the canonical ideal is generated by degree 2 relations. Using this fact, one needs only to determine the quadratic relations among differentials (cusp forms); it is not necessary to look for any higher relations.

3.3 Trigonal $X_0^+(N)$

In this section we make some further remarks about the gonality of modular curves, and in particular study when $X_0^+(N)$ is trigonal for N a prime number.

Notice a pattern one observes in looking at hyperelliptic and trigonal $X_0(N)$. In the latter case, one may rephrase Theorem 3.12 in a loose way by saying that $X_0(N)$ is trigonal if and only if "it has to be", in the sense that *any* (nonhyperelliptic) curve of

that genus is trigonal. And similarly, Ogg's result can be phrased as saying that $X_0(N)$ is only hyperelliptic when it has to be, meaning either that it has genus 2 (which explains $X_0(37)$ being hyperelliptic, for example), or a quotient by some Atkin–Lehner involution has genus zero. So one is tempted to speculate that the gonality of $X_0(N)$ is governed by only two factors: the gonality of a general curve of genus g, and the presence of Atkin–Lehner involutions. We remark that the general curve of genus g has gonality equal to $[\frac{g+3}{2}]$ (which is a special case of the results of [22]).

Let us restrict ourselves now to the case when N is prime. It seems plausible that when $X_0^+(N)$ has genus at least 3 it will have no automorphisms. (This seems to be an open question.) As a special case of this statement, it is a theorem that $X_0^+(N)$ is hyperelliptic precisely when it has genus 2. This is proved in [29], where the reader can in fact find a more general statement which applies when N is square-free. One might ask, more generally, if $X_0^+(N)$ has gonality *equal* to that of the general curve of genus g^+, namely $[\frac{g^+ +3}{2}]$. (The gonality of $X_0^+(N)$ is, in any case, an asymptotically linear function of g^+ by the results of [1]). More generally, one might wonder if the gonality of $X_0(N)$, N prime, is simply the minimum of $[\frac{g+3}{2}]$ and $2 \cdot [\frac{g^+ +3}{2}]$.

In fact, however, the situation is more complicated. More precisely, we have the following result, which shows that the gonality of $X_0^+(N)$ is *not* always the same as that of the general curve of genus g^+:

Theorem 3.14. *Let N be a prime number. Then $X_0^+(N)$ is trigonal iff $2 \leq g^+ \leq 4$ or $g^+ = 5$ and N is either 181 or 227.*

Remark. The exact list of primes N satisfying the conditions of the theorem is as follows:

$g^+ = 2 \Leftrightarrow N \in \{67, 73, 103, 107, 167, 191\}$

$g^+ = 3 \Leftrightarrow N \in \{97, 109, 113, 127, 139, 149, 151, 179, 239\}$

$g^+ = 4 \Leftrightarrow N \in \{137, 173, 199, 251, 311\}$;

To get an upper bound for the set of prime N such that $g_0^+(N) = \alpha$, one can use the Weil bound: if $g_0^+(N) = \alpha$ then $X_0(N)_{\mathbf{F}_2}$ admits a degree 2 map defined over $\mathbf{F}_2$ to the genus α curve $X_0^+(N)_{\mathbf{F}_2}$, so $X_0(N)(\mathbf{F}_4)$ can have at most $2(5 + 4\alpha)$ elements. On the other hand, we know that $X_0(N)(\mathbf{F}_4)$ has at least $2 + \frac{N+1}{12}$ elements. It follows that $N < 96(\alpha + 1)$.

Proof. The theorem follows from the same arguments as in the proof of Theorem

3.12. If $X_0(N)$ is trigonal and $g^+ \geq 5$, then from the Correspondence Lemma $X_0^+(N)$ is actually trigonal over $\mathbb{Q}$, and reducing modulo 2 and estimating points over $\mathbf{F}_4$ gives a bound $N < 335$. Actually counting points on $X_0(N)$ over $\mathbf{F}_4$ and $\mathbf{F}_9$ shows that in fact $N \leq 277$. For the remaining values of N, we decide trigonality by computing generators for the canonical ideal and applying Petri's Theorem. ■

We remark that the degree three coverings $X_0^+(181) \to \mathbf{P}^1$ and $X_0^+(227) \to \mathbf{P}^1$ are not Galois. This follows from the following lemma:

Lemma 3.15. *The automorphism group of $X_0^+(N)$, N a prime number, is an elementary 2-group.*

PROOF. The group $\mathrm{Aut}(X_0^+(N))$ injects into the multiplicative group of the ring $\mathrm{End}(J_0^+(N))$. The latter is known to be a subring of $\prod K_i$, where the K_i are totally real number fields – this follows from the fact that $\mathbf{T} \otimes \mathbb{Q} = \mathrm{End}^0(J_0(N)) = \mathrm{End}^0(J_0^+(N)) \times \mathrm{End}^0(J_0^-(N))$, where $\mathrm{End}^0(J)$ means $\mathrm{End}(J) \otimes \mathbb{Q}$, since the simple factors of J are mutually non-isogenous. Therefore $\mathrm{Aut}(X_0^+(N))$ injects into the group $\prod \mu_i$, where μ_i is the group of roots of unity in K_i. As K_i is totally real, we have $\mu_i = \{\pm 1\}$, which proves the lemma. ■

We also remark that before having computed generators for the canonical ideal of any of the curves $X_0^+(N)$, we had already found evidence that $X_0^+(181)$ and $X_0^+(227)$ might be trigonal by an entirely different method. Since this method could be useful in analyzing the gonality of $X_0^+(N)$ in general, we describe it here. The reader should consider the rest of this section as work in progress, and the details are correspondingly sketchy!

For the rest of this section, let N be a prime number, let R denote the ring $\mathbb{Z}_N$, and let R^{ur} be the ring of integers of the completion of the maximal unramified extension of $\mathbb{Q}_N$ in a fixed algebraic closure $\overline{\mathbb{Q}}_N$.

There exists a model $X_0^+(N)_R$ whose special fiber $X_0^+(N)_{\mathbf{F}_N}$ is a copy of the j-line that intersects itself transversely at the g^+ points corresponding to conjugate pairs of supersingular elliptic curves defined over $\mathbf{F}_{N^2}$ but not $\mathbf{F}_N$ (see the appendix to [29]). Denote by $\widetilde{X_0^+(N)}$ the normalization of $X_0^+(N)_{\mathbf{F}_N}$, which as we just saw is isomorphic to $\mathbf{P}^1_{\mathbf{F}_N}$, and let $\{e_i, e_i'\}$ be the conjugate pairs of j-values of supersingular elliptic curves which are identified in the map $\widetilde{X_0^+(N)} \to X_0^+(N)_{\mathbf{F}_N}$, $(i = 1, \ldots, g^+)$.

Because the elliptic curves with "extra automorphisms" (i.e., those with j-invariant 0 or 1728) are defined over $\mathbf{F}_N$, $X_0^+(N)_R$ should in fact be *regular* over R. (We have not worked out the details of this.)

Assuming $X_0^+(N)_R$ is regular, we can use Lemma 3.5 to define the reduction mod N of a rational function f on $X_0^+(N)$ defined over R^{ur}.

In particular, if there is a rational function on $X_0^+(N)$ of degree d defined over R (resp. R^{ur}), then there is also a rational function on $X_0^+(N)_{\mathbf{F}_N}$ of degree d defined over $\mathbf{F}_N$ (resp. $\overline{\mathbf{F}}_N$). Such an object is simply a rational function $\tilde{f}$ of degree d on the normalization $\widetilde{X_0^+(N)}$ such that $\tilde{f}(e_i) = \tilde{f}(e_i')$ for $i = 1, \ldots, g^+$.

With the above discussion as motivation, we specialize now to the following concrete problem.

Let $\mathbf{F}$ be an algebraically closed field, let d, t be positive integers, and let x_i, y_i ($1 \leq i \leq t$) be distinct elements of $\mathbf{F}^*$ thought of as points in $\mathbf{P}^1_{\mathbf{F}}$. If h is a rational function of degree d on $\mathbf{P}^1_{\mathbf{F}}$, we say that h is a *Castelnuovo function of type (d,t)* if $h(x_i) = h(y_i)$ for $1 \leq i \leq t$ (see [22] for a justification of this nomenclature). What can we say about the existence or nonexistence of Castelnuovo functions of type (d, t) (with respect to the marked points x_i, y_i) on $P^1_{\mathbf{F}}$?

Suppose for example that $d = 1$. Then $h(x_1) = h(y_1)$ is impossible, so no Castelnuovo functions of type $(1, t)$ exist when $t \geq 1$.

If $d = 2$ and $t \geq 2$, then without loss of generality we have $h(x_1) = h(y_1) = 0$ and $h(x_2) = h(y_2) = \infty$. The divisor of h is then $(h) = (x_1) + (y_1) - (x_2) - (y_2)$, so h is determined up to scalars. For generic choices of x_3, y_3, then, we will not have $h(x_3) = h(y_3)$. In other words, for generic (x_i, y_i) we expect a Castelnuovo function of type $(2, t)$ to exist iff $t \leq 2$.

Consider now general d.

A rational function h of degree d on $\mathbf{P}^1$ can be thought of as a ratio

$$h = \frac{a_d z^d + a_{d-1} z^{d-1} + \cdots + a_1 z + a_0}{b_d z^d + b_{d-1} z^{d-1} + \cdots + b_1 z + b_0}$$

of polynomials of degree d. The conditions $h(x_k) = h(y_k)$ for $1 \leq k \leq t$ then give t equations of the form

$$L_k := \sum_{0 \leq j < i \leq d} c_{i,j,k}(a_i b_j - a_j b_i) = 0.$$

for various $c_{i,j,k} \in \mathbf{F}$. If we think of each $z_{ij} := a_i b_j - a_j b_i$ as a separate variable, then each L_k is a linear equation in $r := \binom{d+1}{2}$ variables. Moreover, there are certain quadratic relations between the z_{ij}, namely (fixing α and β) we have

$$Q_{ij} := z_{ij} z_{\beta\alpha} - z_{i\alpha} z_{\beta j} + z_{j\alpha} z_{\beta i} = 0$$

and these give nontrivial independent relations whenever $i > j$ and neither i, j is α or β.

If we think of the z_{ij} as coordinates of a projective space $\mathbf{P}^{r-1}$, then each of the $r - (2d-1)$ forms Q_{ij} is a hypersurface in $\mathbf{P}^{r-1}$. For there to be $a_0, a_1, \ldots, a_d, b_0, \ldots, b_d$ as above, we must have all $Q_{ij} = 0$ and $L_k = 0$ for $k = 1, \ldots, t$. If the points x_i, y_i are chosen generically, then the dimension of this intersection of hyperplanes and quadric hypersurfaces in $\mathbf{P}^{r-1}$ will be $(r-1) - (r - 2d + 1) - t = (2d-2) - t$. So for generic x_i, y_i, one expects this intersection to be empty when $t \geq 2d-1$. The converse is true as well; in other words, for generic values of x_i, y_i, there is a Castelnuovo function of type (d, t) iff $t \leq (2d-2)$.

For a rigorous treatment of the arguments just sketched, and an exposition of how all this relates to Grassmannians, Schubert calculus, and Brill-Noether theory (among other things) see [22] and [26, Sec. 5A].

For example, if we let $\{e_i, e_i'\}$ be the g^+ conjugate pairs of supersingular j-values not in $\mathbf{F}_N$, then putting $x_i = e_i$ and $y_i = e_i'$ we expect (assuming the e_i, e_i' are in sufficiently general position on the j-line) to find a Castelnuovo function h of type (d, g^+) iff $g^+ \leq (2d-2)$. This is equivalent to saying that the smallest d such that there exists a Castelnuovo function of type (d, t) is $[\frac{g^+ + 3}{2}]$.

However, the assumption that the e_i, e_i' are in "general position" is not always satisfied.

Here is a concrete example with $d = 3$. A Castelnuovo function of type $(3, g^+)$ exists on the reduction of $X_0^+(N)$ if and only if the closed subvariety of $\mathbf{P}^5$ defined by the equations

$$Q_{21} = z_{21} z_{30} - z_{20} z_{31} + z_{10} z_{32}$$

and

$$L_k = \sum_{0 \leq j < i \leq 3} (x_k^i y_k^j - x_k^j y_k^i) z_{ij} = 0$$

for $k = 1, \ldots g^+$ is nonempty.

We carried out this computation for a few values of N such that $g^+ = 5$. When $N = 157$ it turns out that there is no admissible function of degree 3, and this corroborates the fact that $X_0^+(157)$ is not trigonal. When $N = 181$, we were surprised to find that there *is* a Castelnuovo function of type $(3, 5)$, namely

$$h(z) = \frac{105z^3 + 51z^2 + z}{93z^2 + 134z + 1}.$$

As we mentioned, it turns out that in fact $X_0^+(181)$ is trigonal. Similarly, one can calculate that the reduction of the curve $X_0^+(227)$ also has a Castelnuovo function of type $(3, 5)$.

In addition to this technique of reducing mod N and thinking about rational functions on the reduced scheme, one could also try to study morphisms of degree $d > 3$ from $X_0(N)$ and $X_0^+(N)$ to $\mathbf{P}^1$ by replacing the systematic use of Petri's theorem by an application of Green's conjectures [21]. These conjectures say roughly that one should be able to see the Clifford index of a curve (and so gather some information about its gonality) by looking at higher syzygies of the canonical ideal.

Chapter 4

Cartier Points on Curves

4.1 Introduction

Throughout this chapter Y will denote a complete, nonsingular, and irreducible algebraic curve of genus $g > 0$ over the algebraically closed field k of characteristic p.

Eventually we will assume that $g \geq 2$.

Denote by W the g-dimensional k-vector space of regular differentials of Y/k, which is just the vector space $H^0(Y, \Omega^1_{Y/k})$. Also let $K(Y)$ denote the function field of Y, and let $\Omega^1_{K(Y)/k}$ be the $K(Y)$-vector space of meromorphic differentials on Y.

Recall (see, for example, [35, A2]) that there is a map $\mathcal{C}$ from $\Omega^1_{K(Y)/k}$ to itself called the *Cartier operator*. The existence of this map only uses the fact that k is perfect. It preserves the space W of regular differentials. The Cartier operator is additive and also satisfies the identity

$$\mathcal{C}(z^p \omega) = z \cdot \mathcal{C}(\omega)$$

for all $z \in K(Y)$ and $\omega \in \Omega^1_{K(Y)/k}$.

In particular, the action of $\mathcal{C}$ on the k-vector space W is σ^{-1}-linear, meaning that $\mathcal{C}(\sigma(\alpha)\omega) = \alpha \cdot \mathcal{C}(\omega)$, where $\alpha \in k$ and σ is the Frobenius (p-th power) map on k.

