THE OPEN BOOK SERIES 1

ANTS X
Proceedings of the Tenth Algorithmic Number Theory Symposium

San Diego 2012

edited by
Everett W. Howe and Kiran S. Kedlaya

msp

ANTS X
Proceedings of the Tenth
Algorithmic Number Theory Symposium

THE OPEN BOOK SERIES

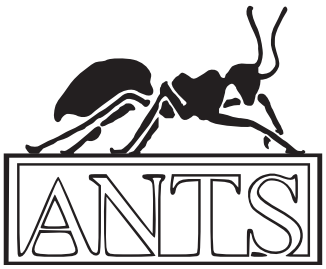
THE OPEN BOOK SERIES 1

ANTS X

Proceedings of the Tenth Algorithmic Number Theory Symposium

San Diego 2012

Edited by
Everett W. Howe and Kiran S. Kedlaya



msp

Everett W. Howe
Center for Communications Research
4320 Westerra Court
San Diego, CA 92121-1969
United States

Kiran S. Kedlaya
Department of Mathematics
University of California, San Diego
9500 Gilman Drive #0112
La Jolla, CA 92093-0112

Front cover artwork based on a detail of
Chicano Legacy 40 Años ©2010 Mario Torero.

Electronic copies can be obtained free of charge from http://msp.org/obs/1 and printed copies can be ordered from MSP (contact@msp.org).

The Open Book Series is a trademark of Mathematical Sciences Publishers.

ISSN: 2329-9061 (print), 2329-907X (electronic)

ISBN: 978-1-935107-00-2 (print), 978-1-935107-01-9 (electronic)

First published 2013.

MATHEMATICAL SCIENCES PUBLISHERS
798 Evans Hall #3840, c/o University of California, Berkeley CA 94720-3840
contact@msp.org **http://msp.org**

Contents

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.7

msp

Preface

The Algorithmic Number Theory Symposium (ANTS), held biennially since 1994, is the premier international forum for research in computational number theory. ANTS is devoted to algorithmic aspects of number theory, including elementary, algebraic, and analytic number theory, the geometry of numbers, arithmetic algebraic geometry, the theory of finite fields, and cryptography.

This volume is the proceedings of the tenth ANTS meeting, held July 9–13, 2012 at the University of California, San Diego. The scientific program of ANTS X consisted of 5 invited lectures, 25 contributed talks, a poster session, and a rump session. The invited speakers were Manjul Bhargava (Princeton University), Nils Bruin (Simon Fraser University), Wen-Ching Winnie Li (Pennsylvania State University), Nils-Peter Skoruppa (Universität Siegen), and Andrew Sutherland (Massachusetts Institute of Technology). Extended abstracts of the presentations of Bruin and Sutherland are included in this volume.

The contributed talks were presentations of papers chosen through a competitive review process. Each of the 55 papers submitted for consideration was reviewed by at least three members of the program committee, often with input from one or more external reviewers as well. Revised and edited versions of the 25 accepted papers are included in this volume.

At each ANTS since 2006, the Number Theory Foundation has sponsored the Selfridge Prize, an award for the best contributed paper, as judged by the program committee. The Selfridge Prize for ANTS X was awarded to Andrew Sutherland for his paper *On the evaluation of modular polynomials*.

Abstracts of all presentations (including invited presentations and posters), PDF slides of many presentations, and the versions of the contributed papers that were presented at the conference can be found on the conference web site:

http://math.ucsd.edu/~kedlaya/ants10/

For each of the previous ANTS conferences, the proceedings volume was produced before the meeting and was available at the meeting. This publication timeline allowed for very little editing and did not permit authors to revise their papers to incorporate insights gained from discussions during the conference. Following a suggestion raised in previous years, the ANTS X organizing committee decided to

produce the proceedings volume *after* the conference. The committee also decided to switch publishers; we are proud to note that this volume is the inaugural volume of the Open Book Series of Mathematical Sciences Publishers.

A word about bibliographic references: The editors tried their best to find online versions of the references that are cited in the ANTS papers. If you are reading a PDF version of one of the papers in this volume, and if one of its references has a title that is colored blue, then the title is a hyperlink to an online copy of the reference. If you are reading a printed copy of an ANTS paper, the hyperlinks will unfortunately no longer work. However, there are still ways to find online versions of cited references. For example, the AMS Digital Mathematics Registry includes a useful list of journal archives:

http://www.ams.org/dmr/JournalList.html

For some of the ANTS references that appear in journals that are not on the AMS list, the editors were nevertheless able to track down online versions. For these references, we spell out the URL of the paper in the bibliographic entry.

The editors are grateful to the authors of the papers in this volume for their flexibility and graciousness during the editing process. The editors are equally grateful to Silvio Levy and Alex Scorpan, our contacts at Mathematical Sciences Publishers, for *their* flexibility and graciousness. We hope that the reader will find the value added by the editing to be sufficient recompense for the extra year's wait for the volume to appear.

Everett Howe and Kiran Kedlaya
San Diego, November 2013

Local organizing committee.

Alina Bucur	University of California, San Diego
Joe Buhler	Center for Communications Research, La Jolla
Dan Gordon	Center for Communications Research, La Jolla
Everett Howe	Center for Communications Research, La Jolla
Kiran Kedlaya	University of California, San Diego
Kristin Lauter	Microsoft Research

Program committee.

Dan Bernstein	University of Illinois, Chicago
Alina Bucur	University of California, San Diego
Joe Buhler	Center for Communications Research, La Jolla
Henri Cohen	Université de Bordeaux 1
Chantal David	Concordia University
Steven Galbraith	University of Auckland

Dan Gordon	Center for Communications Research, La Jolla
Everett Howe (cochair)	Center for Communications Research, La Jolla
Kiran Kedlaya (cochair)	University of California, San Diego
Jürgen Klüners	Universität Paderborn
Kristin Lauter	Microsoft Research
Fernando Rodriguez Villegas	University of Texas, Austin
Peter Stevenhagen	Universiteit Leiden
Michael Stoll	Universität Bayreuth
Bianca Viray	Brown University

Financial sponsors. We are thankful for support from Microsoft Research, the National Science Foundation, the National Security Agency, the Number Theory Foundation, and the University of California, San Diego.

Previous and future ANTS meetings.

	Year	Location	Proceedings
I	1994	Cornell University (Ithaca, NY, USA)	LNCS 877
II	1996	Université Bordeaux 1 (Talence, France)	LNCS 1122
III	1998	Reed College (Portland, OR, USA)	LNCS 1423
IV	2000	Universiteit Leiden (The Netherlands)	LNCS 1838
V	2002	University of Sydney (Australia)	LNCS 2369
VI	2004	University of Vermont (Burlington, VT, USA)	LNCS 3076
VII	2006	Technische Universität Berlin (Germany)	LNCS 4076
VIII	2008	Banff Centre (Banff, Alberta, Canada)	LNCS 5011
IX	2010	INRIA (Nancy, France)	LNCS 6197
X	2012	University of California (San Diego, CA, USA)	OBS 1
XI	2014	Hotel Hyundai (GyeongJu, Korea)	

Proceedings of the previous ANTS meetings have been published in the Springer Lecture Notes in Computer Science series (LNCS). This volume is the first volume of the Mathematical Sciences Publishers' Open Book Series (OBS).

ANTS XI is planned to be held August 7–11, 2014 in GyeongJu, Korea, as a satellite conference to the International Congress of Mathematicians. The chairs of the program committee are Jung Hee Cheon (Seoul National University) and Hyang-Sook Lee (Ewha Womans University).

EVERETT W. HOWE: however@alumni.caltech.edu
Center for Communications Research, 4320 Westerra Court, San Diego, CA 92121-1969, United States

KIRAN S. KEDLAYA: kedlaya@ucsd.edu
Department of Mathematics, University of California, San Diego, 9500 Gilman Drive #0112, La Jolla, CA 92093-0112, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.1

msp

Deterministic elliptic curve primality proving for a special sequence of numbers

Alexander Abatzoglou, Alice Silverberg,
Andrew V. Sutherland, and Angela Wong

We give a deterministic algorithm that very quickly proves the primality or compositeness of the integers N in a certain sequence, using an elliptic curve $E/\mathbb{Q}$ with complex multiplication by the ring of integers of $\mathbb{Q}(\sqrt{-7})$. The algorithm uses $O(\log N)$ arithmetic operations in the ring $\mathbb{Z}/N\mathbb{Z}$, implying a bit complexity that is quasiquadratic in $\log N$. Notably, neither of the classical "$N-1$" or "$N+1$" primality tests apply to the integers in our sequence. We discuss how this algorithm may be applied, in combination with sieving techniques, to efficiently search for very large primes. This has allowed us to prove the primality of several integers with more than 100,000 decimal digits, the largest of which has more than a million bits in its binary representation. At the time it was found, it was the largest proven prime N for which no significant partial factorization of $N-1$ or $N+1$ is known (as of final submission it was second largest).

1. Introduction

With the celebrated result of Agrawal, Kayal, and Saxena [3], one can now unequivocally determine the primality or compositeness of any integer in deterministic polynomial time. With the improvements of Lenstra and Pomerance [27], the AKS algorithm runs in $\widetilde{O}(n^6)$ time, where n is the size of the integer to be tested (in bits). However, it has long been known that for certain special sequences of integers, one can do much better. The two most famous examples are the Fermat numbers $F_k = 2^{2^k}+1$, to which one may apply Pépin's criterion [35], and the Mersenne numbers $M_p = 2^p - 1$, which are subject to the Lucas-Lehmer test [24]. In both cases, the corresponding algorithms are deterministic and run in $\widetilde{O}(n^2)$ time.

MSC2010: primary 11Y11; secondary 11G05, 14K22.
Keywords: primality, elliptic curves, complex multiplication.

In fact, every prime admits a proof of its primality that can be verified by a deterministic algorithm in $\widetilde{O}(n^2)$ time. Pomerance shows in [36] that for every prime $p > 31$ there exists an elliptic curve $E/\mathbb{F}_p$ with an $\mathbb{F}_p$-rational point P of order $2^r > (p^{1/4} + 1)^2$, which allows one to establish the primality of p using just r elliptic curve group operations. Elliptic curves play a key role in Pomerance's proof; the best analogous result using classical primality certificates yields an $\widetilde{O}(n^3)$ time bound (see [38], and compare [9, Theorem 4.1.9]).

The difficulty in applying Pomerance's result lies in finding the pair (E, P), a task for which no efficient method is currently known. Rather than searching for suitable pairs (E, P), we instead fix a finite set of curves $E_a/\mathbb{Q}$, each equipped with a known rational point P_a of infinite order. To each positive integer k we associate one of the curves E_a and define an integer J_k for which we give a necessary and sufficient condition for primality: J_k is prime if and only if the reduction of P_a in $E_a(\mathbb{F}_p)$ has order 2^{k+1} for every prime p dividing J_k. Of course $p = J_k$ when J_k is prime, but this condition can easily be checked without knowing the prime factorization of J_k. This yields a deterministic algorithm that runs in $\widetilde{O}(n^2)$ time (see Algorithm 5.1).

Our results extend the methods used by Gross [20], Denomme and Savin [11], Tsumura [44], and Gurevich and Kunyavskiĭ [22], all of which fit within a general framework laid out by Chudnovsky and Chudnovsky in [8] for determining the primality of integers in special sequences using elliptic curves with complex multiplication (CM). The elliptic curves that we use lie in the family of quadratic twists defined by the equations

$$E_a : y^2 = x^3 - 35a^2x - 98a^3, \tag{1}$$

for squarefree integers a such that $E_a(\mathbb{Q})$ has positive rank. Each curve has good reduction outside of 2, 7, and the prime divisors of a, and has CM by $\mathbb{Z}[\alpha]$, where

$$\alpha = \frac{1 + \sqrt{-7}}{2}.$$

For each curve E_a, we fix a point $P_a \in E_a(\mathbb{Q})$ of infinite order with $P_a \notin 2E_a(\mathbb{Q})$.

For each positive integer k, let

$$j_k = 1 + 2\alpha^k \in \mathbb{Z}[\alpha],$$
$$J_k = j_k\overline{j_k} = 1 + 2(\alpha^k + \bar{\alpha}^k) + 2^{k+2} \in \mathbb{N}.$$

The integer sequence J_k satisfies the linear recurrence relation

$$J_{k+4} = 4J_{k+3} - 7J_{k+2} + 8J_{k+1} - 4J_k,$$

with initial values $J_1 = J_2 = 11$, $J_3 = 23$, and $J_4 = 67$. Then (by Lemma 4.5) J_k is composite for $k \equiv 0 \pmod 8$ and for $k \equiv 6 \pmod{24}$. To each other value of

k we assign a squarefree integer a, based on the congruence class of $k \pmod{72}$, as listed in Table 1. Our choice of a is based on two criteria. First, it ensures that when J_k is prime, the Frobenius endomorphism of $E_a \bmod J_k$ corresponds to complex multiplication by j_k (rather than $-j_k$) and

$$E_a(\mathbb{Z}/J_k\mathbb{Z}) \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{k+1}\mathbb{Z}.$$

Second, it implies that when J_k is prime, the reduction of the point P_a has order 2^{k+1} in $E(\mathbb{Z}/J_k\mathbb{Z})$. The second condition is actually stronger than necessary (in general, one only needs P_a to have order greater than $2^{k/2+1}$), but it simplifies matters. Note that choosing a sequence of the form $j_k = 1 + \Lambda_k$ means that $E_a(\mathbb{Z}[\alpha]/(j_k)) \simeq \mathbb{Z}[\alpha]/\Lambda_k$, whenever J_k is prime and j_k is the Frobenius endomorphism of $E_a \bmod J_k$ (see Lemma 4.6).

We prove in Theorem 4.1 that the integer J_k is prime if and only if the point P_a has order 2^{k+1} on "$E_a \bmod J_k$". More precisely, we prove that if one applies the standard formulas for the elliptic curve group law to compute scalar multiples $Q_i = 2^i P_a$ using projective coordinates $Q_i = [x_i, y_i, z_i]$ in the ring $\mathbb{Z}/J_k\mathbb{Z}$, then J_k is prime if and only if $\gcd(J_k, z_k) = 1$ and $z_{k+1} = 0$. This allows us to determine whether J_k is prime or composite using $O(k)$ operations in the ring $\mathbb{Z}/J_k\mathbb{Z}$, yielding a bit complexity of $O(k^2 \log k \log\log k) = \widetilde{O}(k^2)$ (see Proposition 5.2 for a more precise bound).

We note that, unlike the Fermat numbers, the Mersenne numbers, and many similar numbers of a special form, the integers J_k are not amenable to any of the classical "$N - 1$" or "$N + 1$" type primality tests (or combined tests) that are typically used to find very large primes (indeed, the 500 largest primes currently listed in [7] all have the shape $ab^n \pm 1$ for some small integers a and b).

In combination with a sieving approach described in Section 5, we have used our algorithm to determine the primality of J_k for all $k \le 1.2 \times 10^6$. The prime values of J_k are listed in Table 4. At the time it was found, the prime $J_{1,111,930}$, which has 334,725 decimal digits, was the largest proven prime N for which no significant partial factorization of either $N - 1$ or $N + 1$ was known [1]. On July 4, 2012 it was superseded by a 377,922 digit prime found by David Broadhurst [6] for which no significant factorization of $N - 1$ or $N + 1$ is known; Broadhurst constructed an ECPP primality proof for this prime, but it is not a Pomerance proof.

Generalizations have been suggested to the settings of higher-dimensional abelian varieties with complex multiplication, algebraic tori, and group schemes by Chudnovsky and Chudnovsky [8], Gross [20], and Gurevich and Kunyavskiĭ [21], respectively. In the PhD theses of the first and fourth authors, and in a forthcoming paper, we are extending the results in this paper to a more general framework. In that paper we will also explain why, when restricting to elliptic curves over $\mathbb{Q}$, this method requires curves with CM by $\mathbb{Q}(\sqrt{-D})$ with $D = 1$, 2, 3, or 7.

2. Relation to prior work

In [8], Chudnovsky and Chudnovsky consider certain sequences of integers $s_k = \mathrm{Norm}_{K/\mathbb{Q}}(1 + \alpha_0\alpha_1^k)$, defined by algebraic integers α_0 and α_1 in an imaginary quadratic field $K = \mathbb{Q}(\sqrt{D})$. They give sufficient conditions for the primality of s_k, using an elliptic curve E with CM by K. In our setting, $D = -7$, $\alpha_0 = 2$, $\alpha_1 = (1 + \sqrt{-7})/2$, and $J_k = s_k$. The key difference here is that we give necessary and sufficient criteria for primality that can be efficiently checked by a deterministic algorithm. This is achieved by carefully selecting the curves $E_a/\mathbb{Q}$ that we use, so that in each case we are able to prove that the point $P_a \in E_a(\mathbb{Q})$ reduces to a point of maximal order 2^{k+1} on $E_a \bmod J_k$, whenever J_k is prime. Without such a construction, we know of no way to obtain *any* nontrivial point on $E \bmod s_k$ in deterministic polynomial time.