The action of the Cartier operator on W is dual to the action of the Frobenius map on $H^1(Y, \mathcal{O}_Y)$. So the action of $\mathcal{C}$ on W is intimately related to the structure of the p-torsion on the Jacobian of Y/k. A curve is called *ordinary* if the action of $\mathcal{C}$ on W is an isomorphism, and *superspecial* if $\mathcal{C}$ is the zero operator on W. The following is a nontrivial fact about superspecial curves (see [44]):

Theorem 4.1. *Y is superspecial if and only if the Jacobian of Y is isomorphic to a product of supersingular elliptic curves.*

In Section 4.2, we will give a new proof (and strengthening) of the following theorem of Ekedahl (see [18]):

Theorem 4.2. *Let Y be a curve of genus g over an algebraically closed field k of characteristic p whose Jacobian is isomorphic to a product of supersingular elliptic curves. Then $g \leq \frac{p(p-1)}{2}$. If Y is hyperelliptic then $g \leq \frac{(p-1)}{2}$.*

Remark. Ekedahl shows in [18], considering Fermat curves of degree $p+1$ and hyperelliptic curves of the form $y^2 = x^p - x$ that these bounds are in fact *sharp*.

Our proof of Ekedahl's theorem is based on the hypothesis that the Cartier operator annihilates W, and ultimately relies on general facts about base-point free pencils on curves. Ekedahl's proof, by contrast, has as its direct input the fact that $\mathrm{Jac}(Y)$ is isomorphic to a product of supersingular elliptic curves. He shows that this implies that Y descends down to $\mathbf{F}_{p^2}$ with Frobenius acting on $H^1_{et}(Y_{\mathbf{F}_{p^2}}, \mathbb{Z}_l)$ as multiplication by $\pm p$, which cannot happen (by an observation of Serre) if $g > \frac{p(p-1)}{2}$.

Now we come to the main object of study of this chapter.

Definition. A closed point P of Y is said to be a *Cartier point* if the hyperplane $W(P)$ of regular differentials vanishing at P is stable under the Cartier operator.

Example 4.3. *On an elliptic curve every point is a Cartier point, since $W(P)$ is always zero. On a superspecial curve as well, every point is a Cartier point, since any subspace of a vector space is stable under the zero operator!*

Notice that the Cartier operator $\mathcal{C}$ induces a σ-linear map $\mathcal{C}^*$ on the dual space W^* via the formula

$$(\mathcal{C}^*\lambda, \omega) = (\lambda, \mathcal{C}\omega)^\sigma.$$

There is a natural identification of $\mathcal{C}^*$ with the σ-linear Frobenius map acting on $W^* \cong H^1(Y, \mathcal{O}_Y)$ (see [56, Proposition 9]).

If Y is ordinary then there is also an induced map $\mathbf{P}(\mathcal{C})$ on the projective space $\mathbf{P}(W)$, which is the set of hyperplanes in W.

For a fixed choice of a differential $\nu \in W$, let λ_P be the element of W^* sending ω to $\frac{\omega}{\nu}(P)$. The image $j(P)$ in $\mathbf{P}(W)$ of λ_P depends only on P, not on the differential ν.

The map $j : Y \to \mathbf{P}(W)$ is just the canonical map of Y to projective space. The following lemma can be proved by unwinding definitons (or see the proof of Proposition 4.19):

Lemma 4.4. *Suppose Y is ordinary. A point $P \in Y(k)$ is a Cartier point if and only if $\mathbf{P}(\mathcal{C})(j(P)) = j(P)$.*

In other words, when Y is ordinary a Cartier point can be interpreted geometrically as a point on Y whose image via the canonical embedding is a fixed point of the action of Frobenius on the ambient projective space.

Our original motivation for our study of Cartier points was the following theorem of Coleman:

Theorem 4.5. *Suppose X is a (smooth proper) curve of genus $g \geq 2$ defined over the ring of integers R in an unramified finite extension K of $\mathbb{Q}_p$, $p \geq 3$. Let $\tilde{X}$ denote the special fiber of X, and suppose that $\tilde{X}$ is ordinary. If $P \in X(\bar{K})$ is a torsion point with respect to some embedding of X in $Jac(X)$ defined over K, and if P is ramified (i.e., the extension $K(P)/K$ is ramified), then the reduction $\tilde{P}$ of P is a Cartier point on $\tilde{X}$.*

Proof. Using Lemma 4.4, we see that this theorem is proved in parts (*ii*) and (*iii*) of [12, Proposition 3.6]. ■

In fact, Coleman also proves in [10] that under the hypoetheses of the above theorem, there can be no torsion points at all ramified at p if $p \geq 5$. [He also reaches this conclusion for $p > 2g$ without the hypothesis that $\tilde{X}$ is ordinary.] So really we have found Theorem 4.5 useful only when $p = 3$. Nonetheless, this is an interesting case. For instance, in Section 4.4 and Chapter 5 we apply Theorem 4.5 to some concrete examples pertaining to the Coleman–Kaskel–Ribet conjecture.

As another application of Theorem 4.5, consider the following theorem of Buium (see [7]):

Theorem 4.6. *Let $X \to J$ be an Albanese embedding defined over a number field K of a smooth proper curve of genus $g \geq 2$ into its Jacobian. Let $\mathfrak{p}$ be a prime of K such that $K/\mathbb{Q}$ is unramified at $\mathfrak{p}$, X/K has good reduction at $\mathfrak{p}$, and no point of $X(\bar{K})$ ramified at $\mathfrak{p}$ maps to a torsion point of J. Suppose $p = char(\mathfrak{p})$ is odd. Then*

$$\#(X(\bar{K}) \cap J(\bar{K})_{tors}) \leq p^{4g} \cdot 3^g \cdot [p(2g-2) + 6g] \cdot g!.$$

If $p = 2$, then the same formula holds with p^{4g} replaced by 64^g.

Buium applies Coleman's results to conclude that p in the above theorem can be taken to be any prime bigger than $2g$, or any ordinary prime at least 5. If X is ordinary at some prime lying over 3, then one can take p equal to 3 in Buium's bound if one can show that the reduction $\tilde{X}$ of X has no Cartier points. (Even if there are Cartier points on $\tilde{X}$, one can sometimes successfully show that they do not lift to torsion points on X. See Chapter 5 for examples of this.)

We also mention that Cartier points arise naturally when studying the moduli space of hyperelliptic curves in characteristic p. This is illustrated in the work of B. Chisala (see [8]).

4.2 Duality, Linear Systems, and Ekedahl's Theorem

In preperation for our extension of Ekedahl's theorem, we review some of the duality theory for curves and its relation to the Cartier operator in characteristic p. A good reference for this is [55].

Let k be a field and C/k a (complete nonsingular) algebraic curve with function field $K(C)$. Assume the genus g of C is positive. A *repartition* ξ is a family $\{\xi_P\}_{P\in C}$ of elements of $K(C)$ such that $\xi_P \in \mathcal{O}_P$ for almost all $P \in C$. Let R be the k-algebra of repartitions of $K(C)$. $K(C)$ is embedded diagonally in R, allowing one to view $K(C)$ as a subring of R.

If D is a divisor, write $R(D)$ for the vector subspace of R consisting of all $\xi = \{\xi_P\}$ such that $v_P(\xi_P) \geq -v_P(D)$. We set $L(D)$ equal to the vector space consisting of 0 and all rational functions f such that $(f) \geq -D$. We have $L(D) = H^0(C, \mathcal{L}(D))$, where $\mathcal{L}(D)$ is the subsheaf of the constant sheaf $K(C)$ whose stalk $\mathcal{L}(D)_P$ at a point $P \in C(\bar{k})$ is the set of rational functions f such that $v_P(f) \geq -v_P(D)$.

For any divisor D, $I(D) = H^1(C, \mathcal{L}(D))$ is canonically isomorphic to $R/(R(D) + K(C))$.

Let $\Omega^1_{C/k}$ be the sheaf of differentials of C, and let $\Omega^1_{K(C)/k}$ be the $K(C)$-vector space of meromorphic differentials on C. There is a pairing between elements ω of $\Omega^1_{K(C)/k}$ and repartitions $\xi \in R$ via

$$\langle \omega, \xi \rangle = \sum_{P\in C} \mathrm{Res}_P(\xi_P \omega).$$

Let $\Omega(D)$ be the k-vector space formed by 0 and the elements $\omega \in \Omega^1_{K(C)/k}$ such that

$(\omega) \geq D$. Then by [55, II.8 Theorem 2], for every divisor D the scalar product $\langle \omega, \xi \rangle$ puts the vector spaces $\Omega(D)$ and $I(D)$ in duality.

Now suppose that $C = Y$ is a curve over k, with k algebraically closed of characteristic $p > 0$. Then the Cartier operator $\mathcal{C}$ on the meromorphic differentials of Y satisfies

$$\langle \mathcal{C}\omega, \xi \rangle^p = \langle \omega, \xi^p \rangle.$$

Using this duality theory, we have the following proposition giving an alternate characterization of Cartier points.

Proposition 4.7. *A point $P \in Y(k)$ is a Cartier point if and only if there exists a rational function f on Y and a local parameter T at P such that f is regular outside P, and at P has the Laurent expansion*

$$f = T^{-p} + cT^{-1} + \text{holomorphic terms},$$

where $c \in k$.

Proof. By the duality between differentials and repartitions, we see that $\mathcal{C}$ preserves $W(P) \subset W$ if and only if the p-th power map on R preserves the subspace $R(P)+K(Y)$. Since $R(0)+K(Y)$ has codimension one in $R(P)+K(Y)$, and $R(0)+K(Y)$ is stable under p-th powers, this occurs if and only if the repartition $(0, 0, \ldots, (T^{-1})^p, 0, 0, \ldots)$ is in $R(P)+K(Y)$, where the nonzero component corresponds to the place P. This is easily seen to be equivalent to the existence of an f as in the statement of the proposition. ■

This proposition shows that every Cartier point P on Y, together with a choice of a local parameter T at P, gives rise to degree p rational function $f_{P,T}$ on Y. If $f'_{P,T}$ is another degree p rational function regular outside P and with Laurent expansion $T^{-p} + c'T^{-1} +$ holomorphic terms, then $f'_{P,T} = f_{P,T} + a$ for some constant $a \in k$ by the residue theorem. So we get a uniquely determined base-point free g^1_p on Y, depending only on P and T. It is interesting to consider when different Cartier points can give rise to equivalent g^1_p's.

Lemma 4.8. *With notation as above, suppose P, Q are distinct Cartier points on Y and choose local parameters T and T' at P and Q, respectively. If the corresponding g^1_p's are equivalent, then $[(P) - (Q)]$ is a p-torsion point of Jac(Y).*

PROOF. We can normalize $f_{P,T}$ so that it vanishes at Q. The hypothesis that the g^1_p's are equivalent means that there is an automorphism α of $\mathbf{P}^1$ such that $f_{Q,T'} = \alpha \circ f_{P,T}$. It follows that α maps 0 to ∞, and that the inverse image of 0 under $f_{P,T}$ consists of the point Q with multiplicity p. Hence the divisor of $f_{P,T}$ is $p(Q) - p(P)$, which proves what we want. ■

We will use this in conjuction with a general result concerning base-point free pencils on curves. We will say that a base-point free g^1_e and a base-point free g^1_f on C are *independent* if the corresponding map from C to its image C' in $\mathbf{P}^1 \times \mathbf{P}^1$ has degree one. Thinking about the geometry of $\mathbf{P}^1 \times \mathbf{P}^1$, we see that a g^1_e and g^1_f are independent iff there exists a Zariski-dense open set $U \subseteq C$ such that for all $P \in U$, there are divisors $E \in g^1_e$ and $F \in g^1_f$ with $E \cdot F = (P)$. Indeed, if the g^1_e and g^1_f are dependent, then $E \cdot F$ will always have degree at least 2. On the other hand, if the g^1_e and g^1_f are independent, we can take U to be the inverse image of the complement of the singular locus on C'.

The proof of the following result is very similar to results found in [15] and [3]. Unlike [15, Proposition 3], we do not assume the g^1_d's in question are complete, nor do we assume that the ground field has characteristic zero. We include a slight variant of the proof in [15] here to convince the reader that these extra assumptions are not necessary.

Proposition 4.9. *Let C be a (smooth) curve of genus g defined over an algebraically closed field k. If C has at least d mutually independent base-point free g^1_d's, then $g \leq d(d-1)/2$.*

Remark. This result is sharp, since a plane curve of degree $d+2$ with $d-1$ nodes has $d-1$ independent g^1_d's and genus equal to $d(d-1)/2 - 1$. Also, the independence hypothesis is necessary, since there exist double covers of elliptic curves of arbitrary genus and these have infinitely many g^1_4's.

Before giving the proof, we need two lemmas. Notation will be as follows: if $A \subset \mathbf{P}^r$, then $\langle A \rangle$ will denote the linear span of A. If $\phi : C \to \mathbf{P}^r$ is a morphism and D is an effective divisor on C, then $\langle \phi(D) \rangle$ is the intersection of all hyperplanes H in $\mathbf{P}^r$ such that either $\phi(C) \subset H$ or $\phi^*(H) \geq D$. If ϕ is an embedding we often write simply $\langle D \rangle$ for $\langle \phi(D) \rangle$.

The geometric Riemann–Roch theorem ([4, p. 12]) says that for any effective divisor D of degree d on C,

$$l(D) = 1 + \dim(|D|) = d - \dim(\langle j(D) \rangle)$$

where j is the canonical map to $\mathbf{P}^{g-1}$.

Lemma 4.10. *Let C be a non-hyperelliptic smooth curve of genus g and let g^1_d be a special base-point free linear system on C. Let g^1_e and g^1_f be independent base-point free linear systems on C. If dim($|g^1_d|$)=b, dim($|g^1_d + g^1_e|$)=a_1, and dim($|g^1_d + g^1_f|$)=a_2, then dim($|g^1_d + g^1_e + g^1_f|$) $\geq a_1 + a_2 + 1 - b$. Also, if r denotes the dimension of $|g^1_e|$ and s denotes the dimension of $|g^1_f|$, then* $\dim(|g^1_e + g^1_f|) \geq r + s + 1$.

Proof. Assume that C is canonically embedded. For any $x \in C$, denote by E_x, F_x the unique divisors in g^1_e, g^1_f, respectively, which contain x. As we saw previously, the assumption that g^1_e and g^1_f are independent is equivalent to the assertion that for almost all $x \in C$ we have $E_x \cdot F_x = (x)$. Let D be a divisor in the linear system g^1_d. The fact that D is special means that $\langle D \rangle$ is not all of $\mathbf{P}^{g-1}$. So for a point $x \in C$ to lie outside $\langle D \rangle$ is a nonempty Zariski-open condition. Also the condition that $x \in C$ is such that $E_x \cdot D = F_x \cdot D = \emptyset$ is non-empty and open. We conclude that there exists a point $x_0 \in C$ and divisors E, F in g^1_e, g^1_f, respectively, such that $E \cdot D = F \cdot D = \emptyset, E \cdot F = (x_0)$, and $x_0 \notin \langle D \rangle$.

We must have $\langle D + E + F - x_0 \rangle = \langle \langle D + E \rangle \cup \langle F \rangle \rangle$ since $D + E$ and $F - x_0$ are disjoint, and so

$$\dim(\langle D + E + F - x_0 \rangle) = \dim(\langle D + E \rangle) + \dim(\langle F \rangle) - \dim(\langle D + E \rangle \cap \langle F \rangle).$$

By the geometric Riemann–Roch theorem, we have

$$\dim(\langle D \rangle) = d - b - 1,$$

$$\dim(\langle D + E \rangle) = d + e - a_1 - 1,$$

$$\dim(\langle D + F \rangle) = d + f - a_2 - 1,$$

$$\dim(\langle F \rangle) = f - 1 - s.$$

So we have $\dim(\langle D \rangle \cap \langle F \rangle) = a_2 - b - 1 - s$. And as

$$\langle F \rangle \cap \langle D + E \rangle \supset \langle (\langle D \rangle \cap \langle F \rangle) \cup \{x_0\} \rangle$$

and $x_0 \notin \langle D \rangle$, we have $\dim(\langle F \rangle \cap \langle D + E \rangle) \geq a_2 - b - s$. Therefore

$$\dim(\langle D + E + F - x_0 \rangle) \leq d + e + f - a_1 - a_2 + b - 2.$$

So by geometric Riemann–Roch, we get

$$\dim(|D + E + F - x_0|) \geq a_1 + a_2 - b.$$

Since x_0 is not a base point of $|D + E + F|$, we see that

$$\dim(|D + E + F|) \geq a_1 + a_2 - b + 1.$$

as desired.

Finally, we have

$$\dim(\langle E \rangle) = e - 1 - r$$

and

$$\dim(\langle F - x_0 \rangle) = f - 2 - (s - 1),$$

so

$$\begin{aligned}\dim(\langle E + F - x_0 \rangle) &= \dim(\langle E \rangle) + \dim(\langle F - x_0 \rangle) - \dim(\langle E \rangle \cap \langle F - x_0 \rangle) \\ &\leq e + f - r - s - 2.\end{aligned}$$

Therefore by geometric Riemann–Roch,

$$\dim(|E + F - x_0|) \geq e + f - 2 - (e + f - r - s - 2) = r + s.$$

Since x_0 is not a base point for $|E + F|$, it follows that $\dim(|E + F|) \geq r + s + 1$ as claimed. ■

Lemma 4.11. *Let C/k be hyperelliptic. Then any special g^1_d factors through the (unique) g^1_2. In other words, the map $f : C \to \mathbf{P}^1$ coming from the g^1_d factors as $f = g \circ h$, where $h : C \to \mathbf{P}^1$ has degree 2.*

Proof. It suffices to show that $f(P) = f(Q)$ whenever $(P) + (Q) \in g^1_2$, since then f factors through the quotient of C by the hyperelliptic involution w. So if D is an effective divisor in g^1_d, we have to show for all $P \in C$ that $Q = w(P)$ is a base point of $|D - P|$. This is equivalent to the statement that $l(D - P - Q) = l(D - P)$, which by Riemann–Roch is equivalent to $l(E + P) - l(E + P + Q) = 1$, where E is a residual effective divisor to D, i.e., $D + E = K$. The desired equality follows immediately from the geometric Riemann–Roch theorem. ■

Corollary 4.12. *Suppose C is a hyperelliptic curve of genus g defined over an algebraically closed field k. Then any rational function of degree at most g on C must have even degree.*

We now give the promised proof of Proposition 4.9.

PROOF. Assume $g > \frac{d(d-1)}{2}$. and $d \geq 3$. Without loss of generality we may assume that $d \geq 3$, because for $d = 1$ the Proposition is trivial and for $d = 2$ it is equivalent to the fact that the g^1_2 on a hyperelliptic curve of genus at least 2 is unique, which is proved in [28, IV, 5.3]. Let $D_1, \ldots D_d$ be effective divisors on C corresponding to the d independent g^1_d's. As $d \geq 3$, it follows that $g \geq d$, and in particular each D_i is special by Riemann–Roch. Thus C is not hyperelliptic, since on a hyperelliptic curve no two special g^1_d's can be independent by Lemma 4.11.

Using Lemma 4.10 and induction on k, we see that if $k \geq 3$ and if $|D_1 + \ldots + D_{k-2}|$ is special then

$$\dim(|D_1 + \ldots + D_k|) \geq \binom{k+1}{2}.$$

By the induction hypothesis, we have

$$\dim(|D_1 + \ldots + D_{k-2}|) \geq \binom{k-1}{2}$$

and so $|D_1 + \ldots D_{k-2}|$ is special if $g > k(k-2) - \binom{k-1}{2} = \frac{(k+1)(k-2)}{2}$. Therefore we find that

$$\dim(|D_1 + \ldots + D_d|) \geq \frac{d(d+1)}{2}.$$

Clifford's theorem implies that $|D_1 + \ldots + D_d|$ is nonspecial, so by Riemann–Roch we find that $g \leq \frac{d(d-1)}{2}$, a contradiction. ∎

Using Corollary 4.12, we deduce the following proposition, which generalizes results of Coleman (see [9, 5.5]).

Proposition 4.13. *Let Y be an algebraic curve of genus $g \geq 2$ over an algebraically closed field k of characteristic p.*

1. *If $p = 2$, then $P \in Y(k)$ is a Cartier point if and only if Y is hyperelliptic and P is a hyperelliptic branch point.*

2. *If Y/k is hyperelliptic and $p \neq 2$, then:*

   (a) *If $p \leq g$, then there are no Cartier points on Y.*

   (b) *If $p < 2g$, the hyperelliptic branch points on Y are not Cartier points.*

PROOF. The first part of the proposition follows immediately from Proposition 4.7. Now consider the second part. If P is a Cartier point on Y then there exists a rational function f on Y regular outside P and with a pole of order p at P. As p is odd, the statement in part (a) follows from Corollary 4.12, and the statement in part (b) follows from the fact that the Weierstrass gap sequence at a hyperelliptic branch point P is $1, 3, \ldots, 2g-1$, so in particular there is no rational function regular outside P and with a pole of order $p < 2g$ at P. ■

We now state our strengthening of Ekedahl's results.

Theorem 4.14. *Let Y be a curve of genus g over an algebraically closed field of characteristic p.*

1. *If Y has at least p distinct Cartier points, no two of which differ by a p-torsion point on Jac(Y), then $g \leq \frac{p(p-1)}{2}$.*

2. *If Y is hyperelliptic of genus g, p is odd, and some hyperelliptic branch point of Y is a Cartier point, then $g \leq \frac{p-1}{2}$.*

PROOF. The first assertion follows from Lemma 4.8 and Proposition 4.9 after noting that any two nonequivalent g^1_d's are independant if d is a prime number. The second follows immediately from Proposition 4.13(b). ■

Corollary 4.15 (Ekedahl). *Let Y be a superspecial curve of genus g over an algebraically closed field of characteristic p. Then $g \leq \frac{p(p-1)}{2}$. If Y is hyperelliptic then $g \leq \frac{p-1}{2}$ or $p = 2$ and $g = 1$.*

PROOF. If Y is superspecial then $\mathcal{C} = 0$ on W, so every point on Y is a Cartier point. Also the Jacobian of Y has no p-torsion, so each Cartier point gives rise to a distinct g^1_p on Y by Lemma 4.8. The conclusions of the corollary now follow from the theorem. ■

Remark. If the results of [16] are valid in arbitrary characteristic (they should be, but we have not checked this), then it follows from our proof that if Y is superspecial and $g = \frac{p(p-1)}{2}$, then Y is a nonsingular plane curve of degree $p+1$. In [63] it is proved that if Y is superspecial and hyperelliptic with $g = \frac{p-1}{2}$, then Y is defined by an equation $y^2 = x^p - x$.

4.3 Bounds

We keep the same general notations as before. In particular Y will denote a curve of genus at least 2 over the algebraically closed field k of characteristic p, though we may repeat this information for emphasis.

It is clear from Lemma 4.4 that an ordinary curve can only possess finitely many Cartier points. On the other hand, every point on a superspecial curve is a Cartier point. What is the general picture? In this section we show that a curve has infinitely many Cartier points if and only if it is superspecial. We also give some bounds for the number of Cartier points a non-superspecial curve can possess.

Before we begin, we need a few facts from σ-linear algebra which can be found in [56, Sec. 9] and [5].

Let k be an algebraically closed field of characteristic p and let σ be the Frobenius endomorphism of k. Let T be a σ or σ^{-1}-linear operator on the k-vector space V of dimension g. Then the kernel of T is a linear subspace of V, and the rank s of T is well-defined as g minus the dimension of the kernel of T. The vector space V has a unique direct-sum decomposition $V = V_1 \oplus V_2$, where the action of T is bijective on V_1 and nilpotent on V_2. The dimension r of V_1 is called the Hasse–Witt invariant of T. Furthermore, there exists a basis $\{v_1, \dots v_r\}$ of V_1 such that $Tv_i = v_i$ for $i = 1, 2, \dots, r$. The solutions $x \in V$ to $Tx = x$ are exactly the p^r elements of the $\mathbf{F}_p$-vector space generated by $v_1, \dots, v_r$. The induced operator T^* on V^* has the same invariants r and s as T.

By convention, we say that the Hasse–Witt invariant of the curve Y is the Hasse–Witt invariant of the Cartier operator acting on W, or equivalently of the Frobenius operator acting on $W^* = H^1(Y, \mathcal{O}_Y)$.

We make the following definition.

Definition. Suppose Y is ordinary. A set of *canonical coordinates* for Y is a choice of a basis for $H^1(Y, \mathcal{O}_Y)$ consisting of fixed points of $\mathcal{C}^*$. Equivalently, we say that the map from Y to $\mathbf{P}^{g-1}$ given by sending P to $[\omega_1(P) : \omega_2(P) : \cdots : \omega_g(P)]$ is given in canonical coordinates if $\omega_1, \dots, \omega_g$ is a basis for W consisting of differentials fixed by $\mathcal{C}$ (whose existence is guaranteed by our facts from σ-linear algebra).

The following fact follows immediately from Lemma 4.4.

Lemma 4.16. *Assume Y is ordinary. A point P on Y is a Cartier point if and only if it*

maps to an $\mathbf{F}_p$-rational point with respect to some canonical coordinates for Y.

If Y is a curve of genus $g \geq 2$, we will define δ_Y to be the degree of the canonical map $Y \to \mathbf{P}^{g-1}$, namely 2 if Y is hyperelliptic and 1 otherwise.

Proposition 4.17. *Suppose Y/k is ordinary. Then the number of Cartier points on Y is bounded above by*

$$\delta_Y \cdot \mathrm{Max}(p + 1 + 2g\sqrt{p}, (\frac{2g-2}{\delta_Y})^2).$$

PROOF. (See Proposition 5.1 of [9]). Map Y to $\mathbf{P}^{g-1}$ using canonical coordinates and call the image curve E, so that the Cartier points on Y are just those points of Y which map to $\mathbf{F}_p$-rational points of E by Lemma 4.16. Now if E is defined over $\mathbf{F}_p$, then the Riemann hypothesis gives a bound of $p+1+2g\sqrt{p}$ for the number of rational points on E. Otherwise, $E^\sigma \neq E$, where σ is the Frobenius map on $\mathbf{P}^{g-1}$. In this case, by projecting E down to a suitable two-dimensional plane we can apply Bezout's theorem to estimate the number of points of E fixed by the Frobenius of $\mathbf{P}^{g-1}$. The degree of both E and E^σ is $\frac{2g-2}{\delta_Y}$, so the cardinality of $E \cap E^\sigma$ is bounded by $(\frac{2g-2}{\delta_Y})^2$. ■

Remark. We do not know of any examples of E as in the above proof which are defined over the prime field $\mathbf{F}_p$. It would be interesting to find such a curve or to prove that none exists. To get a flavor for this remark, see Example 4.26 in the next section.

When p is small relative to g, which is the case in our applications to torsion points on curves (where $p = 3$), the following bound is better.

Proposition 4.18. *Suppose Y/k is ordinary. Then there are at most $(g-2)+(p+1)g$ Cartier points on Y.*

PROOF. Let $\mathbf{F} := \mathbf{F}_p$. The assumption that Y is ordinary implies that $W^{\mathcal{C}} \otimes_{\mathbf{F}} k \cong W$, where $W^{\mathcal{C}}$ denotes the space of logarithmic differentials, i.e., the elements of $W := H^0(Y, \Omega^1_{Y/k})$ fixed by the Cartier operator. We have that P is a Cartier point iff $W(P)^{\mathcal{C}} \otimes_{\mathbf{F}} k \cong W(P)$.

If there are fewer than $g-2$ Cartier points on Y then we are done. Otherwise, let $Q_1, \ldots, Q_{g-2}$ be distinct Cartier points. The k-vector space of differentials vanishing at each of the Q_i has dimension at least 2 by Riemann–Roch, and has a basis $\{\omega_i\}$ of logarithmic differentials.

Pick any two such logarithmic ω_i; they generate a 2-dimensional $\mathbf{F}$-vector space V. Let $\{v_1, \ldots, v_{p+1}\}$ be elements of V representing distinct elements of $\mathbf{P}^1(V)$.

Now let P be a Cartier point different from $Q_1, \ldots, Q_{g-2}$. As $\dim_{\mathbf{F}} W(P)^{\mathcal{C}} + \dim_{\mathbf{F}} V = g + 1$, we have $W(P)^{\mathcal{C}} \cap V \neq (0)$, so some nonzero element of V is in $W(P)$.

But then one of $v_1, \ldots, v_{p+1} \in W(P)$. So

$$P \in \bigcup_{i=1}^{p+1} \{\text{zeros of } v_i\}.$$

Now each of the v_i can vanish at most at g points different from the Q_i. So we find that a Cartier point on Y must be one of the $g-2$ points $Q_1, \ldots, Q_{g-2}$, or else one of at most $(p+1)g$ other points where some v_i has a zero. This gives the desired bound. ■

A curve Y is called *extraordinary* if it is neither ordinary nor superspecial (see [10]).