Our work is a direct extension of the techniques developed by Gross [20; 45], Denomme and Savin [11], Tsumura [44], and Gurevich and Kunyavskiĭ [22], who use elliptic curves with CM by the ring of integers of $\mathbb{Q}(i)$ or $\mathbb{Q}(\sqrt{-3})$ to test the primality of Mersenne, Fermat, and related numbers. However, as noted by Pomerance [37, §4], the integers considered in [11] can be proved prime using classical methods that are more efficient and do not involve elliptic curves, and the same applies to [20; 44; 45; 22]. But this is not the case for the sequence we consider here.

3. Background and notation

3A. ***Elliptic curve primality proving.*** Primality proving algorithms based on elliptic curves have been proposed since the mid-1980s. Bosma [5] and Chudnovsky and Chudnovsky [8] considered a setting similar to the one employed here, using elliptic curves to prove the primality of numbers of a special form; Bosma proposed the use of elliptic curves with complex multiplication by $\mathbb{Q}(i)$ or $\mathbb{Q}(\sqrt{-3})$, while Chudnovsky and Chudnovsky considered a wider range of elliptic curves and other algebraic varieties. Goldwasser and Kilian [16; 17] gave the first general purpose elliptic curve primality proving algorithm, using randomly generated elliptic curves. Atkin and Morain [4; 32] developed an improved version of the Goldwasser-Kilian algorithm that uses the CM method to construct the elliptic curves used, rather than generating them at random (it does rely on probabilistic methods for root-finding). With asymptotic improvements due to Shallit, the Atkin-Morain algorithm has a heuristic expected running time of $\widetilde{O}(n^4)$, which makes it the method of choice for general purpose primality proving [33]. Gordon [18] proposed a general purpose compositeness test using supersingular reductions of CM elliptic curves over $\mathbb{Q}$.

Throughout this paper, if $E \subset \mathbb{P}^2$ is an elliptic curve over $\mathbb{Q}$, we shall write points $[x, y, z] \in E(\mathbb{Q})$ so that $x, y, z \in \mathbb{Z}$ and $\gcd(x, y, z) = 1$, and we may use

(x, y) to denote the projective point $[x, y, 1]$.

We say that a point $P = [x, y, z] \in E(\mathbb{Q})$ is *zero mod* N when N divides z; otherwise P is *nonzero mod* N. Note that if P is zero mod N then P is zero mod p for all primes p dividing N.

Definition 3.1. Given an elliptic curve E over $\mathbb{Q}$, a point $P = [x, y, z] \in E(\mathbb{Q})$, and $N \in \mathbb{Z}$, we say that P is *strongly nonzero mod* N *if* $\gcd(z, N) = 1$.

If P is strongly nonzero mod N, then P is nonzero mod p for every prime p dividing N, and if N is prime, then P is strongly nonzero mod N if and only if P is nonzero mod N.

We rely on this fundamental result, which can be found in [16; 26; 17]:

Proposition 3.2. *Let $E/\mathbb{Q}$ be an elliptic curve, let N be a positive integer prime to* $\operatorname{disc}(E)$, *let $P \in E(\mathbb{Q})$, and let $m > (N^{1/4} + 1)^2$. Suppose mP is zero mod N and $(m/q)P$ is strongly nonzero mod N for all primes $q \mid m$. Then N is prime.*

To make practical use of Proposition 3.2, one needs to know the prime factorization of m. For general elliptic curve primality proving this presents a challenge; the algorithms of Goldwasser-Kilian and Atkin-Morain use different approaches to ensure that m has an easy factorization, but both must then recursively construct primality proofs for the primes q dividing m. In our restricted setting we effectively fix the prime factorization of $m = 2^{k+1}$ ahead of time.

Next we give a variant of Proposition 3.2 that replaces "strongly nonzero" with "nonzero", at the expense of m being a prime power with a larger lower bound.

Proposition 3.3. *Let $E/\mathbb{Q}$ be an elliptic curve, let p be a prime, let N be an odd positive integer prime to* $p \operatorname{disc}(E)$, *and let $P \in E(\mathbb{Q})$. Suppose b is a positive integer such that $p^b > (\sqrt{N/3} + 1)^2$ and $p^b P$ is zero mod N and $p^{b-1} P$ is nonzero mod N. Then N is prime.*

Proof. Since $p^{b-1} P$ is nonzero mod N, there are a prime divisor q of N and a positive integer r such that q^r exactly divides N and $p^{b-1} P$ is nonzero mod q^r. Let $E_1(\mathbb{Z}/q^r\mathbb{Z})$ denote the kernel of the reduction map $E(\mathbb{Z}/q^r\mathbb{Z}) \to E(\mathbb{F}_q)$. It follows, for example, from [29, Theorem 4.1] that $E_1(\mathbb{Z}/q^r\mathbb{Z})$ is a q-group. Let $P' \in E(\mathbb{Z}/q^r\mathbb{Z})$ be the reduction of P mod q^r and let P'' be the image of P' in $E(\mathbb{F}_q)$. If $p^{b-1} P'' = 0$ then $p^{b-1} P' \in E_1(\mathbb{Z}/q^r\mathbb{Z})$, so $p^{b-1} P'$ has order a power of q. But by assumption it has order p, which is prime to N. This is a contradiction, so P'' has order p^b. If N were composite, then $q \le N/3$ since N is odd, so by the Hasse bound,

$$p^b \le |E(\mathbb{F}_q)| \le (\sqrt{q} + 1)^2 \le (\sqrt{N/3} + 1)^2,$$

contradicting the hypothesis that $p^b > (\sqrt{N/3} + 1)^2$. □

3B. ***Complex multiplication and Frobenius endomorphism.*** For any number field F, let $\mathcal{O}_F$ denote its ring of integers. If E is an elliptic curve over a field K, and Ω_K is the space of holomorphic differentials on E over K, then Ω_K is a one-dimensional K-vector space, and there is a canonical ring homomorphism

$$\mathrm{End}_K(E) \to \mathrm{End}_K(\Omega) = K. \tag{2}$$

Suppose now that E is an elliptic curve over an imaginary quadratic field K, and that E has complex multiplication (CM) by $\mathcal{O}_K$, meaning that $\mathrm{End}_K(E) \simeq \mathcal{O}_K$. Then the image of the map in (2) is $\mathcal{O}_K$. Let $\psi : \mathcal{O}_K \to \mathrm{End}_K(E)$ denote the inverse map. Suppose that $\mathfrak{p}$ is a prime ideal of K at which E has good reduction and let $\widetilde{E}$ denote the reduction of E mod $\mathfrak{p}$. Then the composition

$$\mathcal{O}_K \xrightarrow{\sim} \mathrm{End}_K(E) \hookrightarrow \mathrm{End}_{\mathcal{O}_K/\mathfrak{p}}(\widetilde{E}),$$

where the first map is ψ and the second is induced by reduction mod $\mathfrak{p}$, gives a canonical embedding

$$\mathcal{O}_K \hookrightarrow \mathrm{End}(\widetilde{E}). \tag{3}$$

The Frobenius endomorphism of $\widetilde{E}$ is $(x, y) \mapsto (x^q, y^q)$ where $q = \mathrm{Norm}_{K/\mathbb{Q}}(\mathfrak{p})$; under the embedding in (3), the Frobenius endomorphism is the image of a particular generator π of the (principal) ideal $\mathfrak{p}$. By abuse of notation, we say that the Frobenius endomorphism is π.

4. Main theorem

In this section we state and prove our main result, Theorem 4.1, which gives a necessary and sufficient condition for the primality of the numbers J_k.

Fix a particular square root of -7 and let $K = \mathbb{Q}(\sqrt{-7})$. Let

$$\alpha = \frac{1 + \sqrt{-7}}{2} \in \mathcal{O}_K,$$

and for each positive integer k, let

$$j_k = 1 + 2\alpha^k \in \mathbb{Z}[\alpha] \quad \text{and} \quad J_k = \mathrm{Norm}_{K/\mathbb{Q}}(j_k) = j_k \overline{j_k} \in \mathbb{N}.$$

Note that J_k is prime in $\mathbb{Z}$ if and only if j_k is prime in $\mathcal{O}_K$. Note also that $\mathrm{Norm}_{K/\mathbb{Q}}(\alpha) = \alpha\bar{\alpha} = 2$.

Recall the family of elliptic curves E_a defined by (1). Lemma 4.5 below shows that J_k is composite if $k \equiv 0 \pmod 8$ or $k \equiv 6 \pmod{24}$, so we omit these cases from our primality criterion. For each remaining value of k, Table 1 lists the twisting parameter a and the point $P_a \in E_a(\mathbb{Q})$ we associate to k. For each of these a, the elliptic curve E_a has rank one over $\mathbb{Q}$, and the point P_a is a generator for $E_a(\mathbb{Q})$ modulo torsion.

k	a	P_a
$k \equiv 0$ or $2 \pmod 3$	-1	$(1, 8)$
$k \equiv 4, 7, 13, 22 \pmod{24}$	-5	$(15, 50)$
$k \equiv 10 \pmod{24}$	-6	$(21, 63)$
$k \equiv 1, 19, 49, 67 \pmod{72}$	-17	$(81, 440)$
$k \equiv 25, 43 \pmod{72}$	-111	$(-633, 12384)$

Table 1. The twisting parameters a and points P_a.

Theorem 4.1. *Fix $k > 1$ such that $k \not\equiv 0 \pmod 8$ and $k \not\equiv 6 \pmod{24}$. Let $P_a \in E_a(\mathbb{Q})$ be as in Table 1 (depending on k). The following are equivalent*:

(i) *$2^{k+1} P_a$ is zero mod J_k and $2^k P_a$ is strongly nonzero mod J_k;*

(ii) *J_k is prime.*

Remark 4.2. Applying Proposition 3.3 with $N = J_k$, $p = 2$, and $b = k + 1$, we can add an equivalent condition in Theorem 4.1 as long as $k \geq 6$, namely:

(iii) $2^{k+1} P_a$ is zero mod J_k and $2^k P_a$ is nonzero mod J_k.

We shall prove Theorem 4.1 via a series of lemmas, but let us first outline the proof. One direction is easy: Since $2^{k+1} > (J_k^{1/4} + 1)^2$ for all $k > 1$, if (i) holds then so does (ii), by Proposition 3.2 (where the hypothesis $\gcd(J_k, \operatorname{disc}(E_a)) = 1$ holds by Lemma 4.5 below).

Now fix a and P_a as in Table 1, and let $\widetilde{P}_a$ denote the reduction of P_a modulo j_k. We first compute a set S_a such that if $k \in S_a$ and j_k is prime, then $E_a(\mathbb{O}_K/(j_k)) \simeq \mathbb{O}_K/(2\alpha^k)$ as $\mathbb{O}_K$-modules. We then compute a set T_a such that if $k \in T_a$ and j_k is prime, then $\widetilde{P}_a$ does not lie in $\alpha E_a(\mathbb{O}_K/(j_k))$ if and only if $k \in T_a$ (note that $\alpha \in \mathbb{O}_K \hookrightarrow \operatorname{End}(E_a)$). For $k \in S_a \cap T_a$, the point $\widetilde{P}_a$ has order 2^{k+1} whenever J_k is prime.

We now fill in the details. Many of the explicit calculations below were performed with the assistance of the Sage computer algebra system [43].

4A. ***The linear recurrence sequence J_k.*** As noted in the introduction, the sequence J_k satisfies the linear recurrence relation

$$J_{k+4} = 4J_{k+3} - 7J_{k+2} + 8J_{k+1} - 4J_k. \tag{4}$$

We now prove this, and also note some periodic properties of this sequence. See [12] or [28, Chapter 6] for basic properties of linear recurrence sequences.

Definition 4.3. We call a sequence a_k (*purely*) *periodic* if there exists an integer m such that $a_k = a_{k+m}$ for all k. The minimal such m is the *period* of the sequence.

Lemma 4.4. *The sequence J_k satisfies* (4). *If p is an odd prime and $\mathfrak{p} \subset \mathbb{O}_K$ is a prime ideal above (p), then the sequence $J_k \bmod p$ is periodic, with period equal to the least common multiple of the orders of* 2 *and* α *in* $(\mathbb{O}_K/\mathfrak{p})^*$.

Proof. The characteristic polynomial of the linear recurrence in (4) is

$$f(x) = x^4 - 4x^3 + 7x^2 - 8x + 4 = (x-1)(x-2)(x^2 - x + 2),$$

whose roots are $1, 2, \alpha$, and $\bar{\alpha}$. It follows that the sequences 1^k, 2^k, α^k, and $\bar{\alpha}^k$, and any linear combination of these sequences, satisfy (4). Thus J_k satisfies (4).

One easily checks that the lemma is true for $p = 7$, so assume $p \neq 7$. Let A be the 4×4 matrix with $A_{i,j} = J_{i+j-1}$. Then $\det A = -2^{12} \cdot 7$ is nonzero mod p, hence its rows are linearly independent over $\mathbb{F}_p$. It follows from Theorems 6.19 and 6.27 of [28] that the sequence $J_k \bmod p$ is periodic, with period equal to the lcm of the orders of the roots of f in $\overline{\mathbb{F}}_p^*$ (which we note are distinct). These roots all lie in $\mathbb{O}_K/\mathfrak{p} \simeq \mathbb{F}_{p^d}$, where $d \in \{1, 2\}$ is the residue degree of $\mathfrak{p}$. Since $\bar{\alpha} = 2/\alpha$, the order of $\bar{\alpha}$ in $(\mathbb{O}_K/\mathfrak{p})^*$ divides the lcm of the orders of 2 and α. The lemma follows. □

When p is an odd prime, let m_p denote the period of the sequence $J_k \bmod p$. Lemma 4.4 implies that m_p always divides $p^2 - 1$, and it divides $p - 1$ whenever p splits in K.

Lemma 4.5.

(i) J_k *is divisible by* 3 *if and only if* $k \equiv 0 \pmod 8$.

(ii) J_k *is divisible by* 5 *if and only if* $k \equiv 6 \pmod{24}$.

(iii) $J_k \equiv 2 \pmod 7$ *if* $k \equiv 0 \pmod 3$, *and* $J_k \equiv 4 \pmod 7$ *otherwise.*

(iv) *For* $k > 1$, *we have* $J_k \equiv 3 \pmod 8$ *if* k *is even, and* $J_k \equiv 7 \pmod 8$ *if* k *is odd.*

(v) J_k *is divisible by* 17 *if and only if* $k \equiv 54 \pmod{144}$.

(vi) J_k *is not divisible by* 37.

Proof. Lemma 4.4 allows us to compute the periods $m_3 = 8$, $m_5 = 24$, $m_7 = 3$, $m_{17} = 144$, and $m_{37} = 36$. It then suffices to check, for $p = 3, 5, 17$, and 37, when $J_k \equiv 0 \pmod p$ for $1 \leq k \leq m_p$, and to determine the values of $J_k \pmod 7$ for $1 \leq k \leq 3$.

It is easy to check that $\alpha^k + \bar{\alpha}^k \equiv 3 \pmod 4$ for odd $k > 1$, and $\alpha^k + \bar{\alpha}^k \equiv 1 \pmod 4$ otherwise. Since $J_k = 1 + 2(\alpha^k + \bar{\alpha}^k) + 2^{k+2}$, we have (iv).

As an alternative proof for one direction of (i) and (ii), note that α and $\bar{\alpha}$ each has order 8 in $(\mathbb{O}_K/(3))^\times$. Hence if $k \equiv 0 \pmod 8$, then $J_k = 1 + 2(\alpha^k + \bar{\alpha}^k) + 2^{k+2} \equiv 1 + 2(1 + 1) + 1 \equiv 0 \pmod 3$. Similarly, $\alpha^6 \equiv 2 \equiv \bar{\alpha}^6 \pmod 5$, so $J_k \equiv 1 + 2(4) + 1 \equiv 0 \pmod 5$ when $k \equiv 6 \pmod{24}$. □

4B. ***The set S_a.*** For each squarefree integer a we define the set of integers

$$S_a := \left\{ k > 1 : \left(\frac{a}{J_k}\right)\left(\frac{j_k}{\sqrt{-7}}\right) = 1 \right\},$$

where $\left(\text{-}\right)$ denotes the (generalized) Jacobi symbol.