Proposition 4.19. *Let Y be an extraordinary curve with Hasse–Witt invariant r. Then the number of Cartier points on Y is bounded by*

$$\text{Min}(4g - 4, 2g - 2 + \delta_Y \cdot \frac{p^r - 1}{p - 1}).$$

Furthermore, let s be the rank of $\mathcal{C}$ acting on W. If $s = 1$ then there is at least one Cartier point on Y.

Proof. Given a hyperplane H in W, let $f_H \in W^*$ denote any nonzero element of the 1-dimensional subspace of W^* which annihilates H under the natural pairing. Also let $\bar{f}_H$ denote the image of f_H in $\mathbf{P}(W)$. It follows from the definition of $\mathcal{C}^*$ that H is stable under $\mathcal{C}$ if and only if $\mathcal{C}^* f_H = c \cdot f_H$ for some $c \in k$.

If $c \neq 0$ then the solutions of $\mathcal{C}^* y = cy$ correspond bijectively to the solutions of $\mathcal{C}^* x = x$ by setting $x = c^{\frac{-1}{p-1}} y$ for some fixed $(p-1)$-st root of c. So the set S of hyperplanes $H \subset W$ such that $\mathcal{C}^* f_H = c \cdot f_H$ for some $c \neq 0$ corresponds bijectively to the $\frac{p^r - 1}{p-1}$ values of $\bar{f}_H$ such that $\mathcal{C}^* f_H = f_H$, where r is the Hasse–Witt invariant for $\mathcal{C}^*$. Now pick a basis $\omega_1, \omega_2, \ldots, \omega_g$ for W such that $\omega_1, \ldots, \omega_r$ are fixed by the Cartier operator, and map Y to its image E in $\mathbf{P}^{g-1}$ using these coordinates. Then the points of S which lie on E are just the points of E with coordinates $(x_1 : \cdots : x_g)$ such that $x_1, \ldots, x_r \in \mathbf{F}_p$ and $x_{r+1} = \cdots = x_g = 0$. In particular, since $r < g$, the intersection $S \cap E$ lies in the

intersection of E with the hyperplane $x_g = 0$. Since the degree of E in $\mathbf{P}^{g-1}$ is $\frac{2g-2}{\delta_Y}$, we have $\#(S \cap E) \leq \text{Min}(\frac{p^r-1}{p-1}, \frac{2g-2}{\delta_Y})$.

On the other hand, the set of H such that $\mathcal{C}^* f_H = 0$ is exactly the set of H such that $f_H \in K$, where K is the kernel of $\mathcal{C}^*$, a vector space of dimension $g - s$. This set forms a projective space L inside $\mathbf{P}(W)$ of dimension $g - s - 1$. Note that the intersection $L \cap E$ is empty, since every point in L has $x_1 = \cdots = x_r = 0$ (notation as above). By hypothesis the dimension of L is between 0 and $g - 2$, inclusive. Since the image E of Y inside $\mathbf{P}(W)$ does not lie in any hyperplane, the intersection $L \cap E$ is finite, and its cardinality is bounded by the degree of E, which is $\frac{2g-2}{\delta_Y}$.

We are now done, because the Cartier points on Y are exactly those points which map to points in $E \cap (S \cup L)$. If $s = 1$ then L is a hyperplane in $\mathbf{P}(W)$ and therefore intersects E nontrivially. ■

Corollary 4.20. *The set of Cartier points on Y is Zariski-closed, and is infinite if and only if Y is superspecial.*

Remark. We mention an application to modular forms. If we fix p it follows from the above proposition that the rank s of the Hecke operator T_p mod p on weight two cusp forms for $\Gamma_0(N)$ is at least 2 for all prime numbers N sufficiently large with respect to p. For T_p mod p is the Cartier operator on $H^0(X_0(N), \Omega^1_{X_0(N)/\bar{\mathbf{F}}_p})$ for $p \neq N$ by the Eichler-Shimura relation. So if $s \leq 1$ then there exist Cartier points on $X_0(N)_{\bar{\mathbf{F}}_p}$, hence $X_0(N)_{\bar{\mathbf{F}}_p}$ admits a degree p morphism to $\mathbf{P}^1$. But one can show using the techniques of Chapter 3 that this is not the case for $N >> p$.

4.4 Examples

In this section we provide some examples of Cartier points on curves, along with some techniques for calculating them. We already know from Proposition 4.13 that if $p \neq 2$ and $p \leq g$, there are no Cartier points on a hyperelliptic curve, and that when $p = 2$ the Cartier points are just the hyperelliptic branch points. By explicit calculations one can determine exactly what the Cartier points on a given hyperelliptic curve are.

Example 4.21. *Cartier points on ordinary hyperelliptic curves.*

Let Y be an ordinary hyperelliptic curve over the (algebraically closed) field k of characteristic $p \neq 2$. Suppose f is given by the equation $y^2 = f(x)$, with f of degree $2g+2$ and having distinct roots. Write $f^{\frac{p-1}{2}} = \sum_{i=0}^{p-1} g_i^p x^{p-1-i}$, where the g_i are elements of $k[x]$. Let $\omega = \frac{dx}{2y}$, so that $\{\omega, x\omega, \ldots, x^{g-1}\omega\}$ is a basis for the regular differentials on Y. It is well-known that $\mathcal{C}(x^{pk+i}\omega) = x^k g_i \omega$ for $0 \leq i \leq p-1$ (see [65]). Since $V = \{\omega, x\omega, \ldots, x^{g-2}\omega\}$ is a basis for the differentials which vanish at ∞ (and the same for ∞'), the two points at infinity are Cartier points if and only if V is stable under $\mathcal{C}$, if and only if $\deg(x^k g_i) \leq g-2$ whenever $0 \leq pk+i \leq g-2$. For example, when $p=3$ and $g=2$, the points at infinity are Cartier points if and only if g_0 is a constant, i.e. $\deg(f'(x)) \leq 4$.

Now let $P = (x(P), y(P))$ be a point on Y with $x(P), y(P) \in k$. Let $h_P \in W^*$ be the function given by the rule $h_P(\nu) = \frac{\nu}{\omega}(P)$. Then

$$\mathcal{C}^*(h_P)(x^{pk+i}\omega) = h_P(\mathcal{C}(x^{pk+i}\omega))^p = (h_P(x^k g_i \omega))^p = (x^k g_i)(P)^p$$

for $0 \leq pk+i \leq g-1$. Using Lemma 4.4, we see that P is a Cartier point if and only if there exists a constant $c \in k^*$ such that $x(P)^{pk+i} = c \cdot x(P)^{pk} g^i(P)^p$ for $0 \leq pk+i \leq g-1$. This happens if and only if there exists $c \in k^*$ such that $x(P)^i = c \cdot g_i(P)^p$ for all $i = 1, 2, \ldots, m := \min(p-1, g-1)$, if and only if $g_0(P) \neq 0$ and the polynomials $h_i := x g_i^p - g_{i+1}^p$ vanish at P for $i = 0, 1, \ldots, m-1$.

Remark. One can adapt these techniques to the non-ordinary case. One can also use the calculations given in this example to give a computational proof of part 2 of Proposition 4.13.

As a special case of our computations, we get the following proposition, which slightly generalizes [9, Theorem 5.5] and is relevant to the study of torsion points on genus 2 curves.

Proposition 4.22. *Let Y be an ordinary genus 2 curve over a field k of characteristic 3 given by the equation $y^2 = f(x)$, with f of degree 6 and having distinct roots. Then the Cartier points on Y are the points whose x-coordinates are roots of $f'(x)$, plus the two points at infinity when* $\deg(f'(x)) \leq 4$.

PROOF. The only things to note in addition to what we have already done are that $f'(x) = g_1^3 - x g_0^3$ and that g_0 and g_1 cannot simultaneously vanish at any point when Y is ordinary. ■

We now give some examples of Cartier points on nonhyperelliptic ordinary curves. We need the following lemma on plane curves.

Lemma 4.23. *Let Y/k be a plane curve given by the degree d homogeneous polynomial f. Then the following diagram commutes:*

$$\begin{array}{ccccc} H^1(Y,\mathcal{O}_Y) & \xrightarrow{a\mapsto a^p} & H^1(Y,\mathcal{O}_Y^p) & \xrightarrow{\quad i\quad} & H^1(Y,\mathcal{O}_Y) \\ {\scriptstyle\cdot f}\uparrow & & & & \uparrow{\scriptstyle\cdot f} \\ H^2(\mathbf{P}^2,\mathcal{O}_{\mathbf{P}^2}(-d)) & \xrightarrow{g\mapsto g^p} & H^2(\mathbf{P}^2,\mathcal{O}_{\mathbf{P}^2}(-dp)) & \xrightarrow{\cdot f^{p-1}} & H^2(\mathbf{P}^2,\mathcal{O}_{\mathbf{P}^2}(-d)) \end{array}$$

Proof. From the exact sequence of sheaves

$$0 \to \mathcal{I}_Y \to \mathcal{O}_{\mathbf{P}^2} \to \mathcal{O}_Y \to 0$$

we see that

$$H^1(Y,\mathcal{O}_Y) \xrightarrow{\sim} H^2(\mathbf{P}^2,\mathcal{I}_Y).$$

As $\mathcal{O}_{\mathbf{P}^2}(-d) \xrightarrow{\sim} \mathcal{I}_Y$ (the map is multiplication by f), we have

$$H^2(\mathbf{P}^2,\mathcal{I}_Y) \cong H^2(\mathbf{P}^2,\mathcal{O}_{\mathbf{P}^2}(-d)).$$

The remaining details are explained in the proof of [28, IV, 4.21]. ■

Remark. Note that the map $F : \mathcal{O}_Y \to \mathcal{O}_Y$ is the p-th power map $\mathcal{O}_Y \to \mathcal{O}_Y^p$ followed by the inclusion map $\mathcal{O}_Y^p \hookrightarrow \mathcal{O}_Y$. So the top row of the above exact sequence gives the action of F on $H^1(Y,\mathcal{O}_Y)$, which is dual to the action of $\mathcal{C}$ on $H^0(Y,\Omega_Y)$. The composite map in the bottom row of the diagram can be calculated quite explicitly using the basis described in [28, Section III.5] for the cohomology of projective space.

Assume that Y is ordinary, and choose a set of coordinates for $\mathbf{P}(W)$, in other words fix a basis B for $H^1(Y,\mathcal{O}_Y)$. We can then consider the matrix $A = A_B$ giving the action of $\mathcal{C}^*$ with respect to the basis B. A is called the Hasse–Witt matrix of Y with respect to the basis B. It is easy to see from σ-linear algebra that there will be exactly $\frac{p^g-1}{p-1}$ projective solutions to the equation $Av^{(p)} = v$, where $v \in H^1(Y,\mathcal{O}_Y)$ is expressed in terms of the basis B, and $v^{(p)}$ means the vector obtained from v by raising all coordinates to the p-th power (See Section 4.3 for a more general statement). Using Lemma 4.4, we have the following proposition.

Proposition 4.24. *Assume Y is ordinary. With notation as in the preceding paragraph, a point $P \in Y(k)$ is a Cartier point if and only if its coordinate vector $v = j(P)$ with respect to the basis B satisfies the equation $Av^{(p)} = v$.*

Remark. This proposition is useful for calculations because given A one can algorithmically find the vectors v such that $Av^{(p)} = v$. This is discussed in [33, Proof of Satz 12] and [65, Proof of Theorem 2.1].

Example 4.25. *The points (0:0:1) and (1:0:0) are the Cartier points in characteristic 3 on the genus three plane curve Y given by the equation*

$$f(x, y, z) = y^3z + xy^2z + x^3y + 2x^2yz + xz^3 = 0.$$

PROOF. It is straightforward to check that the given curve is nonsingular.

We can compute the action of Frobenius on $H^1(Y, \mathcal{O}_Y)$ with respect to the basis $\{x^{-2}y^{-1}z^{-1}, x^{-1}y^{-2}z^{-1}, x^{-1}y^{-1}z^{-2}\}$ for $H^2(\mathbf{P}^2, \mathcal{O}_{\mathbf{P}^2}(-4)))$ using Lemma 4.23. The discussion in [28, Section III.5] shows that this basis corresponds to the canonical embedding given by f with respect to the coordinates x, y, z for $\mathbf{P}^2$.

With respect to this basis, Frobenius acts on $H^1(Y, \mathcal{O}_Y)$ via the matrix A equal to

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

Thus Y is ordinary. By Proposition 4.24, the Cartier points on Y are just the points with projective coordinates $(a : b : c)$ such that $f(a, b, c) = 0$ and $Av = v^{(p)}$, where v is the column vector $(a, b, c)^t$. Using the algorithm from [65], or simply computing by hand, one finds that the solutions to $Av^{(p)} = v$ correspond to the 13 projective points

$$\{(0 : 0 : 1), (1 : 0 : 0), (1 : 0 : i), (1 : 0 : -i), (\alpha : i : \gamma)\}$$

where i is a fixed square root of -1 and $\alpha^3 - \alpha = -i$, $\gamma^3 + \gamma = \alpha$. A computation now shows that of these points, only $(0 : 0 : 1)$ and $(1 : 0 : 0)$ lie on the curve.

We verify from another point of view that these two points are Cartier points. To see that $P = (0 : 0 : 1)$ is a Cartier point, we think of P as the origin on the affine plane curve given by the equation $y^3 + xy^2 + x^3y - x^2y + x = 0$ and apply Lemma 4.7. A calculation shows that the divisors of the degree three affine coordinate functions x and y are

$$\begin{aligned}(x) &= 3(0:0:1) - (1:0:0) - 2(0:1:0)\\ (y) &= (0:0:1) + 2(1:0:0) - 3(0:1:0).\end{aligned}$$

So y is a uniformizing parameter at P, and since

$$\frac{y^3}{x} = -1 - y^2 + y(x - x^2) = -1 - y^2 + y^4(\text{unit}),$$

we have

$$\frac{-1}{x} = y^{-3} + y^{-1} + \text{holomorphic}$$

is the Laurent expansion of $\frac{-1}{x}$ in terms of y, and this function has a pole only at P. Thus P is a Cartier point.

One can check in a similar way that $\frac{y}{z}$ is a uniformizer at $Q = (1:0:0)$, that $\frac{z-1}{y^2}$ is regular outside Q, and that at Q it has a Laurent expansion

$$\frac{z-1}{y^2} = (\frac{y}{z})^{-3} + (\frac{y}{z})^{-1} + \text{holomorphic}$$

so that Q is also a Cartier point. ■

Example 4.26. *The plane curve Y in characteristic 3 given by the equation*

$$y^4 - z^4 + x^2y^2 - x^2z^2 + \alpha \cdot xy^3 = 0.$$

with $\alpha \notin \mathbf{F}_3$ defines a nonsingular curve of genus 3 possessing exactly 5 Cartier points.

PROOF. One easily checks that the given curve is nonsingular; this would be false if we allowed α to lie in $\mathbf{F}_3$.

We can compute the action of Frobenius on $H^1(Y, \mathcal{O}_Y)$ with respect to the ordered basis $\{x^{-2}y^{-1}z^{-1}, x^{-1}y^{-2}z^{-1}, x^{-1}y^{-1}z^{-2}\}$ for $H^2(\mathbf{P}^2, \mathcal{O}_{\mathbf{P}^2}(-4)))$ using Lemma 4.23. The discussion in [28, Section III.5] shows that this is exactly the ordered basis of $H^1(Y, \mathcal{O}_Y)$ determined by choosing the coordinates $\{x, y, z\}$ for $\mathbf{P}(W) \cong \mathbf{P}^2$.

We find that Frobenius acts on the vector space $H^1(Y, \mathcal{O}_Y)$ via the 3×3 identity matrix. Thus not only is Y is ordinary, but in fact it is embedded in $\mathbf{P}^2$ via canonical coordinates. So by Lemma 4.16, the Cartier points on Y are just the points whose coordinates lie in $\mathbf{F}_3$. The only such points are $\{(0:0:1), (0:1:0), (0:1:1), (0:1:-1), (1:0:0)\}$.

■

The following example is motivated by the Coleman–Kaskel–Ribet conjecture (see Chapter 2).

Example 4.27. *There are no Cartier points on the modular curves $X_0(43)$ or $X_0(61)$ in characteristic 3.*

PROOF. From [20], we find models for the canonical embeddings of these two curves in characteristic 3. The curve $Y = X_0(43)$ in characteristic 3 is given by the homogeneous quartic equation

$$f(x,y,z) = x^4 + 2x^3y + 2x^2y^2 + 2x^2yz + x^2z^2 + xy^3 + 2xy^2z + xyz^2 + y^3z + 2y^2z^2 + z^4 = 0.$$

With respect to the basis $\{x^{-2}y^{-1}z^{-1}, x^{-1}y^{-2}z^{-1}, x^{-1}y^{-1}z^{-2}\}$ for $H^2(\mathbf{P}^2, \mathcal{O}_{\mathbf{P}^2}(-4))$ we find that Frobenius acts on $H^1(Y, \mathcal{O}_Y)$ via the matrix A equal to

$$\begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

In particular, Y is ordinary. It follows from Lemma 4.4 that the Cartier points on Y are just the points on Y whose projective coordinates (a,b,c) satisfy the equation $Av^{(p)} = v$, where $v = (a,b,c)^t$. The algorithm given in [33] (which can also be found in [65]), which is easily implemented on a computer, allows one to calculate the $\frac{3^3-1}{3-1} = 13$ projective points v such that $Av^{(p)} = v$. Having done this, one discovers that none of these points $v = (a,b,c)^t$ satisfy $f(a,b,c)$=0. In other words, there are no Cartier points on Y.

The curve $Y := X_0(61)$ in characteristic 3 is a complete intersection in $\mathbf{P}^3$ of the two hypersurfaces:

$$g(x,y,z,w) = w^2 + 2x^2 + 2xy + z^2 = 0,$$

and

$$h(x,y,z,w) = x^2z + xy^2 + xyz + 2xz^2 + y^2z + 2yz^2 = 0.$$

An argument similar to the proof of Lemma 4.23 shows that the action of Frobenius on $H^1(Y, \mathcal{O}_Y)$ with respect to the basis B which corresponds to the given coordinates x, y, z, w for $\mathbf{P}(W) \cong \mathbf{P}^3$ can be computed as follows.

Taking long exact sequences of cohomology shows that $H^1(Y, \mathcal{O}_Y)$ is isomorphic to $H^3(\mathbf{P}^3, \mathcal{O}_{\mathbf{P}^3}(-5))$, and under this isomorphism B corresponds to the basis

$$B' = \{x^{-2}y^{-1}z^{-1}w^{-1}, \ldots, x^{-1}y^{-1}z^{-1}w^{-2}\}$$

for the cohomology of projective space. The Frobenius map induces the map on $H^3(\mathbf{P}^3, \mathcal{O}_{\mathbf{P}^3}(-5))$ which takes $b \in B'$ to the class of $b^p(gh)^{p-1}$, where in this example $p = 3$. Recall that monomials of the form $x^a y^b z^c w^d$ are zero in $H^3(\mathbf{P}^3, \mathcal{O}_{\mathbf{P}^3}(-5))$ whenever any of a, b, c, d are nonnegative.

Using the basis B', one computes the matrix A of Frobenius on $H^1(Y, \mathcal{O}_Y)$ as

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Again, a computer-aided computation shows that none of the solutions to $Av^{(p)} = v$ correspond to points on the curve. ■

Chapter 5

Additional Techniques

5.1 Purpose

In this chapter we present additional methods for studying torsion points on curves. The purpose of this is two-fold: first, we expect that these techniques may be of some general use in trying to explicitly find torsion points on curves. In addition, the results of this chapter prove Theorem 2.9, which was used in our first proof of Theorem 2.15. Specifically, in this chapter we determine the full cuspidal torsion packet $T(N)$ on the trigonal modular curves $X_0(N)$ with $N = 29, 31, 43, 53, 61$. In each case we show that $T(N)$ is indeed the conjectured set $S(N)$, namely $\{0, \infty\}$ if $N = 43, 53, 61$ and $\{0, \infty\} \cup$ {hyperelliptic branch points} if $N = 29, 31$. This determination is independent of our second proof out the CKR conjecture.

The trigonal curve $X_0(37)$ is dealt with in [12], and a determination of $T(23)$ is made in [13] by methods similar to those used here for $N = 29$ and $N = 31$.

5.2 Useful Facts

For each prime p, fix a prime of $\overline{\mathbb{Q}}$ lying over p, and a corresponding decomposition group D_p of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

We will have use for the following two facts in the arguments that follow. The first we learned in conversation with Ken Ribet.

Lemma 5.1. *Suppose N does not divide the discriminant of the Hecke algebra for $X_0(N)$,*

and that P is a torsion point on $J_0(N)$ of order divisible by N. Then P has at least N^2-1 Galois conjugates.

PROOF. Let $\mathbf{T}_N := \mathbf{T} \otimes \mathbb{Z}_N$, and $\tilde{\mathbf{T}}_N := \mathbf{T}_N/N\mathbf{T}_N$. Without loss of generality we may replace P by a multiple Q of P having order N, and we choose a basis for the N-torsion of $J_0(N)$ as a 2-dimensional $\mathbf{T}[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-module.

By [52, Proposition 6.3], the image of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ in $\mathrm{GL}(2, \tilde{\mathbf{T}}_N)$ equals $\{M \in \mathrm{GL}(2, \tilde{\mathbf{T}}_N) | \det(M) \in (\mathbb{Z}/N\mathbb{Z})^*\}$. It is then easy to see that the lemma is true if Q is of the form $(u, 0)$ with respect to our given basis.

So it suffices to show that any vector (u, v) with coefficients in $\tilde{\mathbf{T}}_N$ can be transformed into something with second component 0 by a matrix in the image of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Since any nonzero (u, v) is a multiple of a vector with one of u, v a unit in $\tilde{\mathbf{T}}_N$, we may assume the latter property holds.

Now if u is a unit, then $\left(\begin{smallmatrix} 1 & 0 \\ -vu^{-1} & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = \left(\begin{smallmatrix} u \\ 0 \end{smallmatrix}\right)$.

And if v is a unit, then we have $\left(\begin{smallmatrix} 0 & 1 \\ 1 & -uv^{-1} \end{smallmatrix}\right)\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) = \left(\begin{smallmatrix} v \\ 0 \end{smallmatrix}\right)$. ■

For a proof of the following theorem, see [23, Proposition 12.1].

Theorem 5.2. *Suppose $f = \sum a_n q^n$ is a normalized weight 2 cusp form for $\Gamma_0(N)$ with coefficients in a field E of characteristic $p \neq N$. Assume that $a_p \neq 0$, and let D_p be a decomposition group at p in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Then the restriction of the 2-dimensional Galois representation ρ_f associated to f to D_p has the form (after picking an appropriate basis)*

$$\begin{pmatrix} \chi\lambda(\frac{1}{a_p}) & * \\ 0 & \lambda(a_p) \end{pmatrix}.$$

Here χ is the mod p cyclotomic character and $\lambda(x)$ is the unramified character which takes a Frobenius element at p to x.

Remark. If $p \neq 2$ and p does not divide the discriminant of the Hecke algebra $\mathbf{T}$, then the representation ρ is the direct sum of the representations ρ_{f_i} above, where the f_i are representatives for each conjugacy class of weight two cusp forms for $\Gamma_0(N)$.

5.3 The Curve $X_0(29)$

The curve $X := X_0(29)$ is hyperelliptic of genus 2. Embed X in its Jacobian via the map i sending x to the class of $(x) - (\infty)$, and let Q be a point of X such that

$P = i(Q)$ is a torsion point in this embedding. We can decompose P into a sum of l-primary components for various primes l,

$$P = \sum P_l,$$

where the P_l are points in the Jacobian of l-power order.

Proposition 5.3. *Suppose that $P_3 = 0$. Then we have $T(29) = S(29)$.*

PROOF. Suppose $Q \in T(29), Q \notin S(29)$. By Theorem 2.2, we have $P = P_2 + P_3 + P_{29} + P_C$, with P_C in the cuspidal group, and by hypothesis $P_3 = 0$. Also, since the characteristic polynomial of T_2 has discriminant 8, all primes other than 2 (and in particular 29) are unramified in $\mathbf{T}$. So Theorem 2.5 shows that $P_2 = 0$. Thus $P = P_{29} + P_C$.

We can now see that $P_{29} = 0$ as follows. By [52, Theorem 6.4], there is a $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which acts on P_{29} as the homothety -6, since $6^2 - 1$ is prime to 29. This σ fixes the rational point P_C of order dividing 7, hence acts as multiplication by -6 on P itself. By the spirit-of-Lang estimate [12, Lemma 4.1], P can have at most $g \cdot 6^2 = 72$ conjugates. But an element of order divisible by 29 in $\mathrm{Jac}(X)$ has at least $29^2 - 1$ conjugates by Lemma 5.1, so $P_{29} = 0$. Thus $P \in C$. By Proposition 2.1, it now follows that Q is one of the cusps, contradicting $Q \notin S(29)$. ■

Theorem 5.4. $T(29) = S(29)$.

PROOF. Suppose $P \in T(29), P \notin S(29)$. By Proposition 5.3, we may assume that $P_3 \neq 0$. Then P_3 is in fact ramified at 3 by Theorem 2.3, since the order $n = 7$ of the cuspidal subgroup is not divisible by 3.

To obtain a contradiction, we use explicit data pertaining to the curve X. First, we need some data concerning the Hecke algebra $\mathbf{T}$, which we gather using the tables of [60]. $\mathbf{T}$ is isomorphic to $\mathbb{Z}[\sqrt{2}]$, so the prime 3 is inert; also, the characteristic polynomial of T_3 is $x^2 - 2x - 1$, which shows that X mod 3 is ordinary (since ordinariness is equivalent to the constant term of this characteristic polynomial being a 3-adic unit).

In addition, we use an explicit equation for X, provided in [20] as

$$y^2 = x^6 - 4x^5 - 12x^4 + 2x^3 + 8x^2 + 8x - 7.$$

Let $f(x)$ denote the right-hand side of this equation and $\tilde{f}(x)$ its reduction mod 3. Since $\tilde{f}$ is irreducible over $\mathbf{F}_3$, this equation provides a model for X over $\mathbb{Z}_3$. By Theorem 4.5, the

ramified torsion point P on the ordinary curve X must reduce to a Cartier point mod 3. By Proposition 4.22, the x-coordinate of the reduction of P must then be a root of $f'(x)$ mod 3. One sees easily that $\tilde{f}'(x) = x^4 + x + 2$ is *irreducible* over $\mathbf{F}_3$. If α is a root of this polynomial defined over $\mathbf{F}_{81}$, then $f(\alpha)$ is a square in $\mathbf{F}_{81}$ (A MAPLE computation shows that its norm to $\mathbf{F}_3$ is a square, which is equivalent). This shows that all torsion points on X ramified at 3 have their reduction defined over $\mathbf{F}_{81}$.

Now according to Theorem 5.2, the action of a decomposition group at 3 on 3-torsion points of $\mathrm{Jac}(X)$ is given matricially by

$$\begin{pmatrix} \chi\lambda(\frac{1}{a_3}) & * \\ 0 & \lambda(a_3) \end{pmatrix}.$$

Here χ is the mod 3 cyclotomic character and $\lambda(x)$ is the unramified character which takes a Frobenius element at 3 to x. The order of λ is thus the order of x in the multiplicative group of $\overline{\mathbf{F}}_3$. In this case, a_3 is a root of $\mathrm{char}(T_3) = x^2 - 2x - 1$, so has order 8. The action of Galois on the 3-torsion in the kernel of reduction is given by the character $\chi\lambda(\frac{1}{a_3})$, and the action of Galois on the reduction of the 3-torsion by the character $\lambda(a_3)$.

Now we may assume that P_3 is not in the kernel of reduction. This is because some conjugate of P_3 under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ must lie outside the kernel of reduction or else the representation $J[3]$ would not be irreducible. This representation *is* irreducible since n is not divisible by 3, i.e., 3 is not an Eisenstein prime.

An alternative argument shows in fact that P_3 cannot be in the kernel of reduction at all. First, we note that $P_N = 0$. For otherwise some multiple P' of P having order N has at least $N^2 - 1$ global Galois conjugates by Lemma 5.1. These conjugates all reduce to different points mod 3, since reduction is injective on prime-to-3 torsion. Thus $\tilde{P}$ itself would have at least $N^2 - 1 = 840$ conjugates, which contradicts the fact that $\tilde{P}$ is defined over F_{81}. So $P = P_3 + P_C$, and if P_3 were in the kernel of reduction, it would have a multiple of 8 conjugates under a decomposition group at 3, and hence $P = P_3 + P_C$ would have a multiple of 8 conjugates in a single residue class (namely that of P_C). This violates Corollary 3.7 of [12].

At any rate, without loss of generality we now suppose that 3 divides the order of the reduction of P_3; $\tilde{P}_3$ thus has a multiple of 8 conjugates, i.e. it is minimally defined over an extension of $\mathbf{F}_3$ of degree a multiple of 8. But then so is the reduction of P, a contradiction ($\tilde{P}$ is defined over $\mathbf{F}_{81}$).