If j_k is prime in $\mathbb{O}_K$, then the Frobenius endomorphism of E_a over the finite field $\mathbb{O}_K/(j_k)$ corresponds to either j_k or $-j_k$. For elliptic curves over $\mathbb{Q}$ with complex multiplication, one can easily determine which is the case.

Lemma 4.6. *Suppose a is a squarefree integer, $k > 1$, and j_k is prime in $\mathbb{O}_K$. Then*:

(i) *$k \in S_a$ if and only if the Frobenius endomorphism of E_a over the finite field $\mathbb{O}_K/(j_k)$ is j_k;*

(ii) *if $k \in S_a$, then $E_a(\mathbb{O}_K/(j_k)) \simeq \mathbb{O}_K/(2\alpha^k)$ as $\mathbb{O}_K$-modules.*

Proof. The elliptic curve E_a is the curve in Theorem 1 of [42, p. 1117], with $D = -7$ and $\pi = j_k$. By [42, p. 1135], the Frobenius endomorphism of E_a over $\mathbb{O}_K/(j_k)$ is

$$\left(\frac{a}{J_k}\right)\left(\frac{j_k}{\sqrt{-7}}\right) j_k \in \mathbb{O}_K.$$

Part (i) then follows from the definition of S_a. For (ii), note that (i) implies that if $k \in S_a$, then

$$E_a(\mathbb{O}_K/(j_k)) \simeq \ker(j_k - 1) = \ker(2\alpha^k) \simeq \mathbb{O}_K/(2\alpha^k),$$

which completes the proof. □

The next lemma follows directly from Lemma 4.5(iv).

Lemma 4.7. *Let $k > 1$.*

$$\text{(i)}\ \left(\frac{-1}{J_k}\right) = -1. \quad \text{(ii)}\ \left(\frac{2}{J_k}\right) = \begin{cases} 1 & \text{if } k \text{ is odd,} \\ -1 & \text{if } k \text{ is even.} \end{cases}$$

We now explicitly compute the sets S_a for the values of a used in Theorem 4.1.

Lemma 4.8. *For $a \in \{-1, -5, -6, -17, -111\}$ the sets S_a are as in Table 2.*

Proof. Since $j_k = 1 + 2\alpha^k$, and $\alpha \equiv 4 \pmod{\sqrt{-7}}$, and $2^3 \equiv 1 \pmod 7$, we have

$$\left(\frac{j_k}{\sqrt{-7}}\right) = \left(\frac{1 + 2^{2k+1}}{7}\right) = \begin{cases} 1 & \text{if } k \equiv 1 \pmod 3, \\ -1 & \text{if } k \equiv 0, 2 \pmod 3. \end{cases}$$

We now need to compute $\left(\frac{a}{J_k}\right)$ for $a = -1, -5, -6, -17$, and -111. The case $a = -1$ is given by Lemma 4.7(i). As in the proof of Lemma 4.5, applying Lemma 4.4 to the odd primes $p = 3, 5, 17, 37$ that can divide a, we found that

a	m	$S_a = \{k > 1 : k \bmod m$ is as below$\}$
-1	3	0, 2
-5	24	0, 2, 4, 5, 7, 9, 12, 13, 16, 18, 21, 22, 23
-6	24	3, 7, 9, 10, 11, 12, 13, 17, 20, 22
-17	144	0, 1, 5, 7, 9, 10, 13, 14, 15, 18, 19, 20, 22, 23, 27, 30, 31, 33, 34, 36, 42, 43, 44, 45, 49, 50, 53, 56, 61, 62, 63, 66, 67, 68, 70, 71, 72, 73, 75, 76, 78, 79, 80, 81, 82, 83, 90, 91, 92, 93, 97, 99, 100, 104, 106, 108, 110, 111, 112, 114, 117, 118, 121, 122, 123, 125, 126, 128, 129, 133, 135, 136, 137, 138, 139, 141, 143
-111	72	2, 4, 6, 9, 14, 15, 18, 20, 22, 23, 25, 30, 33, 34, 35, 37, 38, 39, 41, 42, 43, 47, 49, 50, 52, 53, 54, 55, 57, 58, 63, 65, 66, 67, 68, 70

Table 2. The sets S_a.

the periods m_p of the sequences $J_k \bmod p$ are $m_3 = 8$, $m_5 = 24$, $m_{17} = 144$, and $m_{37} = 36$. Since $\left(\frac{-1}{J_k}\right) = -1$, it follows from quadratic reciprocity that for $a = -5, -17$, and -111, the period of the sequence $\left(\frac{a}{J_k}\right)$ divides the least common multiple of the periods m_p for p dividing a. For $a = -6$, by Lemma 4.7(ii) the period of $\left(\frac{2}{J_k}\right)$ is 2, which already divides $m_3 = 8$. Since the period of the sequence $\left(\frac{j_k}{\sqrt{-7}}\right)$ is 3, we find the period m of $\left(\frac{a}{J_k}\right)\left(\frac{j_k}{\sqrt{-7}}\right)$ listed in Table 2 by taking the least common multiple of 3 and the m_p for p dividing a. To compute S_a, it then suffices to compute $\left(\frac{a}{J_k}\right)$ and check when $\left(\frac{a}{J_k}\right) = \left(\frac{j_k}{\sqrt{-7}}\right)$, for $1 < k \le m + 1$. □

4C. ***The set T_a.*** We now define the sets T_a.

Definition 4.9. Let a be a squarefree integer, and suppose that $P \in E_a(K)$. Then the field $K(\alpha^{-1}(P))$ has degree 1 or 2 over K, so it can be written in the form $K(\sqrt{\delta_P})$ with $\delta_P \in K$. Let

$$T_P := \left\{ k > 1 : \left(\frac{\delta_P}{j_k}\right) = -1 \right\}.$$

For the values of a listed in Table 1, let $T_a = T_{P_a}$ and let $\delta_a = \delta_{P_a}$.

Lemma 4.10. *Suppose that $k > 1$, j_k is prime in $\mathbb{O}_K$, and a is a squarefree integer. Suppose that $P \in E_a(K)$, and let $\widetilde{P}$ denote the reduction of P mod j_k. Then $\widetilde{P} \notin \alpha E_a(\mathbb{O}_K/(j_k))$ if and only if $k \in T_P$.*

Proof. Let $L = K(\alpha^{-1}(P)) = K(\gamma)$ for some $\gamma \in L$ such that $\gamma^2 = \delta_P$. Fix a $Q \in E_a(\overline{\mathbb{Q}})$ such that $\alpha Q = P$. Since $\ker(\alpha) \subset E_a[2] \subset E_a(K)$, we have $K(Q) = L = K(\gamma)$. Fix a prime ideal $\mathfrak{p}$ of L above (j_k), let $\mathbb{F} = \mathbb{O}_K/(j_k)$, let

$\widetilde{Q} \in E_a(\overline{\mathbb{F}})$ be the reduction of Q mod $\mathfrak{p}$, and let $\widetilde{\gamma}$ be the reduction of γ mod $\mathfrak{p}$. Then $\mathbb{F}(\widetilde{Q}) = \mathbb{F}(\widetilde{\gamma})$.

Now $\widetilde{P} \in \alpha E_a(\mathbb{F})$ if and only if $\widetilde{Q} \in E_a(\mathbb{F})$. By the above, this happens if and only if $\widetilde{\gamma} \in \mathbb{F}$, that is, if and only if δ_P is a square modulo j_k. □

Lemma 4.11. *We can take*

$$\delta_{-1} = \alpha, \quad \delta_{-5} = -5\alpha, \quad \delta_{-6} = -3\sqrt{-7}, \quad \delta_{-17} = \alpha, \quad \delta_{-111} = -3.$$

Proof. The action of the endomorphism α on the elliptic curve E_a and its reductions is as follows (see Proposition II.2.3.1 of [41, p. 111]). For $(x, y) \in E_a$, we have

$$\alpha(x, y) = \left(\frac{2x^2 + a(7-s)x + a^2(-7-21s)}{(-3+s)x + a(-7+5s)}, \frac{y\left(2x^2 + a(14-2s)x + a^2(28+14s)\right)}{-(5+s)x^2 - a(42+2s)x - a^2(77-7s)} \right),$$

where $s = \sqrt{-7}$. Solving for R in $\alpha R = P_a$ yields δ_a in each case. □

Lemma 4.12. *If $k > 1$ then $\left(\frac{\alpha}{j_k}\right) = -1$.*

Proof. Let $M = K\left(\sqrt{\alpha}\right)$. By the reciprocity law of global class field theory we have

$$\prod_{\mathfrak{p}} (j_k, M_{\mathfrak{p}}/K_{\mathfrak{p}}) = 1,$$

where $(j_k, M_{\mathfrak{p}}/K_{\mathfrak{p}})$ is the norm residue symbol.

Let $f(x) = x^2 - j_k \in \mathbb{O}_{K_\alpha}[x]$. For $k > 1$ we have

$$|f(1)|_\alpha = \left|2\alpha^k\right|_\alpha = 2^{-(k+1)} < 2^{-2} = |4|_\alpha = \left|f'(1)^2\right|_\alpha,$$

and Hensel's lemma implies that $f(x)$ has a root in $\mathbb{O}_{K_\alpha}$. Thus j_k is a square in K_α and $(j_k, M_\alpha/K_\alpha) = 1$.

Identify $K_{\bar{\alpha}}$ with $\mathbb{Q}_2$. Applying Theorem 1 of [40, p. 20] with $a = j_k$ and $b = \alpha$, and using $\bar{\alpha}^5 = 5 + \alpha$, gives $(j_k, \alpha) = -1$, where (j_k, α) is the Hilbert symbol. Thus $j_k \notin \mathrm{Norm}_{M_{\bar{\alpha}}/K_{\bar{\alpha}}}(M_{\bar{\alpha}}^*)$, and therefore $(j_k, M_{\bar{\alpha}}/K_{\bar{\alpha}}) = -1$.

If $\mathfrak{p}$ is a prime ideal of $\mathbb{O}_K$ that does not divide 2, then $M_{\mathfrak{p}}/K_{\mathfrak{p}}$ is unramified. By local class field theory we then have

$$(j_k, M_{\mathfrak{p}}/K_{\mathfrak{p}}) = \left(\frac{\alpha}{\mathfrak{p}}\right)^{\mathrm{ord}_{\mathfrak{p}}(j_k)}.$$

Since j_k is prime to 2, we have $\mathrm{ord}_\alpha(j_k) = \mathrm{ord}_{\bar{\alpha}}(j_k) = 0$, hence

$$\prod_{\mathfrak{p}\nmid 2} (j_k, M_{\mathfrak{p}}/K_{\mathfrak{p}}) = \prod_{\mathfrak{p}\nmid 2} \left(\frac{\alpha}{\mathfrak{p}}\right)^{\mathrm{ord}_{\mathfrak{p}}(j_k)} = \prod_{\text{all } \mathfrak{p}} \left(\frac{\alpha}{\mathfrak{p}}\right)^{\mathrm{ord}_{\mathfrak{p}}(j_k)} = \left(\frac{\alpha}{j_k}\right).$$

Therefore

$$1 = \prod_{\mathfrak{p}} (j_k, M_{\mathfrak{p}}/K_{\mathfrak{p}}) = \Big(\frac{\alpha}{j_k}\Big)(j_k, M_\alpha/K_\alpha)(j_k, M_{\overline{\alpha}}/K_{\overline{\alpha}}) = -\Big(\frac{\alpha}{j_k}\Big),$$

as desired. □

Lemma 4.13. *For $a \in \{-1, -5, -6, -17, -111\}$ the sets T_a are as follows*:

$$\begin{aligned}
T_{-1} &= \{k > 1\},\\
T_{-5} &= \{k > 1 : k \equiv 3, 4, 7, 8, 11, 13, 14, 15, 16, 17, 20, 22 \pmod{24}\},\\
T_{-6} &= \{k > 1 : k \equiv 1, 5, 10, 12, 15, 19, 20, 21, 22, 23 \pmod{24}\},\\
T_{-17} &= \{k > 1\},\\
T_{-111} &= \{k > 1 : k \equiv 1, 2, 3, 6 \pmod{8}\}.
\end{aligned}$$

Proof. We apply Lemma 4.11 and the definition of T_a. Lemma 4.12 implies that $T_{-1} = T_{-17} = \{k > 1\}$. For $a = -6$ we use quadratic reciprocity in quadratic fields (see Theorem 8.15 of [25, p. 257]) to compute $\big(\frac{\sqrt{-7}}{j_k}\big)$. For the remaining cases we compute $\big(\frac{-3}{j_k}\big) = \big(\frac{-3}{J_k}\big)$ and $\big(\frac{-5}{j_k}\big) = \big(\frac{-5}{J_k}\big)$ as in the proof of Lemma 4.8, and apply $\big(\frac{\alpha}{j_k}\big) = -1$ from Lemma 4.12. □

4D. *Proof of Theorem 4.1.*

Lemma 4.14. *Let a be a squarefree integer. Suppose that $P \in E_a(K)$, $k \in S_a \cap T_P$, and j_k is prime. Let $\widetilde{P}$ denote the reduction of P mod j_k. Then the annihilator of $\widetilde{P}$ in $\mathbb{O}_K$ is divisible by α^{k+1}.*

Proof. We have $E_a(\mathbb{O}_K/(j_k)) \simeq \mathbb{O}_K/(2\alpha^k) = \mathbb{O}_K/(\overline{\alpha}\alpha^{k+1})$, by Lemma 4.6(ii). It then suffices to show $\widetilde{P} \notin \alpha E_a(\mathbb{O}_K/(j_k))$, which follows from Lemma 4.10. □

The congruence conditions for k in Table 1 come from taking $S_a \cap T_a$, excluding the cases handled by Lemma 4.5, and adjusting to give disjoint sets.

We now prove Theorem 4.1. Suppose that $k > 1$, $k \not\equiv 0 \pmod 8$, $k \not\equiv 6 \pmod{24}$, and J_k is prime. Let a and P_a be as listed in Table 1. Then $k \in S_a \cap T_a$. Let $\widetilde{P}$ denote the reduction of P_a mod j_k. We have $E_a(\mathbb{O}_K/(j_k)) \simeq \mathbb{O}_K/(2\alpha^k)$ by Lemma 4.6(ii), and therefore the annihilator of $\widetilde{P}$ in $\mathbb{O}_K$ divides $2\alpha^k$. By Lemma 4.14, the annihilator of $\widetilde{P}$ in $\mathbb{O}_K$ is divisible by α^{k+1}. Since $2\alpha^k$ divides 2^{k+1} but α^{k+1} does not divide 2^k, we must have $2^{k+1}\widetilde{P} = 0$ and $2^k\widetilde{P} \neq 0$. Therefore $2^{k+1}P_a$ is zero mod J_k and $2^k P_a$ is strongly nonzero mod J_k.

For the converse, note that $\operatorname{disc}(E_a) = -2^{12} \cdot 7^3 \cdot a^6$, so Lemma 4.5 shows that $\gcd(J_k, \operatorname{disc}(E_a)) = 1$ if $k \not\equiv 0 \pmod 8$ and $k \not\equiv 6 \pmod{24}$. We can therefore apply Proposition 3.2 with $m = 2^{k+1}$, noting that

$$2^{k+1} > ((3 \cdot 2^{k+1})^{\frac{1}{4}} + 1)^2 > (J_k^{1/4} + 1)^2$$

for all $k > 2$, and for $k = 2$ we have $2^{k+1} = 8 > (11^{1/4} + 1)^2 = (J_k^{1/4} + 1)^2$. This proves Theorem 4.1.

Remark. As pointed out by Richard Pinch, $P_a \in 2E_a(\mathbb{O}_K/(j_k))$ if and only if all $x(P_a) - e_i$ are squares mod j_k, where E_a is $y^2 = \prod_{i=1}^{3}(x - e_i)$ and $x(P_a)$ is the x-coordinate. We tested for divisibility by α instead of by 2, to make it clearer how this approach (as initiated by Gross in [20]) makes use of the $\mathbb{O}_K$-module structure of $E_a(\mathbb{O}_K/(j_k))$. Such an approach is useful for further generalizations.

5. Algorithm

A naïve implementation of Theorem 4.1 is entirely straightforward, but here we describe a particularly efficient implementation and analyze its complexity. We then discuss how the algorithm may be used in combination with sieving to search for prime values of J_k, and give some computational results.

5A. ***Implementation.*** There are two features of the primality criterion given by Theorem 4.1 worth noting. First, it is only necessary to perform the operation of adding a point on the elliptic curve to itself (doubling), no general additions are required. Second, testing whether a projective point $P = [x, y, z]$ is zero or strongly nonzero modulo an integer J_k only involves the z-coordinate: P is zero mod J_k if and only if $J_k \mid z$, and P is strongly nonzero mod J_k if and only if $\gcd(z, J_k) = 1$.