This proves that $P_3 = 0$ as desired. ■

5.4 The Curve $X_0(31)$

One would hope to treat the curve $X_0(31)$ in a similar manner as we did $X_0(29)$ in the previous section. Indeed, most of the arguments go through again.

The Hecke algebra $\mathbf{T} = \mathbb{Z}[\frac{1+\sqrt{5}}{2}]$ is ramified only at 5, and the order of the cuspidal group is also 5. The reasoning used to prove Proposition 5.3 shows that it suffices to prove $P_3 = 0$ for an arbitrary torsion point Q on the curve.

We have an equation

$$y^2 = g(x) := x^6 - 8x^5 + 6x^4 + 18x^3 - 11x^2 - 14x - 3$$

for $X := X_0(31)$. This model has good reduction mod 3, and $g'(x)$ mod 3 is congruent to $2x^4 + 2x + 1$, which is irreducible. But a root β of this irreducible polynomial has $g(\beta)$ *not* a square in $\mathbf{F}_{81}$, so we can say only that the reduction of P is defined over an extension of $\mathbf{F}_3$ of degree 8, not 4. The contradiction in the previous argument for $X_0(29)$ vanishes due to this mercurial change! (Otherwise, life would be easy: 3 is still inert in $\mathbf{T}$, and the characteristic polynomial of T_3 is $x^2 + 2x - 4$, so not only is X mod 3 ordinary, but a_3 still has order 8 in the multiplicative group). So another argument is needed to show that P_3 is zero.

Theorem 5.5. $T(31) = S(31)$.

PROOF. Let Q be a torsion point on $X_0(31)$ ramified at 3, and let $P = i(Q)$. We know we can write $P = P_3 + P_{31} + P_C$ by Theorems 2.2 and 2.5. (Note that the discriminant of the Hecke algebra is prime to 31). Here P_C is a point in the cuspidal subgroup, which has order 5. Also, Q must be a Cartier point mod 3, and we have seen that this implies that the reduction of Q mod 3 is defined over $\mathbf{F}_{3^8}$.

Using the characteristic polynomial of T_3, we can compute the zeta function of X mod 3, and in particular $|J_0(31)(\mathbf{F}_{3^8})|$. Using, for example, [45], we see that the zeta function of $X/\mathbf{F}_3$ is

$$\frac{T^4 + 2T^3 + 2T^2 + 6T + 9}{(1-T)(1-3T)},$$

and then one sees from standard properties of the zeta function that

$$|J_0(31)(\mathbf{F}_{3^8})| = \prod_{\zeta} f(\zeta),$$

where f is the numerator of the zeta function over $\mathbf{F}_3$ and ζ runs through all 8th roots of unity. Indeed, such a formula holds whenever 8 is replaced by a power of 2; e.g. $|J_0(31)(\mathbf{F}_3)| = f(1)$.

In particular, we find that $|J_0(31)(\mathbf{F}_{3^8})| = 2^{12} \cdot 3^2 \cdot 5^2 \cdot 7^2$. The mod 3 reduction of P thus has order prime to 31, and since the reduction map is injective on prime-to-3 torsion, we find that in fact $P_{31} = 0$. The order of P mod 3, call it m, is then $3^a \cdot 5^b$, where $b = 0$ or 1, and $a = 1$ or 2. (If $a = 0$ then P_3 is in the kernel of reduction, but as with $X_0(29)$ we may assume that this is not the case.)

We now claim that $a = 1$. For since 3 is inert in the Hecke algebra, $\mathbf{T}/3\mathbf{T} \cong \mathbf{F}_9$ and consequently $J_0(31)[3]$ is an $\mathbf{F}_9$-vector space; it is then easy to see that $a = 2$ is impossible.

Thus $m = 3$ or 15, so $15\tilde{P} = 0$. What we do now is the following: we have an explicit hyperelliptic model for X (in which the cusps corresponds to the two points at infinity) and we know the x-coordinate of $\tilde{P}$ explicitly, so we should be able to compute the order of $\tilde{P}$ in the Jacobian of X. In particular, if $15\tilde{P} \neq 0$ then we get the desired contradiction.

We have done exactly this computation, and in fact $15\tilde{P} \neq 0$ on the Jacobian of X mod 3. This gives us $P_3 = 0$, as we set out to prove. ■

For lack of a suitable reference, we briefly sketch our method for carrying out the above computation that $15\tilde{P} \neq 0$.

5.5 A Computational Method

The general setup for this section is as follows: Let C be a genus 2 hyperelliptic curve given by the equation $y^2 = g(x)$, with g of degree 6 having coefficients in a field of characteristic $p \neq 2$. [The restriction to genus 2 is merely for the sake of concreteness in notation, it is not essential to our method.] For simplicity, we assume that $g(x)$ is monic, so we can write

$$g(x) = x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0.$$

Label the two (non-Weierstrass) points at infinity on this curve ∞, ∞'. Let P be a non-Weierstrass point on the affine part of C with x-coordinate x_0, and let $m \geq 3$ be an integer. We want to know if $mP = 0$ in $\mathrm{Jac}(C)$ with respect to an Albanese embedding of C into $\mathrm{Jac}(C)$ via one of the points at infinity.

To do this, we find an explicit basis for the vector space $L(m\infty)$, which has dimension $m-1$ by Riemann–Roch (and our assumption $m \geq 3$). Now a basis for the $2m-1$-dimensional space $L(m\infty + m\infty')$ is easy to find, using the fact that $(x) = (0) + (0') - (\infty) - (\infty')$ and $(y) = \sum_{i=1}^{6}(P_i) - 3(\infty) - 3(\infty')$. Such a basis is

$$\{1, x, x^2, \ldots, x^m, y, xy, x^2y, \ldots, x^{m-3}y\}.$$

To distinguish between the two points at infinity comes down to choosing a branch of the square root function, in the following sense. Since $g(x)$ is monic, we can write $g(x) = x^6(1 + G(1/x))$, with G a polynomial having no constant term. Using the binomial theorem, we can then expand y as a Laurent series in $1/x$, obtaining

$$\begin{aligned} y &= \pm x^3(1 + \frac{1}{2}G(\frac{1}{x}) - \frac{1}{8}G(\frac{1}{x})^2 + \cdots) \\ &= \pm(g_1(x) + \sum_{i=1}^{\infty} b_i x^{-i}) \end{aligned}$$

Here $g_1(x)$ is a monic polynomial of degree 3 in x and $b_i \in k$. Note that $g_1(x)$ is the polar part of y as an expansion in $\frac{1}{x}$, which is a local parameter at ∞ (and also at ∞'). Without loss of generality, suppose that $y - g_1(x)$ (as opposed to $y + g_1(x)$) is holomorphic (i.e. regular) at ∞'. This is equivalent to choosing a branch of the square root. If for $2 \leq n \leq m-2$ we let

$$g_n(x) = \text{polar part of } x^{n-1}(g_1(x) + \sum_{i=0}^{\infty} b_i x^{-i})$$

then the function $x^{n-1}y - g_n(x)$ is also regular at ∞', and the set

$$\{1, y - g_1(x), xy - g_2(x), x^2y - g_3(x), \ldots, x^{m-3}y - g_{m-2}(x)\}$$

is easily seen to form a basis for $L(m\infty)$.

We are trying to find out if there exists a function $f \in L(m\infty)$ which vanishes to order m at our point P. Since by assumption P is not a Weierstrass point, $w = x - x_0$ is a local parameter at P. Changing variables, let $Y(w) = +\sqrt{g(w + x_0)}$ be a power series

in w representing the expansion of the regular function y around the parameter w. Note that this involves taking a square root of $g(x_0)$ in k (and $g(x_0) \neq 0$ by hypothesis). Let $h_0(w) = 1$, and for $1 \leq n \leq m-2$ let

$$h_n(w) = (w + x_0)^{n-1} Y(w) - g_n(w + x_0)$$

be power series in w representing the expansions of our basis elements for $L(m\infty)$ around w.

Without loss of generality, we are trying to see if there exists a nontrivial linear combination

$$\sum_{i=0}^{m-2} c_i h_i(w) = w^m + d_{m+1} w^{m+1} + \cdots$$

as power series in w, where the d_n for $n \geq m+1$ are in k. This is simply a linear algebra problem: writing

$$h_i(w) = \sum_{j=0}^{m-1} a_{ij} w^j$$

for $0 \leq i \leq m-2$, we want to determine whether or not the $m \times (m-1)$ matrix (a_{ji}) has a nontrivial kernel.

Even when k is a finite field (over which the curve C and also P are defined), this entire computation can be effectively carried out on a computer. For example, we implemented this algorithm using the software package GAP. In the case where C is the hyperelliptic model

$$y^2 = g(x) = x^6 + x^5 + x^2 + x$$

for $X_0(31)$ in characteristic 3 (see the previous section) and P is a point on C with x-coordinate x_0 equal to a root of $g'(x)$, the program assures us that the divisors $15(P) - 15(\infty)$ and $15(P) - 15(\infty')$ are not principal. As explained in the last section, this proves that $T(31) = S(31)$.

5.6 The Curves $X_0(43)$ and $X_0(61)$

In this section we use the computations from Example 4.27 to determine the cuspidal torsion packet on $X_0(43)$ and $X_0(61)$.

Theorem 5.6. *The Coleman–Kaskel–Ribet conjecture (Conjecture 1.2) is true for $X_0(43)$ and $X_0(61)$.*

PROOF. Let X be either $X_0(43)$ or $X_0(61)$, let J be its Jacobian, let g be the genus of X, and suppose that Q is a point of X such that $P := i_\infty(Q)$ is a torsion point of J. We want to prove that P is in the cuspidal subgroup C of J, for then we will be done by Proposition 2.1. Now by Theorem 2.8(1), X is not hyperelliptic. It follows from Theorem 2.5 that $P_2 \in C$. Moreover, it follows from Theorem 4.5, together with Example 4.27, that P is unramified at 3. By Theorem 2.3, this implies that $P_3 \in C$ as well. Since X has ordinary reduction mod l for all $5 \leq l \leq 2g$ by [60], it now follows from Theorem 2.2 that $P = P_N + P_C$, where P_N has N-power order (N being the unique prime of bad reduction for X), and $P_C \in C$. In fact, P_N must be zero. For suppose not; on the one hand, it follows from Lemma 5.1 that P has at least $N^2 - 1$ Galois conjugates. On the other hand, if we let $n = \mathrm{Num}\frac{N-1}{12}$, it follows from [52, Theorem 6.4] that there exists an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acting on P_N (and therefore on all of P, since the order of P_C divides n) as the homothety $1-n$. It then follows from [12, Lemma 4.1] that P has at most $g(n-1)^2 < N^2 - 1$ conjugates, a contradiction. So $P_N = 0$, and therefore $P \in C$ and we are done. ■

5.7 The Curve $X_0(53)$

The technique we use to study $X_0(53)$ is quite different from those employed thus far. Since $X_0(53)$ does not have ordinary reduction mod 3, we cannot use Theorem 4.5. Instead, we make some technical refinements of the proof of [12, Proposition 3.12] and exploit the existence of an elliptic curve defined over $\mathbb{Q}$ with conductor 53 and supersingular reduction mod 3. It should be noted that this method makes no use of an explicit equation for the modular curve $X_0(53)$.

Let $Q \in T(53)$ be a torsion point on $X = X_0(53)$. The discriminant of the characteristic polynomial of T_2 is $2^4 \cdot 37$; in particular, the Hecke algebra $\mathbf{T}$ is unramified at 3,5, and 53. By Theorems 2.2 and 2.5, we have $P = P_3 + P_{53} + P_C$, where P_C has order dividing 13. Also, the same trick as before shows that to prove $T(53) = S(53)$, it is enough to prove that $P_3 = 0$. For in that case there exists an element $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which acts as the homothety -12 on P, so P has at most $12^2 \cdot g = 576$ conjugates. But if $P_{53} \neq 0$, then P_{53} (and hence P) must have at least $53^2 - 1 > 576$ conjugates.

It remains to prove that $P_3 = 0$. Suppose not, so that (since $n = 13$ is not divisible

by 3) Q is a torsion point ramified at 3.

Exploiting the tables of [60], one sees for $p = 3$ (using the notation of [12]) that dim $W_1 = 3$ and dim $W_2 = 1$. This means that $W := H^0(X, \Omega_{X/\mathbb{Z}_3})$ has a 3-dimensional ordinary part and a 1-dimensional non-ordinary part. Since 3 is unramified in $\mathbf{T}$ (or simply because dim $W_2 = 1$ and $\mathcal{C}$ is nilpotent on $\tilde{W}_2$), H_2 (and by abuse of notation W_2) is in fact superspecial. (See [12, Proposition 3.10])

Pick $\omega_0 \in W_2$ (it is unique up to a constant). The following proposition will be important for our upcoming analysis.

Proposition 5.7. *The differential $\omega_0 \in W_2$ has only simple zeros mod 3. These zeros occur exactly at the branch points of the modular parametrization map from X to the unique rational elliptic curve of conductor 53.*

PROOF. According to the tables of [60], the Jacobian $J_0(53)$ is isogenous to a product $A \times E$ of a simple 3-dimensional abelian variety A with ordinary reduction mod 3 and an elliptic curve E with supersingular reduction mod 3. By [17], the degree d of the modular parametrization map $\phi : X_0(53) \longrightarrow E$ is 2. Up to constants, ω_0 is the pullback $\phi^*\nu$ of an invariant differential on E. Since ν has no zeros mod 3, the zeros of ω_0 mod 3 occur exactly at the ramification points of ϕ, and (since the ramification must be tame given $d < 3$) have multiplicity $d - 1 = 1$. ■

Setting $k_\omega(x) = \text{ord}_{\tilde{x}}(\tilde{\omega}) + 1$ for $\omega \in W$, we see that $k_{\omega_0}(x) \leq 2$ for all $x \in X$, where the tilde denotes reduction mod 3.

Let Q' be a point unramified at 3 in the same mod 3 residue class as Q, and let T be a parameter at Q' which gives an analytic isomorphism, defined over $\mathbb{Q}_p^{\text{nr}}$, from $U(Q')$ (the residue class of Q') to the open unit ball in $\mathbb{C}_p$. Let $s = v(T(P))$, where v is the normalized valuation of $\mathbb{C}_p$ taking p to 1.

By Riemann–Roch, $\tilde{Q}$ is not a base point for one of $\tilde{W}_1, \tilde{W}_2$. Our argument for proving that $P_3 = 0$ will fall into two cases accordingly.

Case 1: $\tilde{Q}$ is not a base point for W_2, i.e. $k_{\omega_0}(Q) = 1$.

By [12, Lemma 3.11], $M_P(H_2) = 0$. So by [12, Proposition 3.6],

$$s = \frac{1}{p^{2n}(p^2 - 1)} = \frac{1}{8 \cdot 3^{2n}} \tag{5.1}$$

for some integer $n \geq 0$.

Now suppose, momentarily, that $k_\nu(Q) \geq p = 3$ for all $\nu \in W_1$. Since it is ordinary, $\tilde{W}_1$ has a basis of differentials fixed by the Cartier operator $\mathcal{C}$. For these differentials, $k_\nu(Q) \geq 3$ implies that in fact $k_\nu(Q) \geq 4$ by [12, Lemma 11]. Thus in fact $k_\nu(Q) \geq 4$ for all $\nu \in W_1$. As no differential in $\tilde{W}_2$ vanishes at $\tilde{Q}$, we see that every differential of $\tilde{W}$ that vanishes at $\tilde{Q}$ vanishes to order at least 3. This is impossible by Riemann–Roch. (See [9, Proof of Theorem 5.5(a)]).

So some element ν of W_1 has $k_\nu(Q) \leq p-1 = 2$, i.e. $k_Q(H_1) \leq p-1$ in the notation of [12]. By [12, Proposition 3.5], then, we see that s is either of the form $\frac{r}{t}$ with $t \leq 2$, $\frac{1}{kp^n} = \frac{1}{k \cdot 3^n}$, or $\frac{1}{kp^n(p-1)} = \frac{1}{2k \cdot 3^n}$ with $n \geq 0$ and k equal to 1 or 2. But these possibilities are all clearly incompatible with (5.1), so this case does not occur.

Case 2: $\tilde{Q}$ is not a base point for W_1, so $k_Q(H_1) = 1$ and (by Proposition 5.7 and Case 1) $k_Q(H_2) = 2$.

Then by [12, Proposition 3.5,3.6] applied to H_1, we have that s is either of the form $\frac{1}{(p-1)p^n} = \frac{1}{2 \cdot 3^n}$ or $\frac{1}{p^M} = \frac{1}{3^M}$ with $n \geq 0$ and $M > 0$.

Applying the same proposion to H_2 (and noting again that $M_Q(H_2) = 0$), we see that s must be of the form $\frac{1}{k(p^2-1)p^n} = \frac{1}{16 \cdot 3^n}$ with $n \geq 0$ or $\frac{r}{t}$ with $t \leq k = 2$. The only conceivable overlap with the results derived from analyzing H_1 is that $s = \frac{1}{2}$.

So we will have shown that $P_3 = 0$, and thus that $T(53) = S(53)$, if we can manage to rule out $s = \frac{1}{2}$. We remark that our analysis shows, using [12, Lemma 14], that there are at most two torsion points ramified at 3 in any single residue class of X. Since there are at most $2g - 2 = 6$ residue classes where $\tilde{\omega_0}(\tilde{Q}) = 0$, there can be at most 12 torsion points on X ramified at 3. This shows in particular that P_3 ramified at 3 implies $P_{53} = 0$, since otherwise by Lemma 5.1, P would have at least $53^2 - 1$ Galois conjugates, all ramified at 3.

Now we know that $J_0(53)$ is isogenous to $A \times E$ with A ordinary mod 3 and E supersingular. Write the image of P_3 in $A \times E$ componentwise as (P_A, P_E). Now since E is a supersingular elliptic curve and P_E a point of 3-power order on E, the residue class mod 3 of all conjugates of P_E is the same (namely 0). Now according to [54, Proposition 2], the image of an inertia group at 3 in $\mathrm{Aut}(E[3])$ is a non-split Cartan subgroup, hence cyclic of order 8. So if $P_A = 0$, then $P = P_E + P_C$ has at least 8 distinct conjugates in the same residue class. This is impossible by our above remark.

Now suppose $P_A \neq 0$. Then we can exploit our knowledge of the Galois representation on 3-torsion of the ordinary part of $J_0(53)$. As the characteristic polynomial of

T_3 acting on the ordinary cusp forms is $x^3 - 3x^2 - x + 1$, a MAPLE computation shows (using the shape of the Galois representation in the ordinary case, see Theorem 5.2) that a nontrivial 3-torsion point in the ordinary part of $J_0(53)$ is either in the kernel of reduction and has at least 26 conjugates in one residue class, or else its reduction has at least 26 conjugates. So our point P, or its reduction mod 3, must have at least 26 Galois conjugates, which contradicts our observation that there are at most 12 torsion points on X ramified at 3.

Thus $P_A = P_E = 0$, i.e., P_3 is in the kernel of the isogeny $f : J_0(53) \to A \times E$. But the degree of this isogeny can only be divisible by primes which divide the discriminant of the Hecke algebra. In particular, the degree of f is prime to 3. So P_3, which is of 3-power order, cannot be in the kernel of f unless it is zero.

We have thus proved the following theorem:

Theorem 5.8. $T(53) = S(53)$.

Chapter 6

Appendix

In this appendix, we give proofs of some results concerning torsion packets on curves. Some of these results seem to be known to the experts, but to our knowledge the proofs we present have not appeared in the literature.

Let us briefly set up some notation and recall a few definitions. Let X be an algebraic curve defined over a field K of characteristic zero. As usual, when we say that X is a curve we require it to be smooth, proper, and geometrically irreducible over K. We will denote by J the Jacobian variety of X, and by g the genus of X.

Let $\bar{K}$ be an algebraic closure of K. We define an equivalence relation on $X(\bar{K})$ by defining $P \sim Q$ iff the divisor $m(P) - m(Q)$ on X is principal for some positive integer m. We call an equivalence class under $\sim$ a *torsion packet* on X. Clearly the torsion packet containing $P \in X(\bar{K})$ is nothing more than the set of points of $X(\bar{K})$ which map to torsion points of J via the Albanese map from X to J sending $Q \in X(\bar{K})$ to the class of the divisor $(Q) - (P)$. A torsion packet is said to be *trivial* if it has only one element.

Recall that the Manin–Mumford conjecture (proved by M. Raynaud in [49]) states that if $g \geq 2$, then every torsion packet on X is finite. In fact, Raynaud [50] proved the following more general result about torsion points on subvarieties of abelian varieties:

Theorem 6.1. *Suppose A is an abelian variety over the algebraically closed field L, and that V is a geometrically irreducible closed K-subvariety of A. If the torsion points of A are Zariski-dense in V, then X is the translate, by a torsion point, of an abelian subvariety of A.*

Remark. Raynaud's theorem is equivalent to the following assertion: With hypotheses as in Theorem 6.1, the set of torsion points of A lying on X is contained in a finite union $\cup Z_j$, where each $Z_j \subseteq V$ and is a translate of a (possibly zero-dimensional) abelian subvariety of A by a torsion point.

We show how Raynaud's theorem can be used to prove some interesting results about the size of torsion packets on a curve X.

First of all, we have the following result, which says that if $g \geq 2$ then there is a uniform bound on the size of torsion packets on X. This result is stated without proof in [59]. The result seems to be *nearly* proved in the literature — Raynaud [50, Proposition 9.1] gives a proof when $g > 2$ — but for completeness we provide a proof due to Bjorn Poonen.

Theorem 6.2. *Suppose $g \geq 2$. Then there exists a constant M depending only on the curve X such that every torsion packet on X has size at most M.*

PROOF. Assume that there are torsion packets on X of arbitrarily large size. Fix a positive integer $n \geq 3$, and consider the map $\alpha_n : X^n \to J^{n-1}$ given by sending $(P_1, \ldots, P_n)$ to $([(P_1) - (P_2)], [(P_2) - (P_3)], \ldots, [(P_{n-1}) - (P_n)])$. The map α_n is just the composition of the Albanese map $X^n \to J^n$ given by $(P_1, \ldots, P_n) \mapsto ([(P_1) - (P_0)], [(P_2) - (P_0)], \ldots, [(P_n) - (P_0)])$ for some fixed point P_0 on X with the map $J^n \twoheadrightarrow J^{n-1}$ obtained by quotienting J^n by J, embedded diagonally. Since the image of X^n under the Albanese map generates J^n as a group, the image of X^n under α_n generates J^{n-1}.

So X^n cannot be the translate by a torsion point of an abelian subvariety of J^{n-1}, because the former variety has dimension n while the latter has dimension $(n-1)g > n$. By Theorem 6.1, it is therefore enough to show that the torsion points of J^{n-1} are Zariski-dense in X^n.

Notice that a point $(P_1, \ldots, P_n)$ on X^n maps to a torsion point on J^{n-1} iff $P_1, \ldots, P_n$ all lie in a common torsion packet.

For each positive integer i, let T_i be the set of n-tuples $(P_1, \ldots, P_n)$ such that the P_i all lie in a common torsion packet having at least N elements.

The Zariski closures of the sets T_i in X^n form a descending chain

$$\bar{T}_1 \supseteq \bar{T}_2 \supseteq \cdots \supseteq \bar{T}_i \supseteq \cdots$$

and since X^n is Noetherian this chain stabilizes, starting with say T_M, so that

$$\bar{T}_M = \bar{T}_{M+1} = \cdots$$

Let V be the Zariski closure of T_M in X^n. We claim that $V = X^n$. As previously noted, this claim implies the theorem. In any case, our initial assumption shows that V has dimension at least one.

To prove the claim, suppose that V is contained in but not equal to X^n. Let $0 < k < n$ be its dimension. A consideration of tangent spaces shows that there is some coordinate projection $X^n \to X^k$ so that the induced map π from V to X^k is dominant. To see this, choose a nonsingular point P of V; by linear algebra there is some projection $\pi_0 : X^n \to X^k$ such that the corresponding map $\pi : V \to X^k$ has $\pi_* : T_P(V) \to T_{\pi(P)}(X^k)$ an isomorphism. The map π_* is then an isomorphism on tangent spaces in a whole open neighborhood U_0 of P. If V' denotes the scheme-theoretic image of V in X^k under π, it follows that there is some nonsingular point Q in V' whose tangent space has dimension at least k. Therefore $V' = X^k$, i.e., π is dominant.

It follows that π is generically finite of degree d. This means that there exists a Zariski open set $U \subseteq X^k$ such that the induced map $\pi : U' := \pi^{-1}(U) \to U$ has finite fibers all of size d.

Since T_i is dense in V for $i \geq M$, for each i there is some element P^i of T_i in U'. Also we must have $P' \in U'$ for all P' such that $\pi(P') = \pi(P^i)$, since $U' = \pi^{-1}(U)$. So if $i > k + d$, then there are at least $d + 1$ elements in $T_i \cap U'$ lying in the fiber of π over $\pi(P^i)$. This contradiction establishes the claim, and hence the theorem. ■

Notice that Theorem 6.2 does not directly address the issue of how many nontrivial torsion packets a curve can have. With this question the picture is slightly more complicated. For example, since the subtraction map $\phi : X \times X \to J$ given by $(P, Q) \mapsto [(P) - (Q)]$ is surjective when $g = 2$, it follows easily that every curve of genus 2 has infinitely many nontrivial torsion packets.

In [11, Example (iv)], one finds that statement that it follows from [50] that for $g \geq 3$, there are only finitely many nontrivial torsion packets. This conclusion is false, as B. Poonen pointed out: we claim that if X has genus three and is both hyperelliptic and bielliptic (for example, X could be the projective curve corresponding to the equation $y^2 = x^8 + 1$), then there are infinitely many nontrivial torsion packets on X.

To see this, let h be the hyperelliptic involution on X, and choose a Weierstrass point S on X, and note that $2(S)$ is linearly equivalent to $(P) + (hP)$ for all points P on X. Let $f : X \to E$ be a degree 2 map. Then the image of the homomorphism $E \to J$ given

by $P \mapsto [f^*(P) - 2(S)]$ is contained in the image of ϕ, because if $f^*(P) = (Q) + (Q')$, then $[f^*(P) - 2(S)] = [(Q) - (hQ')]$. The claim follows.

In fact, we can show that the above example is the *only* way a curve of genus greater than 2 can have infinitely many nontrivial torsion packets.

Theorem 6.3. *Let X be a curve of genus $g \geq 2$. Then X has infinitely many nontrivial torsion packets if and only if either $g = 2$, or $g = 3$ and X is both hyperelliptic and bielliptic.*

PROOF. We have already established the "if" implication, so we need to prove the "only if" assertion.

We continue to denote by ϕ the subtraction morphism $X \times X \to J$. Suppose X is not hyperelliptic; then ϕ is birational to its image W. Indeed, it is easy to see that W is two-dimensional, so that ϕ is generically finite; on the other hand, the generic degree of ϕ is clearly 1, so ϕ must be a birational isomorphism.

More precisely, we claim that ϕ induces an isomorphism between the open subvarieties $U = X \times X - \Delta$ and $U' = W - 0$, where $\Delta \subset X \times X$ is the diagonal. To see this, we use the fact (see [4, Exer. VI.A-1]) that if $K = \bar{K}$ and $\psi : X \to \mathbf{P}(\Omega)$ denotes the canonical embedding, with Ω the vector space of regular differentials on X, then the projectivized image of the differential map

$$\phi_* : T_{(P,Q)}(X \times X) \to T_{[(P)-(Q)]}J \xrightarrow{\sim} \Omega^*$$

is the linear subspace of $\mathbf{P}(\Omega)$ spanned by $\psi(P)$ and $\psi(Q)$. (Here $\mathbf{P}(\Omega)$ denotes the space of hyperplanes in Ω, which is isomorphic to the space of lines in Ω^*.)

Using this description, we see easily that ϕ_* is an isomorphism on tangent spaces when restricted to U. Since $\phi : U \to U'$ is clearly bijective when X is not hyperelliptic, our claim follows from Lemma 14.8 and Theorem 14.9 of [25] ("inverse function theorem" for varieties).

Now suppose that W contains infinitely many torsion points of J. If $g > 2$, then W is not an abelian surface, because it generates J, which has dimension at least 3, as an abelian group. (This follows from the stronger fact that any Albanese embedding of X generates J). Therefore, by Raynaud's theorem, W must contain the translate of an elliptic curve E by a torsion point of J. Since W contains the origin of J, we can in fact assume that E is an abelian subvariety of J.

If we assume that X is not hyperelliptic, then there is a birational map $E \hookrightarrow W \to X^2 \twoheadrightarrow X$, where the map $X^2 \twoheadrightarrow X$ is projection onto the second factor. Since E and X are nonsingular curves, this map extends to a morphism $E \to X$. Our assumption that X has infinitely many nontrivial torsion packets implies that this morphism is not constant, which contradicts the fact that $g > 1$.