To reduce the cost of doubling, we transform the curve

$$E_a: \quad y^2 = x^3 - 35a^2x - 98a^3$$

to the Montgomery form [31]

$$E_{A,B}: \quad By^2 = x^3 + Ax^2 + x.$$

Such a transformation is not possible over $\mathbb{Q}$, but it can be done over $\mathbb{Q}(\sqrt{-7})$. In general, one transforms a short Weierstrass equation $y^2 = f(x) = x^3 + a_4x + a_6$ into Montgomery form by choosing a root γ of $f(x)$ and setting $B = (3\gamma^2 - a_4)^{-1/2}$ and $A = 3\gamma B$; see, for example, [34]. For the curve E_a, we choose $\gamma = \frac{1}{2}(-7 + \sqrt{-7})a$, yielding

$$A = \frac{-15 - 3\sqrt{-7}}{8} \quad \text{and} \quad B = \frac{7 + 3\sqrt{-7}}{56a}.$$

With this transformation, the point $P_a = (x_0, y_0)$ on E_a corresponds to the point $(B(x_0 - \gamma), By_0)$ on the Montgomery curve $E_{A,B}$, and is defined over $\mathbb{Q}(\sqrt{-7})$.

In order to apply this transformation modulo J_k, we need a square root of -7 in $\mathbb{Z}/J_k\mathbb{Z}$. If J_k is prime and $d = 7^{(J_k+1)/4}$, then

$$d^2 \equiv 7^{(J_k-1)/2} \cdot 7 \equiv \left(\frac{7}{J_k}\right) 7 \equiv -7 \pmod{J_k},$$

since $J_k \equiv 3 \pmod 4$ and $J_k \equiv 2, 4 \pmod 7$ is a quadratic residue modulo 7. If we find that $d^2 \not\equiv -7 \pmod{J_k}$, then we immediately know that J_k must be composite and no further computation is required.

With the transformation to Montgomery form, the formulas for doubling a point on E_a become particularly simple. If $P = [x_1, y_1, z_1]$ is a projective point on $E_{A,B}$ and $2P = [x_2, y_2, z_2]$, we may determine $[x_2, z_2]$ from $[x_1, z_1]$ via

$$\begin{aligned} 4x_1z_1 &= (x_1 + z_1)^2 - (x_1 - z_1)^2, \\ x_2 &= (x_1 + z_1)^2(x_1 - z_1)^2, \\ z_2 &= 4x_1z_1\big((x_1 - z_1)^2 + C(4x_1z_1)\big), \end{aligned} \tag{5}$$

where $C = \frac{1}{4}(A + 2) = \frac{1}{32}(1 - 3\sqrt{-7})$. Note that C does not depend on P (or even a), and may be precomputed. Thus doubling requires just 2 squarings, 3 multiplications, and 4 additions in $\mathbb{Z}/J_k\mathbb{Z}$.

We now present the algorithm, which exploits the transformation of E_a into Montgomery form. We assume that elements of $\mathbb{Z}/J_k\mathbb{Z}$ are uniquely represented as integers in $[0, J_k - 1]$.

Algorithm 5.1.

Input: Positive integers k and J_k.

Output: *True* if J_k is prime and *false* if J_k is composite.

1. If $k \equiv 0 \pmod 8$ or $k \equiv 6 \pmod{24}$ then return *false*.
2. Compute $d = 7^{(J_k+1)/4} \bmod J_k$.
3. If $d^2 \not\equiv -7 \pmod{J_k}$ then return *false*.
4. Determine a via Table 1, depending on $k \pmod{72}$.
5. Compute $r = (-7 + d)a/2 \bmod J_k$, $B = (7 + 3d)/(56a) \bmod J_k$, and $C = (1 - 3d)/32 \bmod J_k$.
6. Let $x_1 = B(x_0 - r) \bmod J_k$ and $z_1 = 1$, where $P_a = (x_0, y_0)$ is as in Table 1.
7. For i from 1 to $k + 1$, compute $[x_i, z_i]$ from $[x_{i-1}, z_{i-1}]$ via (5).
8. If $\gcd(z_k, J_k) = 1$ and $J_k \mid z_{k+1}$ then return *true*, otherwise return *false*.

The tests in step 1 rule out cases where J_k is divisible by 3 or 5, by Lemma 4.5; J_k is then composite, since $J_k > 5$ for all k. This also ensures $\gcd(a, J_k) = 1$ (see Lemma 4.5), so the divisions in step 5 are all valid (J_k is never divisible by 2 or 7). By Remark 4.2, for $k \geq 6$ the condition $\gcd(z_k, J_k) = 1$ in step 8 can be replaced with $z_k \not\equiv 0 \bmod J_k$.

Proposition 5.2. *Algorithm 5.1 performs* $6k + o(k)$ *multiplications and* $4k$ *additions in* $\mathbb{Z}/J_k\mathbb{Z}$. *Its time complexity is* $O(k^2 \log k \log\log k)$ *and it uses* $O(k)$ *space.*

k	step 2	step 7	k	step 2	step 7	k	step 2	step 7
$2^{10}+1$	0.00	0.01	$2^{14}+1$	0.88	5.50	$2^{17}+1$	133	983
$2^{11}+1$	0.00	0.02	$2^{15}+1$	5.26	32.2	$2^{18}+1$	723	5010
$2^{12}+1$	0.02	0.15	$2^{16}+1$	27.5	183	$2^{19}+1$	3310	23600
$2^{13}+1$	0.15	0.91				$2^{20}+1$	13700	107000

Table 3. Timings for Algorithm 5.1 (CPU seconds on a 3.0 GHz AMD Phenom II 945).

Proof. Using standard techniques for fast exponentiation [46], step 2 uses $k + o(k)$ multiplications in $\mathbb{Z}/J_k\mathbb{Z}$. Steps 5–6 perform $O(1)$ operations in $\mathbb{Z}/J_k\mathbb{Z}$ and step 7 uses $5k$ multiplications and $4k$ additions. The cost of the divisions in step 5 are comparatively negligible, as is the cost of step 8. Multiplications (and additions) in $\mathbb{Z}/J_k\mathbb{Z}$ have a bit complexity of $O(\mathsf{M}(k))$, where $\mathsf{M}(k)$ counts the bit operations needed to multiply two k-bit integers [14, Theorem 9.8]. The bound on the time complexity of Algorithm 5.1 then follows from the Schönhage-Strassen [39] bound: $\mathsf{M}(k) = O(k \log k \log\log k)$. The space complexity bound is immediate: The algorithm only needs to keep track of two pairs $[x_i, z_i]$ and $[x_{i-1}, z_{i-1}]$ at any one time, and elements of $\mathbb{Z}/J_k\mathbb{Z}$ can be represented using $O(k)$ bits. □

Table 3 gives timings for Algorithm 5.1 when implemented using the `gmp` library [19] for all integer arithmetic, including the gcd computations. We list the times for step 2 and step 7 separately (the time spent on the other steps is negligible). In the typical case, where J_k is composite, the algorithm is very likely[1] to terminate in step 2, which effectively determines whether J_k is a strong probable prime base -7, as in [9, Algorithm 3.5.3]. To obtain representative timings at the values of k listed, we temporarily modified the algorithm to skip step 2.

We note that the timings for step 7 are suboptimal due to the fact that we used the `gmp` function `mpz_mod` to perform modular reductions. A lower level implementation (using Montgomery reduction [30], for example) might improve these timings by perhaps 20 or 30 percent.

We remark that Algorithm 5.1 can easily be augmented, at essentially no additional cost, to retain an intermediate point $Q = [x_s, y_s, z_s]$, where $s = k + 1 - r$ is chosen so that the order 2^r of Q is the least power of 2 greater than $(J_k^{1/4}+1)^2$. The value of y_s may be obtained as a square root of $y_s^2 = (x_s^3 + Ax_s^2 z_s + x_s z_s^2)/(Bz_s)$ by computing $(y_s^2)^{(J_k+1)/4}$. When J_k is prime, the algorithm can then output a Pomerance-style certificate $(E_{A,B}, Q, r, J_k)$ for the primality of J_k. This certificate has the virtue that it can be verified using just $2.5k + O(1)$ multiplications in $\mathbb{Z}/J_k\mathbb{Z}$, versus the $6k + o(k)$ multiplications used by Algorithm 5.1, by checking that the point Q has order 2^r on the elliptic curve $E_{A,B} \bmod J_k$.

[1] Indeed, we have yet to encounter even a single J_k that is a strong pseudoprime base -7.

5B. ***Searching for prime values of J_k.*** While one can directly apply Algorithm 5.1 to any particular J_k, when searching a large range $1 \le k \le n$ for prime values of J_k it is more efficient to first *sieve* the interval $[1, n]$ to eliminate values of k for which J_k cannot be prime.

For example, as noted in Lemma 4.5, if $k \equiv 0 \pmod 8$ then J_k is divisible by 3. More generally, for any small prime ℓ, one can very quickly compute $J_k \bmod \ell$ for all $k \le n$ by applying the linear recurrence (4) for J_k, working modulo ℓ. If $\ell < \sqrt{n}$, then the sequence $J_k \bmod \ell$ will necessarily cycle, but in any case it takes very little time to identify all the values of $k \le n$ for which J_k is divisible by ℓ; the

k	J_k	a	k	J_k	a	k	J_k	a
2	11	−1	319	427...247	−5	17807	110...799	−1
3	23	−1	375	307...023	−1	18445	125...407	−5
4	67	−5	467	152...727	−1	19318	793...763	−5
5	151	−1	489	639...239	−1	26207	495...799	−1
7	487	−5	494	204...963	−1	27140	359...907	−1
9	2039	−1	543	115...143	−1	31324	116...867	−5
10	4211	−6	643	145...399	−17	36397	155...007	−5
17	524087	−1	684	321...531	−1	47294	327...963	−1
18	1046579	−1	725	706...551	−1	53849	583...567	−1
28	107...427	−5	1129	291...591	−17	83578	122...491	−6
38	109...043	−1	1428	297...011	−1	114730	593...411	−6
49	225...791	−17	2259	425...023	−1	132269	345...831	−1
53	360...711	−1	2734	415...123	−5	136539	864...023	−1
60	461...451	−1	2828	822...787	−1	147647	599...399	−1
63	368...943	−1	3148	175...227	−5	167068	120...027	−5
65	147...007	−1	3230	849...483	−1	167950	388...883	−5
77	604...191	−1	3779	156...127	−1	257298	104...179	−1
84	773...531	−1	5537	254...887	−1	342647	423...399	−1
87	618...703	−1	5759	171...279	−1	414349	120...207	−5
100	507...507	−5	7069	382...207	−5	418033	118...831	−17
109	259...207	−5	7189	508...207	−5	470053	451...407	−5
147	713...023	−1	7540	233...107	−5	475757	536...791	−1
170	598...611	−1	7729	183...591	−111	483244	347...667	−5
213	526...239	−1	9247	168...687	−5	680337	279...759	−1
235	220...519	−17	10484	398...747	−1	810653	295...711	−1
287	994...999	−1	15795	234...023	−1	857637	115...519	−1
						1111930	767...411	−6

Table 4. Prime values of $J_k \approx 2^{k+2}$ for $k \le 1.2 \times 10^6$. The column labeled a gives the value of the twisting factor.

total time required is just $\widetilde{O}(n \log \ell)$, versus $\widetilde{O}(n^2)$ if one were to instead apply a trial division by ℓ to each J_k.

We used this approach to sieve the interval $[1, n]$ for those k for which J_k is not divisible by any prime $\ell \leq L$. Of course one still needs to consider $J_k \leq L$, but this is a small set consisting of roughly $\log_2 L$ values, each of which can be tested very quickly. With $n = 10^6$ and $L = 2^{35}$, sieving reduces the number of potentially prime J_k by a factor of more than 10, leaving 93,707 integers J_k as candidate primes to be tested with Algorithm 5.1. The prime values of J_k found by the algorithm are listed in Table 4, along with the corresponding value of a. As noted in the introduction, we have extended these results to $n = 1.2 \times 10^6$, finding one additional prime with $k = 1{,}111{,}930$, which is also listed in Table 4. The data in Table 4 suggests that prime values of J_k may be more common than prime values of Mersenne numbers M_n; there are 78 primes J_k with fewer than one million bits, but only 33 Mersenne primes in this range. This can be at least partly explained by the fact that M_n can be prime only when n is prime, whereas the values of k for which J_k can be prime are not so severely constrained. By analyzing these constraints in detail, it may be possible to give a heuristic estimate for the density of primes in the sequence J_k, but we leave this to a future article.

Acknowledgments

We thank Daniel J. Bernstein, François Morain, Carl Pomerance, and Karl Rubin for helpful conversations, and the organizers of ECC 2010, the First Abel Conference, and the AWM Anniversary Conference where useful discussions took place. We thank the reviewers for helpful comments. We also thank Henri Cohen and Richard Pinch for helpful comments given at ANTS-X.

This work was supported by the National Science Foundation under grants CNS-0831004 and DMS-1115455.

References

[1] Alexander Abatzoglou, Alice Silverberg, Andrew V. Sutherland, and Angela Wong, *The Prime Database*: $2^{1111932} + 2 \cdot V(1, 2, 1111930) + 1$, 2012. http://primes.utm.edu/en_US/primes/page.php?id=106847

[2] ACM (ed.), *Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing* (*STOC* '86), New York, Association for Computing Machinery, 1986.

[3] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena, *PRIMES is in P*, Ann. of Math. (2) **160** (2004), no. 2, 781–793. MR 2006a:11170

[4] A. O. L. Atkin and F. Morain, *Elliptic curves and primality proving*, Math. Comp. **61** (1993), no. 203, 29–68. MR 93m:11136

[5] Wieb Bosma, *Primality testing with elliptic curves*, Ph.D. thesis, Mathematisch Instituut, Universiteit van Amsterdam, 1985. http://www.math.ru.nl/~bosma/pubs/PRITwEC1985.pdf

[6] David Broadhurst, *The Prime Database*: $(935695 \cdot 2^{627694} + 3)^2 + (1123581 \cdot 2^{313839})^2$, 2012. http://primes.utm.edu/en_US/primes/page.php?id=108157

[7] Chris Caldwell, *The prime pages*: *prime number research*, *records*, *and resources*, 2012. http://primes.utm.edu/

[8] D. V. Chudnovsky and G. V. Chudnovsky, *Sequences of numbers generated by addition in formal groups and new primality and factorization tests*, Adv. in Appl. Math. **7** (1986), no. 4, 385–434. MR 88h:11094

[9] Richard Crandall and Carl Pomerance, *Prime numbers*: *A computational perspective*, second ed., Springer, New York, 2005. MR 2006a:11005

[10] Jean-Marie De Koninck and Claude Levesque (eds.), *Théorie des nombres*: *Proceedings of the International Conference held at the Université Laval*, *Quebec*, *July* 5–18, 1987, Berlin, de Gruyter, 1989. MR 90f:11002

[11] Robert Denomme and Gordan Savin, *Elliptic curve primality tests for Fermat and related primes*, J. Number Theory **128** (2008), no. 8, 2398–2412. MR 2009c:11208

[12] Graham Everest, Alf van der Poorten, Igor Shparlinski, and Thomas Ward, *Recurrence sequences*, Mathematical Surveys and Monographs, no. 104, American Mathematical Society, Providence, RI, 2003. MR 2004c:11015

[13] Victor G. Ganzha, Ernst W. Mayr, and Evgenii V. Vorozhtsov (eds.), *Computer algebra in scientific computing*: *Proceedings of the* 9*th International Workshop* (*CASC* 2006) *held in Chişinău*, *September* 11–15, 2006, Lecture Notes in Computer Science, no. 4194, Berlin, Springer, 2006. MR 2007j:68005

[14] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, second ed., Cambridge University Press, Cambridge, 2003. MR 2004g:68202

[15] Andrew M. Gleason (ed.), *Proceedings of the International Congress of Mathematicians* (Berkeley, 1986), vol. 1, Providence, RI, American Mathematical Society, 1987. MR 89c:00042

[16] Shafi Goldwasser and Joe Kilian, *Almost all primes can be quickly certified*, in ACM [2], 1986, pp. 316–329.