Therefore X is hyperelliptic with involution h. If S is a hyperelliptic branch point on X, then the composite map

$$X^2 \to X^2 \to X^{(2)} \to J$$

given by

$$(P, Q) \mapsto (P, hQ) \mapsto P + hQ \mapsto [(P) + (hQ) - 2(S)]$$

is just the difference map ϕ. Therefore E is contained in the image of the symmetric square map $X^{(2)} \to J$. It follows from [2, Theorem 3] that X is bielliptic. An application of the Correspondence Lemma (Lemma 3.1) shows in this case that $g \leq 3$. ■

Remark. Suppose Γ is the division group of a finitely generated subgroup of $J(\bar{K})$. We say that P and Q in $J(\bar{K})$ are in the same Γ-packet if $[(P) - (Q)] \in \Gamma$. Using a theorem of Raynaud–Faltings et. al. which generalizes Theorem 6.1 (see [48, Theorem 3] for a statement), we can replace the phrase "torsion packet" in Theorems 6.2 and 6.3 by "Γ-packet" and the resulting statements are still true.

We conclude with an intriguing open problem concerning torsion packets on curves, a conjecture of Coleman first put forth in his paper [10].

Conjecture 6.4. *Suppose K is a number field, and that X is a curve of genus $g \geq 2$ defined over K. Let $\mathfrak{p}$ be a prime of K of characteristic p not dividing 6 for which $K/\mathbb{Q}$ is unramified and X has good reduction. Let T be a torsion packet in $X(\bar{K})$ stable under* $\mathrm{Gal}(\bar{K}/K)$. *Then the extension $K(T)/K$ is unramified above $\mathfrak{p}$.*

A good discussion of what is known about this conjecture can be found in [11]. For example, Coleman proved in [10] that the conjecture is true for $p > 2g$, and he proved it for $p < 2g$ as well when the reduction of X at $\mathfrak{p}$ is either ordinary or superspecial. Another bit of evidence for the conjecture is given by [11, Proposition 5].

Without the condition that $K/\mathbb{Q}$ is unramified at $\mathfrak{p}$, the conjecture is false, as is explained in example (2) at the end of [10].

We now give the details of an example, also due to Coleman, which shows that the other hypotheses of the conjecture are necessary as well. Specifically, we show that the cuspidal torsion packet on $X_1(13)$ (which has at least 22 elements) is stable under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and is ramified above the primes 2, 3, and 13.

For a number of details concerning the structure of $X_1(13)$, we refer the reader to the excellent paper [41].

$X_1(13)$ (which we refer to from now on as X) is a curve of genus 2 defined over $\mathbb{Q}$ which has good reduction everywhere except 13. It has 12 cusps, 6 of them rational over $\mathbb{Q}$ and the other 6 rational over K, the maximal totally real subfield of $\mathbb{Q}(\mu_{13})$. The group Δ, which is dihedral of order 12, acts on X as a group of automorphisms. Δ consists of the six diamond bracket operators $\langle d\rangle$ with $d \in (\mathbb{Z}/13\mathbb{Z})^*/\pm 1$ and the six Atkin–Lehner involutions w_ζ for $\zeta \in \mu_{13}$, where $w_\zeta = w_{\zeta^{-1}}$.

Explicitly, if (E, P) corresponds to a "point" of X, where E is an elliptic curve and P a point of order 13 on E, then $\langle d\rangle : (E, P) \mapsto (E, dP)$ and $w_\zeta : (E, P) \mapsto (E', \phi(Q))$, where $\phi : E \to E'$ is the isogeny with cyclic kernel generated by P and Q is a point on E whose Weil pairing with P is ζ.

It is not hard to see that the hyperelliptic involution on X coincides with the automorphism $\langle 5\rangle$. Indeed, it follows from [41, p. 44] (where $\langle d\rangle$ is called γ_d) that the involution $\langle 5\rangle = \langle 2\rangle^3$ acts on $J = J_1(13)$ as -1, so it must be the hyperelliptic involution. Similarly, the quotient of X by the cubic automorphism $\langle 3\rangle$ has genus zero. One way to see this is that by the Riemann–Hurwitz formula, $\langle 3\rangle$ either has four fixed points and $X/\langle 3\rangle$ has genus zero, or else $\langle 3\rangle$ has 1 fixed point and $X/\langle 3\rangle$ has genus one. The latter is impossible, as can be seen by applying Riemann–Hurwitz to the degree 6 Galois covering $X \to X/\langle 2\rangle$. (Alternatively, we will construct momentarily 4 fixed points of $\langle 3\rangle$). A similar application of Riemann–Hurwitz shows that the 6 hyperelliptic branch points P_i and the four fixed points Q_i of $\langle 3\rangle$ are all distinct.

The 6 rational cusps generate a subgroup of $J(\mathbb{Q})$ of order 19 (in fact, this is the full Mordell-Weil group of J). Since Δ acts freely on the set of cusps, all 12 cusps C_i lie in a common torsion packet, which we refer to as the cuspidal torsion packet T. We also see from the freeness of this action that the 22 points C_i, P_i, Q_i are all distinct. In addition,

they all lie in the torsion packet T. This follows directly from the very useful Lemma 6.1 of [9]. Clearly T is stable under $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$.

It remains to show that the torsion packet T is ramified above the primes 2,3, and 13. It is ramified at 13, for example, because six of the cusps are defined only over K, which is totally ramified at 13. We can see that T is ramified at 2 from the fact that the divisor classes $[(P_i) - (P_j)]$ generate the entire group $J[2]$ of 2-torsion points on J, which is certainly ramified at 2.

It suffices, then, to show that the points Q_i are ramified at 3. To see this, we identify these points quite explicitly. Let $E/\mathbb{Q}$ have complex multiplication by $\mathbb{Z}[\omega]$, where ω is a cube root of unity (so that $j(E) = 0$). The prime ideal 13 splits in $\mathbb{Z}[\omega]$ as a product $(13) = (\pi)(\pi')$ of principal ideals. Since $(\omega - 3)(\omega - 9) \equiv 0 \pmod{\pi}$, we have that either $\omega \equiv 3$ or $\omega \equiv 9 \pmod{\pi}$. Without loss of generality, let us assume that $\omega \equiv 3 \pmod{\pi}$. Then if we choose a point $P \in E[\pi]$, it follows that $[\omega]$ induces an isomorphism between the pairs (E, P) and $(E, 3P)$. So (E, P) corresponds to a fixed point of $\langle 3 \rangle$. (Conversely, it is not hard to see that all four fixed points of $\langle 3 \rangle$ arise in this way, either from $P \in E[\pi]$ or $P \in E[\pi']$.)

We claim that a field of definition for a pair (E, P) as above must contain $\mathbb{Q}(\sqrt{-3})$, hence is ramified at 3. In other words, we want to know that every element of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which fixes P also fixes ω. This is clear, because if $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ fixes P but not ω, then $\tau\pi = \pi'$ and $\tau P = P$. This implies that $P \in E[\pi]$ and $\tau P = P \in E[\pi']$, but the intersection of $E[\pi]$ and $E[\pi']$ is trivial. This proves the claim and finishes the proof that T is ramified at 2, 3, and 13.

Bibliography

[1] D. Abramovich, *A linear lower bound on the gonality of modular curves*, IMRN **20** (1996), 1005–1011.

[2] D. Abramovich and J. Harris, *Abelian varieties and curves in $W_d(C)$*, Comp. Math. **78** (1991), 227–238.

[3] R. Accola, *On Castelnuovo's inequality for algebraic curves. I*, Trans. AMS. **251** (1979), 357–373.

[4] E. Arbarello, M. Cornalba, P. Griffiths, and J. Harris, *Geometry of Algebraic Curves*, Grundlehren der mathematischen Wissenschaften, vol. 267, Springer-Verlag, Berlin and New York, 1985.

[5] P. Bayer and J. González, *On the Hasse-Witt invariants of modular curves*, Experimental Mathematics **6** (1997), 57–76.

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

[7] A. Buium, *Geometry of p-jets*, Duke Math J. **82** (1996), 349–367.

[8] B. Chisala, *Canonical coordinates for hyperelliptic curves in characteristic p*, Ph.D. Dissertation, University of California, Berkeley, 1986.

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

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

[11] R.F. Coleman, *Torsion points on curves*, in Galois Representations and Arithmetic Algebraic Geometry, Advanced Studies in Pure Math. 12 (1987), 235–247.

[12] R.F. Coleman, B. Kaskel, and K. Ribet, *Torsion points on* $X_0(N)$. To appear in Proceedings of a Symposia in Pure Mathematics.

[13] R.F. Coleman, Letter to M. Baker, June 1997.

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

[15] M. Coppens, *Some sufficient conditions for the gonality of a smooth curve*, Journal of Pure and Applied Algebra **30** (1983), 5–21.

[16] M. Coppens, *On G. Martens' characterization of smooth plane curves*, Bull. London Math. Soc. **20** (1988), 217–220.

[17] J. Cremona, *Computing the degree of the modular parametrization of a modular elliptic curve*, Mathematics of Computation **64** (1995), 1235–1250.

[18] T. Ekedahl, *On supersingular curves and Abelian varieties*, Math. Scand. **60** (1987), 151–178.

[19] 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.

[20] S. Galbraith, *Equations for modular curves*, Ph.D. thesis, Oxford University (1996).

[21] M. Green and R. Lazarsfeld, *On the projective normality of complete linear series on an algebraic curve*, Invent. Math. **83** (1985), no. 1, 73–90.

[22] P. Griffiths and J. Harris, *On the variety of special linear systems on a general algebraic curve*, Duke Math. J. **47** (1980), 233–272.

[23] B. Gross, *A tameness criterion for Galois representations associated to modular forms (mod p)*, Duke Math J. **61** (1990), 445–517.

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

[25] J. Harris, *Algebraic Geometry*, Graduate Texts in Mathematics, vol. 133, Springer-Verlag, Berlin and New York, 1992.

[26] J. Harris and I. Morrison, *Moduli of Curves*, Graduate Texts in Mathematics, vol. 187, Springer-Verlag, Berlin and New York, 1998.

[27] J. Harris and J. Silverman, *Bielliptic curves and symmetric products*, Proceedings of the AMS **112** (1991), 347–356.

[28] R. Hartshorne, *Algebraic Geometry*, Graduate Texts in Mathematics, vol. 52, Springer-Verlag, Berlin and New York, 1977.

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

[30] Y. Hasegawa and M. Shimura, *Trigonal modular curves*, to appear in Acta Arith. (1999)

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

[32] Y. Hasegawa, Letter to K. Ribet, Aug. 1998.

[33] H. Hasse and E. Witt, *Zyklische unverzweigte Erweiterungskörper vom Primzahlgrade p über einem algebraischen Funktionenkörper der Charakteristik p*, Monatsch. Math. Phys. **43** (1936), 477–492.

[34] D. Kubert and S. Lang, *Modular Units,* Grundlehren der mathematischen Wissenschaften 244, Springer-Verlag, Berlin and New York, 1981.

[35] S. Lang, *Elliptic Functions* (2nd ed.), Graduate Texts in Mathematics, vol. 112, Springer-Verlag, Berlin and New York, 1987.

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

[37] H. W. Lenstra, Jr. and K. Ribet. (In preparation).

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

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

[40] B. Mazur, *Rational isogenies of prime degree*, Invent. Math. **44** (1978), 129–162.

[41] B. Mazur and J. Tate, *Points of order 13 on elliptic curves*, Invent. Math. **22** (1973), 41–49.

[42] R. Miranda, *Algebraic Curves and Riemann Surfaces*, Graduate Studies in Mathematics, vol. 5, American Mathematical Society, Providence, R.I., 1995.

[43] K.V. Nguyen and M. Saito, *D-Gonality of modular curves and bounding torsions*, Preprint. Available on the Algebraic Geometry Web as alg-geom/9603024.

[44] N. Nygaard, *Slopes of powers of Frobenius on crystalline cohomology*, Ann. Sci. Ecole Norm. Sup. **14** (1981), 369–401.

[45] A.P. Ogg, *Hyperelliptic modular curves*, Bull. Soc. Math. France **102** (1974), 449–462.

[46] A.P. Ogg, *Diophantine equations and modular forms*, Bull. AMS **81** (1975), 14–27.

[47] A.P. Ogg, *On the Weierstrass points of* $X_0(N)$, Illinois J. Math. **22** (1978), no. 1, 31–35.

[48] B. Poonen, *Mordell-Lang plus Bogomolov*, to appear in Invent. Math.

[49] M. Raynaud, *Courbes sur une variété Abélienne et points de torsion*, Invent. Math. **71** (1983), 207–233.

[50] 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.

[51] 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 and New York.

[52] K. Ribet, *Images of semistable Galois representations*, Olga Taussky-Todd: in memoriam. Pacific J. Math. **1997**, Special Issue, 277–297.

[53] J.P. Serre, *Abelian* l*-adic Representations and Elliptic Curves*, Research Notes in Mathematics, 7, A. K. Peters, Ltd., Wellesley, MA, 1998.

[54] J.P. Serre, *Propriétés Galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), 259–331.

[55] J.P. Serre, *Algebraic Groups and Class Fields*, Graduate Texts in Mathematics, vol. 117, Springer-Verlag, Berlin and New York, 1975.

[56] J.P. Serre, *Sur la topologie des variétés algébriques en caractéristique p*, Symp. Int. Top. Alg., Mexico (1958), 24–53.

[57] J.P. Serre, *Cohomologie Galoisienne*, Lecture Notes in Mathematics, vol. 5, Springer–Verlag, Berlin and New York, 1994.

[58] J. Silverman, *The Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics, vol. 106, Springer-Verlag, Berlin and New York, 1986.

[59] J. Silverman, *Hecke points on modular curves*, Duke Math. J. **60** (1990), pp. 401–423.

[60] W. Stein, *Modular Forms Database*. Available via the Number Theory Web: http://www.math.uga.edu/˜ntheory/web.html.

[61] J. Sturm, *On the congruence of modular forms*, in Number theory (New York, 1984–1985), 275–280, Lecture Notes in Math., vol. 1240, Springer-Verlag, Berlin and New York, 1987.

[62] M. Tsfasman and S. Vladut, *Algebraic-Geometric Codes*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1991.

[63] R. Valentini, *Hyperelliptic curves with zero Hasse-Witt matrix*, Man. Mathematica **86** (1995), 185–194.

[64] H. Wada, *A table of Hecke operators (II)*, Proc. Japan Acad. **49** (1973), pp. 380–84.

[65] N. Yui, *On the Jacobian varieties of hyperelliptic curves over fields of characteristic* $p > 2$, Journal of Algebra **52** (1978), 378–410.