[17] ———, *Primality testing using elliptic curves*, J. ACM **46** (1999), no. 4, 450–472. MR 2002e:11182

[18] Daniel M. Gordon, *Pseudoprimes on elliptic curves*, in De Koninck and Levesque [10], 1989, pp. 290–305. MR 91g:11158

[19] Torbjörn Granlund and the GMP development team, *GNU MP*: *The GNU Multiple Precision Arithmetic Library* (version 5.0.1), 2011. http://gmplib.org/

[20] Benedict H. Gross, *An elliptic curve test for Mersenne primes*, J. Number Theory **110** (2005), no. 1, 114–119. MR 2005m:11007

[21] Alexander Gurevich and Boris Kunyavskiĭ, *Primality testing through algebraic groups*, Arch. Math. (Basel) **93** (2009), no. 6, 555–564. MR 2011g:11235

[22] ———, *Deterministic primality tests based on tori and elliptic curves*, Finite Fields Appl. **18** (2012), no. 1, 222–236. MR 2874918

[23] Hideki Imai and Yuliang Zheng (eds.), *Public key cryptography*: *Proceedings of the* 3*rd International Workshop on Practice and Theory in Public Key Cryptosystems* (*PKC* 2000) *held in Melbourne*, *January* 18–20, 2000, Lecture Notes in Computer Science, no. 1751, Berlin, Springer, 2000. MR 2002f:94052

[24] D. H. Lehmer, *An extended theory of Lucas' functions*, Ann. of Math. (2) **31** (1930), no. 3, 419–448. MR 1502953

[25] Franz Lemmermeyer, *Reciprocity laws*: *From Euler to Eisenstein*, Springer, Berlin, 2000. MR 2001i:11009

[26] H. W. Lenstra, Jr., *Elliptic curves and number-theoretic algorithms*, in Gleason [15], 1987, pp. 99–120. http://www.mathunion.org/ICM/ICM1986.1/Main/icm1986.1.0099.0120.ocr.pdf MR 89d:11114

[27] H. W. Lenstra, Jr. and Carl Pomerance, *Primality testing with Gaussian periods*, preprint, 2011. http://www.math.dartmouth.edu/∼carlp/aks041411.pdf

[28] Rudolf Lidl and Harald Niederreiter, *Introduction to finite fields and their applications*, Cambridge University Press, Cambridge, 1994, Revision of the 1986 first edition. MR 95f:11098

[29] J. S. Milne, *Elliptic curves*, BookSurge, Charleston, SC, 2006. MR 2007h:14044

[30] Peter L. Montgomery, *Modular multiplication without trial division*, Math. Comp. **44** (1985), no. 170, 519–521. MR 86e:11121

[31] ———, *Speeding the Pollard and elliptic curve methods of factorization*, Math. Comp. **48** (1987), no. 177, 243–264. MR 88e:11130

[32] François Morain, *Elliptic curves*, *primality proving and some titanic primes*, Journées Arithmétiques (Luminy, 1989), Astérisque, vol. 198-200, 1991, pp. 245–251. MR 92m:11147

[33] ———, *Implementing the asymptotically fast version of the elliptic curve primality proving algorithm*, Math. Comp. **76** (2007), no. 257, 493–505. MR 2007m:11167

[34] Katsuyuki Okeya, Hiroyuki Kurumatani, and Kouichi Sakurai, *Elliptic curves with the Montgomery-form and their cryptographic applications*, in Imai and Zheng [23], 2000, pp. 238–257. MR 2003h:94045

[35] Th. Pépin, *Sur la formule* $2^{2^n} + 1$, C. R. Acad. Sci. Paris **85** (1877), 329–331.

[36] Carl Pomerance, *Very short primality proofs*, Math. Comp. **48** (1987), no. 177, 315–322. MR 88b:11088

[37] ———, *Primality testing*: *variations on a theme of Lucas*, Congr. Numer. **201** (2010), 301–312. MR 2010k:11191

[38] Vaughan R. Pratt, *Every prime has a succinct certificate*, SIAM J. Comput. **4** (1975), no. 3, 214–220. MR 52 #12395

[39] A. Schönhage and V. Strassen, *Schnelle Multiplikation großer Zahlen*, Computing (Arch. Elektron. Rechnen) **7** (1971), 281–292. MR 45 #1431

[40] J.-P. Serre, *A course in arithmetic*, Graduate Texts in Mathematics, no. 7, Springer, New York, 1973. MR 49 #8956

[41] Joseph H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 151, Springer, New York, 1994. MR 96b:11074

[42] H. M. Stark, *Counting points on CM elliptic curves*, Rocky Mountain J. Math. **26** (1996), no. 3, 1115–1138. MR 98b:11060

[43] W. A. Stein et al., *Sage Mathematics Software* (*version* 4.7.1), The Sage Development Team, 2011. http://www.sagemath.org

[44] Yu Tsumura, *Primality tests for* $2^p \pm 2^{(p+1)/2} + 1$ *using elliptic curves*, Proc. Amer. Math. Soc. **139** (2011), no. 8, 2697–2703. MR 2012e:11210

[45] Song Y. Yan and Glyn James, *Testing Mersenne primes with elliptic curves*, in Ganzha et al. [13], 2006, pp. 303–312. MR 2007k:11209

[46] Andrew Chi Chih Yao, *On the evaluation of powers*, SIAM J. Comput. **5** (1976), no. 1, 100–103. MR 52 #16128

ALEXANDER ABATZOGLOU: aabatzog@math.uci.edu
Department of Mathematics, University of California, Irvine, CA 92697, United States

ALICE SILVERBERG: asilverb@math.uci.edu
Mathematics Department, University of California, Irvine, CA 92697-3875, United States

ANDREW V. SUTHERLAND: drew@math.mit.edu
Department of Mathematics, MIT, Cambridge, MA 02139, United States

ANGELA WONG: awong@math.uci.edu
Department of Mathematics, University of California, Irvine, CA 92697, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.21

msp

Imaginary quadratic fields with isomorphic abelian Galois groups

Athanasios Angelakis and Peter Stevenhagen

In 1976, Onabe discovered that, in contrast to the Neukirch-Uchida results that were proved around the same time, a number field K is not completely characterized by its absolute abelian Galois group A_K. The first examples of nonisomorphic K having isomorphic A_K were obtained on the basis of a classification by Kubota of idele class character groups in terms of their infinite families of Ulm invariants, and did not yield a description of A_K. In this paper, we provide a direct "computation" of the profinite group A_K for imaginary quadratic K, and use it to obtain *many* different K that all have the *same minimal* absolute abelian Galois group.

1. Introduction

The absolute Galois group G_K of a number field K is a large profinite group that we cannot currently describe in very precise terms. This makes it impossible to answer fundamental questions on G_K, such as the inverse Galois problem over K. Still, Neukirch [7] proved that normal number fields are completely characterized by their absolute Galois groups: If G_{K_1} and G_{K_2} are isomorphic as topological groups, then K_1 and K_2 are isomorphic number fields. The result was refined by Ikeda, Iwasawa, and Uchida ([8], [9, Chapter XII, §2]), who disposed of the restriction to normal number fields, and showed that every topological isomorphism $G_{K_1} \xrightarrow{\sim} G_{K_2}$ is actually induced by an inner automorphism of $G_{\mathbb{Q}}$. The same statements hold if all absolute Galois groups are replaced by their maximal *prosolvable* quotients.

It was discovered by Onabe [10] that the situation changes if one moves a further step down from G_K, to its maximal *abelian* quotient $A_K = G_K/\overline{[G_K, G_K]}$, which is the Galois group $A_K = \mathrm{Gal}(K^{\mathrm{ab}}/K)$ of the maximal abelian extension K^{ab} of K.

MSC2010: primary 11R37; secondary 20K35.
Keywords: absolute Galois group, class field theory, group extensions.

Even though the Hilbert problem of explicitly generating K^{ab} for general number fields K is still open after more than a century, the group A_K can be described by class field theory, as a quotient of the idele class group of K.

Kubota [5] studied the group X_K of continuous characters on A_K, and expressed the structure of the p-primary parts of this countable abelian torsion group in terms of an infinite number of so-called *Ulm invariants*. It had been shown by Kaplansky [4, Theorem 14] that such invariants determine the isomorphism type of a countable reduced abelian torsion group. Onabe computed the Ulm invariants of X_K explicitly for a number of small imaginary quadratic fields K, and concluded from this that there exist nonisomorphic imaginary quadratic fields K and K' for which the absolute abelian Galois groups A_K and $A_{K'}$ are isomorphic as profinite groups. This may even happen in cases where K and K' have different class numbers, but the explicit example $K = \mathbb{Q}(\sqrt{-2})$, $K' = \mathbb{Q}(\sqrt{-5})$ of this that occurs in Onabe's main theorem [10, Theorem 2] is incorrect. This is because the value of the finite Ulm invariants in [5, Theorem 4] is incorrect for the prime 2 in case the ground field is a special number field in the sense of our Lemma 3.2. As it happens, $\mathbb{Q}(\sqrt{-5})$ and the exceptional field $\mathbb{Q}(\sqrt{-2})$ do have different Ulm invariants at 2. The nature of Kubota's error is similar to an error in Grunwald's theorem that was corrected by a theorem of Wang occurring in Kubota's paper [5, Theorem 1]. It is related to the noncyclic nature of the 2-power cyclotomic extension $\mathbb{Q} \subset \mathbb{Q}(\zeta_{2^\infty})$.

In this paper, we obtain Onabe's corrected results by a direct class field theoretic approach that completely avoids Kubota's dualization and the machinery of Ulm invariants. We show that the imaginary quadratic fields $K \neq \mathbb{Q}(\sqrt{-2})$ that are said to be of 'type A' in [10] share a *minimal* absolute abelian Galois group that can be described completely explicitly as

$$A_K = \widehat{\mathbb{Z}}^2 \times \prod_{n \geq 1} \mathbb{Z}/n\mathbb{Z}.$$

The numerical data that we present suggest that these fields are in fact very common among imaginary quadratic fields: More than 97% of the 2356 fields of odd prime class number $h_K = p < 100$ are of this nature. We believe (Conjecture 7.1) that there are actually *infinitely many* K for which A_K is the minimal group above. Our belief is supported by certain reasonable assumptions on the average splitting behavior of exact sequences of abelian groups, and these assumptions are tested numerically in the final section of the paper.

2. Galois groups as $\widehat{\mathbb{Z}}$-modules

The profinite abelian Galois groups that we study in this paper naturally come with a topology for which the identity has a basis of open neighborhoods that are open subgroups of finite index. This implies that they are not simply $\mathbb{Z}$-modules, but that

the exponentiation in these groups with ordinary integers extends to exponentiation with elements of the profinite completion $\widehat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}$ of $\mathbb{Z}$. By the Chinese remainder theorem, we have a decomposition of the profinite ring $\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$ into a product of rings of p-adic integers, with the index p ranging over all primes. As $\widehat{\mathbb{Z}}$-modules, our Galois groups decompose correspondingly as a product of pro-p-groups.

It is instructive to look first at the $\widehat{\mathbb{Z}}$-module structure of the absolute abelian Galois group $A_{\mathbb{Q}}$ of $\mathbb{Q}$, which we know very explicitly by the Kronecker-Weber theorem. This theorem states that $\mathbb{Q}^{\mathrm{ab}}$ is the maximal cyclotomic extension of $\mathbb{Q}$, and that an element $\sigma \in A_{\mathbb{Q}}$ acts on the roots of unity that generate $\mathbb{Q}^{\mathrm{ab}}$ by exponentiation. More precisely, we have $\sigma(\zeta) = \zeta^u$ for all roots of unity, with u a uniquely defined element in the unit group $\widehat{\mathbb{Z}}^*$ of the ring $\widehat{\mathbb{Z}}$. This yields the well-known isomorphism $A_{\mathbb{Q}} = \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q}) \cong \widehat{\mathbb{Z}}^* = \prod_p \mathbb{Z}_p^*$.

For odd p, the group $\mathbb{Z}_p^*$ consists of a finite torsion subgroup T_p of $(p-1)$-st roots of unity, and we have an isomorphism

$$\mathbb{Z}_p^* = T_p \times (1 + p\mathbb{Z}_p) \cong T_p \times \mathbb{Z}_p$$

because $1 + p\mathbb{Z}_p$ is a free $\mathbb{Z}_p$-module generated by $1 + p$. For $p = 2$ the same is true with $T_2 = \{\pm 1\}$ and $1 + 4\mathbb{Z}_2$ the free $\mathbb{Z}_2$-module generated by $1 + 4 = 5$. Taking the product over all p, we obtain

$$A_{\mathbb{Q}} \cong T_{\mathbb{Q}} \times \widehat{\mathbb{Z}}, \tag{1}$$

with $T_{\mathbb{Q}} = \prod_p T_p$ the product of the torsion subgroups $T_p \subset \mathbb{Q}_p^*$ of the multiplicative groups of the completions $\mathbb{Q}_p$ of $\mathbb{Q}$. More canonically, $T_{\mathbb{Q}}$ is the *closure* of the torsion subgroup of $A_{\mathbb{Q}} = \mathrm{Gal}(\mathbb{Q}^{\mathrm{ab}}/\mathbb{Q})$, and $A_{\mathbb{Q}}/T_{\mathbb{Q}}$ is a free $\widehat{\mathbb{Z}}$-module of rank 1. The invariant field of $T_{\mathbb{Q}}$ inside $\mathbb{Q}^{\mathrm{ab}}$ is the unique $\widehat{\mathbb{Z}}$-extension of $\mathbb{Q}$.

Even though it looks at first sight as if the isomorphism type of $T_{\mathbb{Q}}$ depends on the properties of prime numbers, one should realize that in an infinite product of finite cyclic groups, the Chinese remainder theorem allows us to rearrange factors in many different ways. One has for instance a noncanonical isomorphism

$$T_{\mathbb{Q}} = \prod_p T_p \cong \prod_{n \geq 1} \mathbb{Z}/n\mathbb{Z}, \tag{2}$$

as both of these products, when written as a countable product of cyclic groups of prime power order, have an infinite number of factors $\mathbb{Z}/\ell^k\mathbb{Z}$ for each prime power ℓ^k. Note that, for the product $\prod_p T_p$ of cyclic groups of order $p-1$ (for $p \neq 2$), this statement is not completely trivial: It follows from the existence, by the well-known theorem of Dirichlet, of infinitely many primes p that are congruent to 1 mod ℓ^k, but not to 1 mod ℓ^{k+1}.

Now suppose that K is an arbitrary number field, with ring of integers $\mathbb{O}$. By class field theory, A_K is the quotient of the idele class group $C_K = \big(\prod'_{\mathfrak{p}\leq\infty} K_{\mathfrak{p}}^*\big)/K^*$ of K by the connected component of the identity. In the case of imaginary quadratic fields K, this connected component is the subgroup $K_\infty^* = \mathbb{C}^* \subset C_K$ coming from the unique infinite prime of K, and in this case the Artin isomorphism for the absolute abelian Galois group A_K of K reads

$$A_K = \widehat{K}^*/K^* = \Big(\prod_{\mathfrak{p}}{}' K_{\mathfrak{p}}^*\Big)/K^*. \tag{3}$$

Here $\widehat{K}^* = \prod'_{\mathfrak{p}} K_{\mathfrak{p}}^*$ is the group of *finite* ideles of K, that is, the restricted direct product of the groups $K_{\mathfrak{p}}^*$ at the finite primes $\mathfrak{p}$ of K, taken with respect to the unit groups $\mathbb{O}_{\mathfrak{p}}^*$ of the local rings of integers. For the purposes of this paper, which tries to describe A_K as a profinite abelian group, it is convenient to treat the isomorphism for A_K in (3) as an identity — as we have written it down.

The expression (3) is somewhat more involved than the corresponding identity $A_{\mathbb{Q}} = \widehat{\mathbb{Z}}^*$ for the rational number field, but we will show in Lemma 3.2 that the *inertial part* of A_K, that is, the subgroup $U_K \subset A_K$ generated by all inertia groups $\mathbb{O}_{\mathfrak{p}}^* \subset C_K$, admits a description very similar to (1).

Denote by $\widehat{\mathbb{O}} = \prod_{\mathfrak{p}} \mathbb{O}_{\mathfrak{p}}$ the profinite completion of the ring of integers $\mathbb{O}$ of K. In the case that K is imaginary quadratic, the inertial part of A_K takes the form

$$U_K = \Big(\prod_{\mathfrak{p}} \mathbb{O}_{\mathfrak{p}}^*\Big)/\mathbb{O}^* = \widehat{\mathbb{O}}^*/\mu_K, \tag{4}$$

since the unit group $\mathbb{O}^*$ of $\mathbb{O}$ is then equal to the group μ_K of roots of unity in K. Apart from the quadratic fields of discriminant -3 and -4, which have 6 and 4 roots of unity, respectively, we always have $\mu_K = \{\pm 1\}$, and (4) can be viewed as the analogue for K of the group $\widehat{\mathbb{Z}}^* = A_{\mathbb{Q}}$.

In the next section, we determine the structure of the group $\widehat{\mathbb{O}}^*/\mu_K$. As the approach works for any number field, we will not assume that K is imaginary quadratic until the very end of that section.

3. Structure of the inertial part

Let K be any number field, and $\widehat{\mathbb{O}} = \prod_{\mathfrak{p}} \mathbb{O}_{\mathfrak{p}}$ the profinite completion of its ring of integers. Denote by $T_{\mathfrak{p}} \subset \mathbb{O}_{\mathfrak{p}}^*$ the subgroup of local roots of unity in $K_{\mathfrak{p}}^*$, and put

$$T_K = \prod_{\mathfrak{p}} T_{\mathfrak{p}} \subset \prod_{\mathfrak{p}} \mathbb{O}_{\mathfrak{p}}^* = \widehat{\mathbb{O}}^*. \tag{5}$$

The analogue of (1) for K is the following.

Lemma 3.1. *The closure of the torsion subgroup of* $\widehat{\mathbb{O}}^*$ *is equal to* T_K, *and* $\widehat{\mathbb{O}}^*/T_K$ *is a free* $\widehat{\mathbb{Z}}$*-module of rank* $[K:\mathbb{Q}]$. *Less canonically, we have an isomorphism*

$$\widehat{\mathbb{O}}^* \cong T_K \times \widehat{\mathbb{Z}}^{[K:\mathbb{Q}]}.$$

Proof. As the finite torsion subgroup $T_\mathfrak{p} \subset \mathbb{O}_\mathfrak{p}^*$ is closed in $\mathbb{O}_\mathfrak{p}^*$, the first statement follows from the definition of the product topology on $\widehat{\mathbb{O}}^* = \prod_\mathfrak{p} \mathbb{O}_\mathfrak{p}^*$.

Reduction modulo $\mathfrak{p}$ in the local unit group $\mathbb{O}_\mathfrak{p}^*$ gives rise to an exact sequence

$$1 \longrightarrow 1+\mathfrak{p} \longrightarrow \mathbb{O}_\mathfrak{p}^* \longrightarrow k_\mathfrak{p}^* \longrightarrow 1$$

that can be split by mapping the elements of the unit group $k_\mathfrak{p}^*$ of the residue class field to their Teichmüller representatives in $\mathbb{O}_\mathfrak{p}^*$. These form the cyclic group of order $\#k_\mathfrak{p}^* = N\mathfrak{p}-1$ in $T_\mathfrak{p}$ consisting of the elements of order coprime to $p = \mathrm{char}(k_\mathfrak{p})$. The kernel of reduction $1+\mathfrak{p}$ is by [3, one-unit theorem, p. 231] a finitely generated $\mathbb{Z}_p$-module of free rank $[K_\mathfrak{p}:\mathbb{Q}_p]$ having a finite torsion group consisting of roots of unity in $T_\mathfrak{p}$ of p-power order. Combining these facts, we find that $\mathbb{O}_\mathfrak{p}^*/T_\mathfrak{p}$ is free over $\mathbb{Z}_p$ of rank $[K_\mathfrak{p}:\mathbb{Q}_p]$ or, less canonically, that we have a local isomorphism

$$\mathbb{O}_\mathfrak{p}^* \cong T_\mathfrak{p} \times \mathbb{Z}_p^{[K_\mathfrak{p}:\mathbb{Q}_p]}$$

for each prime $\mathfrak{p}$. Taking the product over all $\mathfrak{p}$, and using the fact that the sum of the local degrees at p equals the global degree $[K:\mathbb{Q}]$, we obtain the desired global conclusion. □

In order to derive a characterization of $T_K = \prod_\mathfrak{p} T_\mathfrak{p}$ for arbitrary number fields K similar to (2), we observe that we have an exact divisibility $\ell^k \parallel \#T_\mathfrak{p}$ of the order of $T_\mathfrak{p}$ by a prime power ℓ^k if and only if the local field $K_\mathfrak{p}$ at $\mathfrak{p}$ contains a primitive ℓ^k-th root of unity, but *not* a primitive ℓ^{k+1}-th root of unity. We may reword this as: The prime $\mathfrak{p}$ splits completely in the cyclotomic extension $K \subset K(\zeta_{\ell^k})$, but *not* in the cyclotomic extension $K \subset K(\zeta_{\ell^{k+1}})$. If such $\mathfrak{p}$ exist at all for ℓ^k, then there are infinitely many of them, by the Chebotarev density theorem.

Thus, T_K can be written as a product of groups $(\mathbb{Z}/\ell^k\mathbb{Z})^\mathbb{Z} = \mathrm{Map}(\mathbb{Z}, \mathbb{Z}/\ell^k\mathbb{Z})$ that are themselves countable products of cyclic groups of order ℓ^k. The prime powers $\ell^k > 1$ that occur for K are *all* but those for which we have an equality

$$K(\zeta_{\ell^k}) = K(\zeta_{\ell^{k+1}}).$$

For $K = \mathbb{Q}$ all prime powers ℓ^k occur, but for general K, there are finitely many prime powers that may disappear. This is due to the fact that the infinite cyclotomic extension $\mathbb{Q} \subset \mathbb{Q}(\zeta_{\ell^\infty})$ with group $\mathbb{Z}_\ell^*$ can partially "collapse" over K.

To describe the exceptional prime powers ℓ^k that disappear for K, we consider, for ℓ an *odd* prime, the number

$$w(\ell) = w_K(\ell) = \#\mu_{\ell^\infty}(K(\zeta_\ell))$$

of ℓ-power roots of unity in the field $K(\zeta_\ell)$. For almost all ℓ, this number equals ℓ, and we call ℓ *exceptional* for K if it is divisible by ℓ^2. Note that no odd exceptional prime numbers exist for imaginary quadratic fields K.

For the prime $\ell = 2$, we consider instead the number

$$w(2) = w_K(2) = \#\mu_{2^\infty}(K(\zeta_4))$$

of 2-power roots in $K(\zeta_4) = K(i)$. If K contains $i = \zeta_4$, or if $w(2)$ is divisible by 8, we call 2 *exceptional* for K. Note that the only imaginary quadratic fields K for which 2 is exceptional are $\mathbb{Q}(i)$ and $\mathbb{Q}(\sqrt{-2})$.

The number $w(K)$ of *exceptional roots of unity* for K is now defined as

$$w(K) = \prod_{\ell \text{ exceptional}} w(\ell).$$

Note that $w(K)$ refers to roots of unity that may or may not be contained in K itself, and that every prime ℓ dividing $w(K)$ occurs with exponent at least 2. The prime powers $\ell^k > 1$ that do *not* occur when T_K is written as a direct product of groups $(\mathbb{Z}/\ell^k\mathbb{Z})^{\mathbb{Z}}$ are the *strict* divisors of $w(\ell)$ at exceptional primes ℓ, with the exceptional prime $\ell = 2$ giving rise to a special case.

Lemma 3.2. *Let K be a number field, and $w = w(K)$ its number of exceptional roots of unity. Then we have a noncanonical isomorphism of profinite groups*

$$T_K = \prod_{\mathfrak{p}} T_{\mathfrak{p}} \cong \prod_{n \geq 1} \mathbb{Z}/nw\mathbb{Z},$$

except when 2 is exceptional for K and $i = \zeta_4$ is not contained in K. In this special case, we have

$$T_K = \prod_{\mathfrak{p}} T_{\mathfrak{p}} \cong \prod_{n \geq 1} (\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/nw\mathbb{Z}).$$

The group T_K is isomorphic to the group $T_{\mathbb{Q}}$ in (2) *if and only if we have $w = 1$.*

Proof. If ℓ is odd, the tower of field extensions

$$K(\zeta_\ell) \subset K(\zeta_{\ell^2}) \subset \cdots \subset K(\zeta_{\ell^k}) \subset K(\zeta_{\ell^{k+1}}) \subset \cdots$$

is a $\mathbb{Z}_\ell$-extension, and the steps $K(\zeta_{\ell^k}) \subset K(\zeta_{\ell^{k+1}})$ with $k \geq 1$ in this tower that are equalities are exactly those for which ℓ^{k+1} divides $w(\ell)$.

Similarly, the tower of field extensions

$$K(\zeta_4) \subset K(\zeta_8) \subset \cdots \subset K(\zeta_{2^k}) \subset K(\zeta_{2^{k+1}}) \subset \cdots$$

is a $\mathbb{Z}_2$-extension in which the steps $K(\zeta_{2^k}) \subset K(\zeta_{2^{k+1}})$ with $k \geq 2$ that are equalities are exactly those for which 2^{k+1} divides $w(2)$. The extension $K = K(\zeta_2) \subset K(\zeta_4)$ that we have in the remaining case $k = 1$ is an equality if and only if K contains $i = \zeta_4$.

Thus, a prime power $\ell^k > 2$ that does not occur when T_K is written as a product of groups $(\mathbb{Z}/\ell^k\mathbb{Z})^{\mathbb{Z}}$ is the same as a *strict* divisor $\ell^k > 2$ of $w(\ell)$ at an exceptional prime ℓ. The special prime power $\ell^k = 2$ does not occur if and only if $i = \zeta_4$ is in K. Note that in this case, 2 is by definition exceptional for K.

It is clear that replacing the group $\prod_{n\geq 1} \mathbb{Z}/n\mathbb{Z}$ from (2) by $\prod_{n\geq 1} \mathbb{Z}/nw\mathbb{Z}$ has the effect of removing cyclic summands of order ℓ^k with $\ell^{k+1} \mid w$, and this shows that the groups given in the Lemma are indeed isomorphic to T_K. Only for $w = 1$ we obtain the group $T_{\mathbb{Q}}$ in which all prime powers ℓ^k arise. □

Lemmas 3.1 and 3.2 tell us what $\widehat{\mathbb{O}}^*$ looks like as a $\widehat{\mathbb{Z}}$-module. In particular, it shows that the dependence on K is limited to the degree $[K : \mathbb{Q}]$, which is reflected in the rank of the free $\widehat{\mathbb{Z}}$-part of $\widehat{\mathbb{O}}^*$, and the nature of the exceptional roots of unity for K. For the group $\widehat{\mathbb{O}}^*/\mu_K$, the same is true, but the proof requires an extra argument, and the following lemma.

Lemma 3.3. *There are infinitely many primes* $\mathfrak{p}$ *of* K *for which we have*

$$\gcd(\#\mu_K, \#T_{\mathfrak{p}}/\#\mu_K) = 1.$$

Proof. For every prime power $\ell^k > 1$ that exactly divides $\#\mu_K$, the extension $K = K(\zeta_{\ell^k}) \subset K(\zeta_{\ell^{k+1}})$ is a cyclic extension of prime degree ℓ. For different prime powers $\ell^k \parallel \#\mu_K$, we get different extensions, so infinitely many primes $\mathfrak{p}$ of K are inert in all of them. For such $\mathfrak{p}$, we have $\gcd(\#\mu_K, \#T_{\mathfrak{p}}/\#\mu_K) = 1$. □

Lemma 3.4. *We have a noncanonical isomorphism* $T_K/\mu_K \cong T_K$.

Proof. Pick a prime $\mathfrak{p}_0$ of K that satisfies the conditions of Lemma 3.3. Then μ_K embeds as a direct summand in $T_{\mathfrak{p}_0}$, and we can write $T_{\mathfrak{p}_0} \cong \mu_K \times T_{\mathfrak{p}_0}/\mu_K$ as a product of two cyclic groups of coprime order. It follows that the natural exact sequence

$$1 \longrightarrow \prod_{\mathfrak{p}\neq\mathfrak{p}_0} T_{\mathfrak{p}} \longrightarrow T_K/\mu_K \longrightarrow T_{\mathfrak{p}_0}/\mu_K \longrightarrow 1$$

can be split using the composed map $T_{\mathfrak{p}_0}/\mu_K \to T_{\mathfrak{p}_0} \to T_K \to T_K/\mu_K$. This makes T_K/μ_K isomorphic to the product of $\prod_{\mathfrak{p}\neq\mathfrak{p}_0} T_{\mathfrak{p}}$ and a cyclic group for which the order is a product of prime powers that already "occur" infinitely often in T_K. Thus T_K/μ_K is isomorphic to a product of exactly the same groups $(\mathbb{Z}/\ell^k\mathbb{Z})^{\mathbb{Z}}$ that occur in T_K, and therefore isomorphic to T_K itself. □

For imaginary quadratic K, where $\widehat{\mathbb{O}}^*/\mu_K$ constitutes the inertial part U_K of A_K from (4), we summarize the results of this section in the following way.

Theorem 3.5. *Let K be an imaginary quadratic field. Then the subgroup T_K/μ_K of U_K is a direct summand of U_K. For $K \neq \mathbb{Q}(i), \mathbb{Q}(\sqrt{-2})$, we have isomorphisms*

$$U_K = \widehat{\mathbb{O}}^*/\mu_K \cong \widehat{\mathbb{Z}}^2 \times (T_K/\mu_K) \cong \widehat{\mathbb{Z}}^2 \times \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z}$$

of profinite groups.

For K equal to $\mathbb{Q}(i)$ or $\mathbb{Q}(\sqrt{-2})$, the prime 2 is exceptional for K, and the groups $T_K/\mu_K \cong T_K$ are different as they do not have cyclic summands of order 2 and 4, respectively.

4. Extensions of Galois groups

In the previous section, all results could easily be stated and proved for arbitrary number fields. From now on, K will denote an imaginary quadratic field. In order to describe the full group A_K from (3), we consider the exact sequence

$$1 \longrightarrow U_K = \widehat{\mathbb{O}}^*/\mu_K \longrightarrow A_K = \widehat{K}^*/K^* \xrightarrow{\psi} \mathrm{Cl}_K \longrightarrow 1 \tag{6}$$

that describes the class group Cl_K of K in idelic terms. Here ψ maps the class of the finite idele $(x_\mathfrak{p})_\mathfrak{p} \in \widehat{K}^*$ to the class of its associated ideal $\prod_\mathfrak{p} \mathfrak{p}^{e_\mathfrak{p}}$, with $e_\mathfrak{p} = \mathrm{ord}_\mathfrak{p}\, x_\mathfrak{p}$.

The sequence (6) shows that U_K is an open subgroup of A_K of index equal to the class number h_K of K. In view of Theorem 3.5, this immediately yields Onabe's discovery that different K can have the same absolute abelian Galois group.

Theorem 4.1. *An imaginary quadratic number field $K \neq \mathbb{Q}(i), \mathbb{Q}(\sqrt{-2})$ of class number 1 has absolute abelian Galois group isomorphic to*

$$G = \widehat{\mathbb{Z}}^2 \times \prod_{n \geq 1} \mathbb{Z}/n\mathbb{Z}.$$

In Onabe's paper [10, §5], the group G, which is not explicitly given but characterized by its infinitely many Ulm invariants, is referred to as 'of type A'. We will refer to G as the *minimal* Galois group, as every absolute abelian Galois group of an imaginary quadratic field $K \neq \mathbb{Q}(i), \mathbb{Q}(\sqrt{-2})$ contains a subgroup isomorphic to G. We will show that there are actually *many* more K having this absolute abelian Galois group than the seven fields K of class number 1 to which the preceding theorem applies.

Now take for K any imaginary quadratic field of class number $h_K > 1$. Then Theorem 3.5 and the sequence (6) show that A_K is an abelian group extension of Cl_K by the minimal Galois group G from Theorem 4.1. If the extension (6) were split, we would find that A_K is isomorphic to $G \times \mathrm{Cl}_K \cong G$; but it turns out that splitting at this level *never* occurs for nontrivial Cl_K, in the following strong sense.

Theorem 4.2. *For every imaginary quadratic field K of class number $h_K > 1$, the sequence* (6) *is totally nonsplit*; *that is*, *there is no nontrivial subgroup $C \subset \mathrm{Cl}_K$ for which the associated subextension $1 \to U_K \to \psi^{-1}[C] \to C \to 1$ is split.*

Proof. Suppose there is a non-trivial subgroup $C \subset \mathrm{Cl}_K$ over which the extension (6) splits, and pick $[\mathfrak{a}] \in C$ of prime order p. Then there exists an element

$$\big((x_\mathfrak{p})_\mathfrak{p} \bmod K^*\big) \in \psi^{-1}([\mathfrak{a}]) \subset A_K = \widehat{K}^*/K^*$$

of order p. In other words, there exists $\alpha \in K^*$ such that we have $x_\mathfrak{p}^p = \alpha \in K_\mathfrak{p}^*$ for all $\mathfrak{p}$, and such that α generates the ideal $\mathfrak{a}^p$. But this implies by [1, Chapter IX, Theorem 1] that α is a p-th power in K^*, and hence that $\mathfrak{a}$ is a principal ideal. Contradiction. □

At first sight, Theorem 4.2 seems to indicate that in the case $h_K > 1$, the group A_K will *not* be isomorphic to the minimal Galois group $G \cong U_K$. However, finite abelian groups requiring no more than k generators do allow extensions by free $\widehat{\mathbb{Z}}$-modules of finite rank k that are again free of rank k, just like they do with free $\mathbb{Z}$-modules in the classical setting of finitely generated abelian groups. The standard example for $k = 1$ is the extension

$$1 \longrightarrow \widehat{\mathbb{Z}} \xrightarrow{\times p} \widehat{\mathbb{Z}} \longrightarrow \mathbb{Z}/p\mathbb{Z} \longrightarrow 1$$

for an integer $p \neq 0$, prime or not. Applying to this the functor $\mathrm{Hom}(-, M)$ for a multiplicatively written $\widehat{\mathbb{Z}}$-module M, we obtain an isomorphism

$$M/M^p \xrightarrow{\sim} \mathrm{Ext}(\mathbb{Z}/p\mathbb{Z}, M) \tag{7}$$

by the Hom-Ext-sequence from homological algebra [6]. We will use it in Section 5.

Lemma 4.3. *Let B be a finite abelian group, F a free $\widehat{\mathbb{Z}}$-module of finite rank k, and*

$$1 \longrightarrow F \longrightarrow E \longrightarrow B \longrightarrow 1$$

an exact sequence of $\widehat{\mathbb{Z}}$-modules. Then E is free of rank k if and only if this sequence is totally nonsplit.

Proof. One may reduce the statement to the familiar case of modules over principal ideal domains by writing $\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$, and consider the individual p-parts of the sequence. As a matter of convention, note that in the degenerate case where B is the trivial group, there are no nontrivial subgroups $C \subset B$ over which the sequence splits, making the sequence by definition totally nonsplit. □

In order to apply the preceding lemma, we replace the extension (6) by the pushout under the quotient map $U_K = \widehat{\mathbb{O}}^*/\mu_K \to U_K/T_K = \widehat{\mathbb{O}}^*/T_K$ from U_K to

its maximal $\widehat{\mathbb{Z}}$-free quotient. This yields the exact sequence of $\widehat{\mathbb{Z}}$-modules

$$1 \longrightarrow \widehat{\mathbb{O}}^*/T_K \longrightarrow \widehat{K}^*/(K^* \cdot T_K) \longrightarrow \mathrm{Cl}_K \longrightarrow 1 \tag{8}$$

in which Cl_K is finite and $\widehat{\mathbb{O}}^*/T_K$ is free of rank 2 over $\widehat{\mathbb{Z}}$ by Lemma 3.1.

Theorem 4.4. *Let K be an imaginary quadratic field of class number $h_K > 1$, and suppose the sequence* (8) *is totally nonsplit. Then the absolute abelian Galois group of K is the minimal group G occurring in Theorem 4.1.*

Proof. If the extension (8) is totally nonsplit, then $\widehat{K}^*/(K^* \cdot T_K)$ is free of rank 2 over $\widehat{\mathbb{Z}}$ by Lemma 4.3. In this case the exact sequence of $\widehat{\mathbb{Z}}$-modules

$$1 \longrightarrow T_K/\mu_K \longrightarrow A_K = \widehat{K}^*/K^* \longrightarrow \widehat{K}^*/(K^* \cdot T_K) \longrightarrow 1$$

is split, and A_K is isomorphic to $U_K = G = \widehat{\mathbb{Z}}^2 \times (T_K/\mu_K)$. □

Remark. We will use Theorem 4.4 in this paper to find many imaginary quadratic fields K having the same minimal absolute abelian Galois group G. It is however interesting to note that this is the only way in which this can be done, as Theorem 4.4 actually admits a converse: If the absolute abelian Galois group of an imaginary quadratic field K of class number $h_K > 1$ is the minimal group G, then the sequence (8) is totally nonsplit. The proof, which we do not include in this paper, will be given in the forthcoming doctoral thesis of the first author.

It is instructive to see what all the preceding extensions of Galois groups amount to in terms of field extensions. The diagram of fields in Figure 1 lists all subfields of the extension $K \subset K^{\mathrm{ab}}$ corresponding to the various subgroups we considered in analyzing the structure of $A_K = \mathrm{Gal}(K^{\mathrm{ab}}/K)$.

We denote by H the Hilbert class field of K. This is the maximal totally unramified abelian extension of K, and it is finite over K with group Cl_K. The inertial part of A_K is the Galois group $U_K = \mathrm{Gal}(K^{\mathrm{ab}}/H)$, which is isomorphic to G for all imaginary quadratic fields $K \neq \mathbb{Q}(i), \mathbb{Q}(\sqrt{-2})$. The fundamental sequence (6) corresponds to the tower of fields

$$K \subset H \subset K^{\mathrm{ab}}.$$

By Theorem 3.5, the invariant field L of the closure T_K/μ_K of the torsion subgroup of U_K is an extension of H with group $\widehat{\mathbb{Z}}^2$. The tower of field extensions

$$K \subset H \subset L$$

corresponds to the exact sequence of Galois groups (8).

We define L_0 as the "maximal $\widehat{\mathbb{Z}}$-extension" of K, that is, as the compositum of the $\mathbb{Z}_p$-extensions of K for *all* primes p. As is well-known, an imaginary quadratic field admits two independent $\mathbb{Z}_p$-extensions for each prime p, so $F = \mathrm{Gal}(L_0/K)$

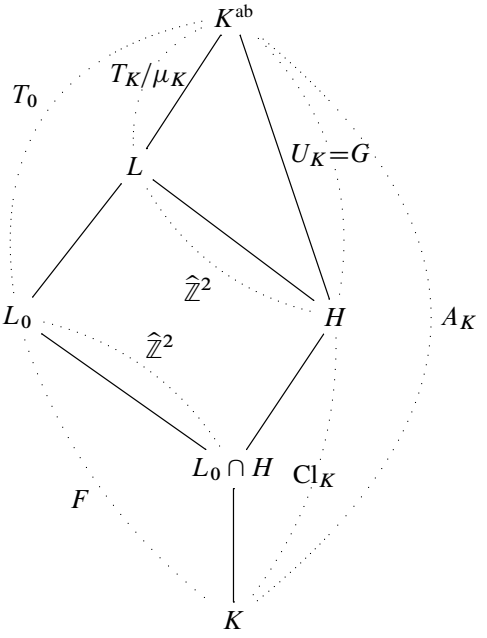


Figure 1. The structure of $A_K = \mathrm{Gal}(K^{\mathrm{ab}}/K)$.

is a free $\widehat{\mathbb{Z}}$-module of rank 2, and L_0 is the invariant field under the closure T_0 of the torsion subgroup of A_K. The image of the restriction map $T_0 \to \mathrm{Cl}_K$ is the maximal subgroup of Cl_K over which (8) splits. The invariant subfield of H corresponding to it is the intersection $L_0 \cap H$. The totally nonsplit case occurs when H is contained in L_0, leading to $L_0 \cap H = H$ and $L_0 = L$. In this case $\mathrm{Gal}(L/K) = \mathrm{Gal}(L_0/K)$ is itself a free $\widehat{\mathbb{Z}}$-module of rank 2, and A_K is an extension of $\widehat{\mathbb{Z}}^2$ by T_K/μ_K that is isomorphic to G.

5. Finding minimal Galois groups

In order to use Theorem 4.4 and find imaginary quadratic K for which the absolute abelian Galois group A_K is the minimal group G from Theorem 4.1, we need an algorithm that can effectively determine, on input K, whether the sequence of $\widehat{\mathbb{Z}}$-modules

$$1 \longrightarrow \widehat{\mathbb{O}}^*/T_K \longrightarrow \widehat{K}^*/(K^* \cdot T_K) \longrightarrow \mathrm{Cl}_K \longrightarrow 1 \tag{8}$$

from Section 4 is totally nonsplit. This means that for every ideal class $[\mathfrak{a}] \in \mathrm{Cl}_K$ of prime order, the subextension $E[\mathfrak{a}]$ of (8) lying over the subgroup $\langle [\mathfrak{a}] \rangle \subset \mathrm{Cl}_K$ is nonsplit.

Any profinite abelian group M is a module over $\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p$, and can be written accordingly as a product $M = \prod_p M_p$ of p-primary parts, where $M_p = M \otimes_{\widehat{\mathbb{Z}}} \mathbb{Z}_p$ is a pro-p-group and $\mathbb{Z}_p$-module. In the same way, an exact sequence of $\widehat{\mathbb{Z}}$-modules is a "product" of exact sequences for their p-primary parts, and splitting over a group of prime order p only involves p-primary parts for that p.

For the free $\widehat{\mathbb{Z}}$-module $M = \widehat{\mathbb{O}}^*/T_K$ in (8), we write T_p for the torsion subgroup of $\mathbb{O}_p^* = (\mathbb{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p)^* = \prod_{\mathfrak{p}|p} \mathbb{O}_{\mathfrak{p}}^*$. Then the p-primary part of M is the pro-p-group

$$M_p = \mathbb{O}_p^*/T_p = \prod_{\mathfrak{p}|p} (\mathbb{O}_{\mathfrak{p}}^*/T_{\mathfrak{p}}) \cong \mathbb{Z}_p^2. \tag{9}$$

In order to verify the hypothesis of Theorem 4.4, we need to check that the extension $E[\mathfrak{a}]$ has nontrivial class in $\mathrm{Ext}(\langle[\mathfrak{a}]\rangle, M)$ for all $[\mathfrak{a}] \in \mathrm{Cl}_K$ of prime order p. We can do this by verifying in each case that the element of $M/M^p = M_p/M_p^p$ corresponding to it under the isomorphism (7) is nontrivial. This yields the following theorem.

Theorem 5.1. *Let K be an imaginary quadratic field, and define for each prime number p dividing h_K the homomorphism*

$$\phi_p : \mathrm{Cl}_K[p] \longrightarrow \mathbb{O}_p^*/T_p(\mathbb{O}_p^*)^p$$

that sends the class of a p-torsion ideal $\mathfrak{a}$ coprime to p to the class of a generator of the ideal $\mathfrak{a}^p$. Then (8) *is totally nonsplit if and only if all maps ϕ_p are injective.*

Proof. Under the isomorphism (7), the class of the extension

$$1 \longrightarrow M \longrightarrow E \xrightarrow{f} \mathbb{Z}/p\mathbb{Z} \longrightarrow 1$$

in $\mathrm{Ext}(\mathbb{Z}/p\mathbb{Z}, M)$ corresponds by [6, Chapter III, Proposition 1.1] to the residue class of the element

$$(f^{-1}(1 \bmod p\mathbb{Z}))^p \in M/M^p.$$

In the case of $E[\mathfrak{a}]$, we apply this to $M = \widehat{\mathbb{O}}^*/T_K$, and choose the identification $\mathbb{Z}/p\mathbb{Z} = \langle[\mathfrak{a}]\rangle$ under which $1 \bmod p\mathbb{Z}$ is the *inverse* of $[\mathfrak{a}]$. Then $f^{-1}(1 \bmod p\mathbb{Z})$ is the residue class in $\widehat{K}^*/(K^* \cdot T_K)$ of any finite idele $x \in \widehat{K}^*$ that is mapped to ideal class of $\mathfrak{a}^{-1}$ under the map ψ from (6).

We pick $\mathfrak{a}$ in its ideal class coprime to p, and take for $x = (x_{\mathfrak{p}})_{\mathfrak{p}}$ an idele that locally generates $\mathfrak{a}^{-1}$ at all p. If $\alpha \in K^*$ generates $\mathfrak{a}^p$, then $x^p\alpha$ is an idele in $\widehat{\mathbb{O}}^*$ that lies in the same class modulo K^* as x^p, and its image

$$(f^{-1}(1 \bmod p\mathbb{Z}))^p = x^p = x^p\alpha \in M/M^p = M_p/M_p^p = \mathbb{O}_p^*/T_p(\mathbb{O}_p^*)^p$$

corresponds to the class of $E[\mathfrak{a}]$ in $\mathrm{Ext}(\langle[\mathfrak{a}]\rangle, \mathbb{O}^*/T_K)$. As the idele $x = (x_{\mathfrak{p}})_{\mathfrak{p}}$ has components $x_{\mathfrak{p}} \in \mathbb{O}_{\mathfrak{p}}^*$ at $\mathfrak{p} \mid p$ by the choice of $\mathfrak{a}$, we see that this image in

$M_p/M_p^p = \mathbb{O}_p^*/T_p(\mathbb{O}_p^*)^p$ is the element $\phi_p([\mathfrak{a}])$ we defined. The map ϕ_p is clearly a homomorphism, and we want it to assume nontrivial values on the elements of order p in $\mathrm{Cl}_K[p]$, for each prime p dividing h_K. The result follows. □

Remark. In Theorem 5.1, it is not really necessary to restrict to representing ideals $\mathfrak{a}$ that are coprime to p. One may take $K_p^*/T_p(K_p^*)^p$ as the target space of ϕ_p to accommodate all $\mathfrak{a}$, with $K_p = K \otimes_{\mathbb{Z}} \mathbb{Z}_p$, and observe that the image of ϕ_p is in the subgroup $\mathbb{O}_p^*/T_p(\mathbb{O}_p^*)^p$, as the valuations of $\mathfrak{a}^p$ at the primes over p are divisible by p.

Remark. It is possible to prove Theorem 5.1 without explicit reference to homological algebra. What the proof shows is that, in order to lift an ideal class of arbitrary order n under (8), it is necessary and sufficient that its n-th power is generated by an element α that is locally everywhere a n-th power *up to multiplication by local roots of unity.* This extra leeway in comparison with the situation in Theorem 4.2 makes it into an interesting splitting problem for the group extensions involved, as this condition on α may or may not be satisfied. Note that at primes outside n, the divisibility of the valuation of α by n automatically implies the local condition.

In Onabe's paper, which assumes throughout that Cl_K itself is a cyclic group of prime order, the same criterion is obtained from an analysis of the Ulm invariants occurring in Kubota's setup [5].

Our Theorem 5.1 itself does not assume any restriction on Cl_K, but its use in finding K with minimal absolute Galois group G does imply certain restrictions on the structure of Cl_K. The most obvious implication of the injectivity of the map ϕ_p in the theorem is a bound on the p-rank of Cl_K, which is defined as the dimension of the group $\mathrm{Cl}_K / \mathrm{Cl}_K^p$ as an $\mathbb{F}_p$-vector space.

Corollary 5.2. *If* Cl_K *has* p*-rank at least* 3 *for some* p*, then the sequence* (8) *splits over a subgroup of* Cl_K *of order* p*.*

Proof. It follows from the isomorphism in (9) that the image of ϕ_p lies in a group that is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^2$. If Cl_K has p-rank at least 3, then ϕ_p will not be injective. Now apply Theorem 5.1. □

As numerical computations in uncountable $\widehat{\mathbb{Z}}$-modules such as $\widehat{K}^*/(K^* \cdot T_K)$ can only be performed with finite precision, it is not immediately obvious that the splitting type of an idelic extension as (8) can be found by a finite computation. The maps ϕ_p in Theorem 5.1 however are linear maps between finite-dimensional $\mathbb{F}_p$-vector spaces that lend themselves very well to explicit computations. One just needs some standard algebraic number theory to compute these spaces explicitly. A high-level description of an algorithm that determines whether the extension (8) is totally nonsplit is then easily written down.

Algorithm 5.3.

Input: An imaginary quadratic number field K.

Output: *No* if the extension (8) for K is not totally nonsplit, *yes* otherwise.

1. Compute the class group Cl_K of K. If Cl_K has p-rank at least 3 for some p, output *no* and stop.
2. For each prime p dividing h_K, compute $n \in \{1, 2\}$ $\mathbb{O}$-ideals coprime to p such that their classes in Cl_K generate $\mathrm{Cl}_K[p]$, and generators x_1 up to x_n for their p-th powers. Check whether x_1 is trivial in $\mathbb{O}_p^*/T_p(\mathbb{O}_p^*)^p$. If it is, output *no* and stop. If $n = 2$, check whether x_2 is trivial in $\mathbb{O}_p^*/T_p \cdot \langle x_1 \rangle \cdot (\mathbb{O}_p^*)^p$. If it is, output *no* and stop.
3. If all primes $p \mid h_K$ are dealt with without stopping, output *yes* and stop.

Step 1 is a standard task in computational algebraic number theory. For imaginary quadratic fields, it is often implemented in terms of binary quadratic forms, and particularly easy. From an explicit presentation of the group, it is also standard to find the global elements x_1 and, if needed, x_2. The rest of Step 2 takes place in a *finite* group, and this means that we only compute in the rings $\mathbb{O}_p$ up to small precision. For instance, computations in $\mathbb{Z}_p^*/T_p(\mathbb{Z}_p^*)^p$ amount to computations modulo p^2 for odd p, and modulo p^3 for $p = 2$.

6. Splitting behavior at 2

The splitting behavior of the sequence (8) depends strongly on the structure of the p-primary parts of Cl_K at the primes $p \mid h_K$. In view of Theorem 5.1 and Corollary 5.2, fields with cyclic class groups and few small primes dividing h_K appear to be more likely to have minimal Galois group G. In Section 7, we will provide numerical data to examine the average splitting behavior.

For odd primes p, class groups of p-rank at least 3 arising in Corollary 5.2 are very rare, at least numerically and according to the Cohen-Lenstra heuristics. At the prime 2, the situation is a bit different, as the 2-torsion subgroup of Cl_K admits a classical explicit description going back to Gauss. Roughly speaking, his theorem on ambiguous ideal classes states that $\mathrm{Cl}_K[2]$ is an $\mathbb{F}_2$-vector space generated by the classes of the primes $\mathfrak{p}$ of K lying over the rational primes that ramify in $\mathbb{Q} \subset K$, subject to a single relation coming from the principal ideal $(\sqrt{D_K})$. Thus, the 2-rank of Cl_K for a discriminant with t distinct prime divisors equals $t - 1$. In view of Corollary 5.2, our method to construct K with absolute abelian Galois group G does not apply if the discriminant D_K of K has more than 3 distinct prime divisors.

If $-D_K$ is a prime number, then h_K is odd, and there is nothing to check at the prime 2.

For D_K with two distinct prime divisors, the 2-rank of Cl_K equals 1, and we can replace the computation at $p = 2$ in Algorithm 5.3 by something that is much simpler.

Theorem 6.1. *Let K be an imaginary quadratic field with even class number, and suppose that its 2-class group is cyclic. Then the sequence* (8) *is nonsplit over* $\mathrm{Cl}_K[2]$ *if and only if the discriminant D_K of K is of one of the following types*:

(1) $D_K = -pq$ *for primes* $p \equiv -q \equiv 5 \bmod 8$;

(2) $D_K = -4p$ *for a prime* $p \equiv 5 \bmod 8$;

(3) $D_K = -8p$ *for a prime* $p \equiv \pm 3 \bmod 8$.

Proof. If K has a nontrivial cyclic 2-class group, then $D_K \equiv 0, 1 \bmod 4$ is divisible by exactly two different primes.

If D_K is odd, we have $D_K = -pq$ for primes $p \equiv 1 \bmod 4$ and $q \equiv 3 \bmod 4$, and the ramified primes $\mathfrak{p}$ and $\mathfrak{q}$ of K are in the unique ideal class of order 2 in Cl_K. Their squares are ideals generated by the integers p and $-q$ that become squares in the genus field $F = \mathbb{Q}(\sqrt{p}, \sqrt{-q})$ of K, which is a quadratic extension of K with group $C_2 \times C_2$ over $\mathbb{Q}$ that is locally unramified at 2.

If we have $D_K \equiv 5 \bmod 8$, then 2 is inert in $\mathbb{Q} \subset K$, and 2 splits in $K \subset F$. This means that K and F have isomorphic completions at their primes over 2, and that p and $-q$ are local squares at 2. In this case ϕ_2 is the trivial map in Theorem 5.1, and is not injective.

If we have $D \equiv 1 \bmod 8$ then 2 splits in $\mathbb{Q} \subset K$. In the case $p \equiv -q \equiv 1 \bmod 8$ the integers p and $-q$ are squares in $\mathbb{Z}_2^*$, and ϕ_2 is again the trivial map. In the other case $p \equiv -q \equiv 5 \bmod 8$, the generators p and $-q$ are nonsquares in $\mathbb{Z}_2^*$, also up to multiplication by elements in $T_2 = \{\pm 1\}$. In this case ϕ_2 is injective.

If D_K is even, we either have $D_K = -4p$ for a prime $p \equiv 1 \bmod 4$ or $D_K = -8p$ for an odd prime p. In the case $D_K = -4p$ the ramified prime over 2 is in the ideal class of order 2. For $p \equiv 1 \bmod 8$, the local field $\mathbb{Q}_2(\sqrt{-p}) = \mathbb{Q}_2(i)$ contains a square root of $2i$, and ϕ_2 is not injective. For $p \equiv 5 \bmod 8$, the local field $\mathbb{Q}_2(\sqrt{-p}) = \mathbb{Q}_2(\sqrt{3})$ does not contain a square root of ± 2, and ϕ_2 is injective. In the case $D_K = -8p$ the ramified primes over both 2 and p are in the ideal class of order 2. For $p \equiv \pm 1 \bmod 8$ the generator $\pm p$ is a local square at 2. For $p \equiv \pm 3 \bmod 8$ it is not. □

In the case where the 2-rank of Cl_K exceeds 1, the situation is even simpler.

Theorem 6.2. *Let K be an imaginary quadratic field for which the 2-class group is noncyclic. Then the map ϕ_2 in Theorem 5.1 is not injective.*

Proof. As every 2-torsion element in Cl_K is the class of a ramified prime $\mathfrak{p}$, its square can be generated by a rational prime number. This implies that the image

of ϕ_2 is contained in the cyclic subgroup

$$\mathbb{Z}_2^*/\{\pm 1\}(\mathbb{Z}_2^*)^2 \subset \widehat{\mathbb{O}}^*/T_2(\widehat{\mathbb{O}}^*)^2$$

of order 2. Thus ϕ_2 is not injective if Cl_K has noncyclic 2-part. □

In view of Theorem 4.4 and the remark following it, imaginary quadratic fields K for which A_K is the minimal Galois group from Theorem 4.1 can only be found among those K for which $-D_K$ is prime, or in the infinite families from Theorem 6.1. In the next section, we will find many of such K.

7. Computational results

In Onabe's paper [10], only cyclic class groups Cl_K of prime order $p \leq 7$ are considered. In this case there are just 2 types of splitting behavior for the extension (8), and Onabe provides a list of the first few K with $h_K = p \leq 7$, together with the type of splitting they represent. For $h_K = 2$ the list is in accordance with Theorem 6.1. In the cases $h_k = 3$ and $h_K = 5$ there are only 2 split examples against 10 and 7 nonsplit examples, and for $h_K = 7$ no nonsplit examples are found. This suggests that ϕ_p is rather likely to be injective for increasing values of $h_K = p$.

This belief is confirmed if we extend Onabe's list by including *all* imaginary quadratic K of odd prime class number $h_K = p < 100$. By the work of Watkins [11], we now know, much more precisely than Onabe did, what the exact list of fields with given small class number looks like. The extended list, with the 65 out of 2356 cases in which the extension (8) splits mentioned explicitly, is given in Table 1.

As the nonsplit types give rise to fields K having the minimal group G as its absolute Galois group, one is inevitably led to the following conjecture.

Conjecture 7.1. *There are infinitely many imaginary quadratic fields K for which the absolute abelian Galois group is isomorphic to*

$$G = \widehat{\mathbb{Z}}^2 \times \prod_{n \geq 1} \mathbb{Z}/n\mathbb{Z}.$$

The numerical evidence may be strong, but we do not even have a theorem that there are infinitely many prime numbers that occur as the class number of an imaginary quadratic field. And even if we had, we have no theorem telling us what the distribution between split and nonsplit will be.

From Table 1, one easily gets the impression that among all K with $h_K = p$, the fraction for which the sequence (8) splits is about $1/p$. In particular, assuming infinitely many imaginary quadratic fields to have prime class number, we would expect 100% of these fields to have the minimal absolute abelian Galois group G.

If we fix the class number $h_K = p$, the list of K will be finite, making it impossible to study the average distribution of the splitting behavior over $\mathrm{Cl}_K[p]$. For

p	$\#\{K : h_K = p\}$	#Nonsplit	$-D_K$ for split K
2	18	8	35, 51, 91, 115, 123, 187, 235, 267, 403, 427
3	16	13	107, 331, 643
5	25	19	347, 443, 739, 1051, 1123, 1723
7	31	27	859, 1163, 2707, 5107
11	41	36	9403, 5179, 2027, 10987, 13267
13	37	34	1667, 2963, 11923
17	45	41	383, 8539, 16699, 25243
19	47	43	4327, 17299, 17539, 17683
23	68	65	2411, 9587, 21163
29	83	80	47563, 74827, 110947
31	73	70	9203, 12923, 46867
37	85	83	20011, 28283
41	109	106	14887, 21487, 96763
43	106	105	42683
47	107	107	—
53	114	114	—
59	128	126	125731, 166363
61	132	131	101483
67	120	119	652723
71	150	150	—
73	119	117	358747, 597403
79	175	174	64303
83	150	150	—
89	192	189	48779, 165587, 348883
97	185	184	130051

Table 1. Splitting types for fields K with $h_K = p < 100$. The second column gives the number of imaginary quadratic fields with class number p; the third column gives the number of such fields for which the sequence (8) does not split; and the fourth column gives $-D_K$ for the fields K for which (8) splits.

this reason, we computed the average splitting behavior over $\mathrm{Cl}_K[p]$ for the set S_p of imaginary quadratic fields K for which the class number has a *single* factor p.

More precisely, Table 2 lists, for the first N_p imaginary quadratic fields $K \in S_p$ of absolute discriminant $|D_K| > B_p$, the fraction f_p of K for which the sequence (8) is split over $\mathrm{Cl}_K[p]$. We started counting for absolute discriminants exceeding B_p to avoid the influence that using many very small discriminants may have on observing the asymptotic behavior. Numerically, the values for $p \cdot f_p \approx 1$ in the table show that the fraction f_p is indeed close to $1/p$.

For the first three odd primes, we also looked at the distribution of the splitting over the three kinds of local behavior in K of the prime p (split, inert or ramified)

p	N_p	B_p	$p \cdot f_p$	p	N_p	B_p	$p \cdot f_p$
3	300	10^7	0.960	43	2150	10^6	1.080
5	500	10^7	0.930	47	470	10^7	0.900
7	700	10^7	0.960	53	530	10^5	1.000
11	1100	10^7	0.990	59	590	10^6	0.900
13	1300	10^7	1.070	61	1830	10^5	0.933
17	1700	10^7	0.920	67	670	10^6	0.900
19	1900	10^7	1.000	71	1000	10^5	1.136
23	2300	10^7	1.030	73	3650	10^5	0.900
29	2900	10^6	1.000	79	1399	10^7	1.130
31	3100	10^6	0.970	83	1660	10^6	1.000
37	3700	10^6	0.930	89	890	10^5	1.100
41	4100	10^6	1.060	97	970	10^8	1.100

Table 2. Splitting fractions at p for h_K divisible by $p < 100$. For the given values of p, N_p, and B_p, we consider the first N_p imaginary quadratic fields K with $|D_K| > B_p$ and whose class numbers are divisible by a single factor of p. The fourth column gives the value of $p \cdot f_p$, where f_p is the fraction of these fields for which the sequence (8) is split over $\mathrm{Cl}_K[p]$.

and concluded that, at least numerically, there is no clearly visible influence; see Table 3.

p	N_p	B_p	$p \cdot f_p$	Split	Inert	Ramified
3	300	10^7	0.960	0.925	0.947	1.025
5	500	10^7	0.930	0.833	0.990	1.022
7	700	10^7	0.960	0.972	0.963	0.897

Table 3. Splitting fractions at p according to local behavior at p. The first four columns are as in Table 2. The remaining columns give the values of p times the quantity analogous to f_p, where we further limit our attention to fields in which p has the prescribed splitting behavior.

We further did a few computations that confirmed the natural hypothesis that the splitting behaviors at different primes p and q that both divide the class number once are independent of each other.

Acknowledgement

We thank Georges Gras for spotting an inaccuracy in part (2) of Theorem 6.1, and for pointing out related results in his textbook on class field theory [2].

References

[1] Emil Artin and John Tate, *Class field theory*, AMS Chelsea, Providence, RI, 2009, reprinted with corrections from the 1967 original. MR 2009k:11001

[2] Georges Gras, *Class field theory*: *from theory to practice*, Springer, Berlin, 2003. MR 2003j:11138

[3] Helmut Hasse, *Number theory*, Grundlehren der mathematischen Wissenschaften, no. 229, Springer, Berlin, 1980, reprint of the 1980 edition. MR 81c:12001b

[4] Irving Kaplansky, *Infinite abelian groups*, University of Michigan Press, Ann Arbor, 1954. MR 16,444g

[5] Tomio Kubota, *Galois group of the maximal abelian extension over an algebraic number field*, Nagoya Math. J. **12** (1957), 177–189. MR 20 #4539

[6] Saunders Mac Lane, *Homology*, Grundlehren der mathematischen Wissenschaften, no. 114, Springer, Berlin, 1975. MR 96d:18001

[7] Jürgen Neukirch, *Kennzeichnung der p-adischen und der endlichen algebraischen Zahlkörper*, Invent. Math. **6** (1969), 296–314. MR 39 #5528

[8] ———, *Über die absoluten Galoisgruppen algebraischer Zahlkörper*, Journées Arithmétiques de Caen (Caen, 1976), Astérisque, no. 41-42, Soc. math. de France, Paris, 1977, pp. 67–79. MR 57 #5954

[9] Jürgen Neukirch, Alexander Schmidt, and Kay Wingberg, *Cohomology of number fields*, 2nd ed., Grundlehren der mathematischen Wissenschaften, no. 323, Springer, Berlin, 2008. MR 2008m:11223

[10] Midori Onabe, *On the isomorphisms of the Galois groups of the maximal abelian extensions of imaginary quadratic fields*, Natur. Sci. Rep. Ochanomizu Univ. **27** (1976), no. 2, 155–161. MR 55 #7999

[11] Mark Watkins, *Class numbers of imaginary quadratic fields*, Math. Comp. **73** (2004), no. 246, 907–938. MR 2005a:11175

ATHANASIOS ANGELAKIS: aangelakis@math.leidenuniv.nl
Mathematisch Instituut, Universiteit Leiden, Postbus 9512, 2300 RA Leiden, The Netherlands

PETER STEVENHAGEN: psh@math.leidenuniv.nl
Mathematisch Instituut, Universiteit Leiden, Postbus 9512, 2300 RA Leiden, The Netherlands

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.41

msp

Iterated Coleman integration for hyperelliptic curves

Jennifer S. Balakrishnan

The Coleman integral is a p-adic line integral. Double Coleman integrals on elliptic curves appear in Kim's nonabelian Chabauty method, the first numerical examples of which were given by the author, Kedlaya, and Kim. This paper describes the algorithms used to produce those examples, as well as techniques to compute higher iterated integrals on hyperelliptic curves, building on previous joint work with Bradshaw and Kedlaya.

1. Introduction

In a series of papers in the 1980s, Coleman gave a p-adic theory of integration on the projective line [8], then on curves and abelian varieties [9; 7]. This integration theory relies on locally defined antiderivatives that are extended analytically by the principle of Frobenius equivariance. In joint work with Bradshaw and Kedlaya [1], we made this construction explicit and gave algorithms to compute single Coleman integrals for hyperelliptic curves.

Having algorithms to compute Coleman integrals allows one to compute p-adic regulators in K-theory [8; 7], carry out the method of Chabauty-Coleman for finding rational points on higher genus curves [15], and utilize Kim's nonabelian analogue of the Chabauty method [14].

Kim's method, in the case of rank-1 elliptic curves, allows one to find integral points via the computation of double Coleman integrals. Indeed, Coleman's theory of integration is not limited to single integrals; it gives rise to an entire class of

MSC2010: primary 11S80; secondary 11Y35, 11Y50.
Keywords: Coleman integration, p-adic integration, iterated Coleman integration, hyperelliptic curves, nonabelian Chabauty, integral points.

locally analytic functions, the *Coleman functions*, on which antidifferentiation is well-defined. In other words, one can define iterated p-adic integrals [4; 8]

$$\int_P^Q \xi_n \cdots \xi_1$$

which behave formally like iterated path integrals

$$\int_0^1 \int_0^{t_1} \cdots \int_0^{t_{n-1}} f_n(t_n) \cdots f_1(t_1)\, dt_n \cdots dt_1.$$

Let us fix some notation. Let C be a genus-g hyperelliptic curve over an unramified extension K of $\mathbb{Q}_p$ having good reduction. Let $k = \mathbb{F}_q$ denote its residue field, where $q = p^m$. We will assume that C is given by a model of the form $y^2 = f(x)$, where f is a monic separable polynomial with $\deg f = 2g + 1$.

Our methods for computing iterated integrals are similar in spirit to those detailed in [1]. We begin with algorithms for tiny iterated integrals, use Frobenius equivariance to write down a linear system yielding the values of integrals between points in different residue disks, and, if needed, use basic properties of integration to correct endpoints. We begin with some basic properties of iterated path integrals.

2. Iterated path integrals

We follow the convention of Kim [14] and define our integrals as follows:

$$\int_P^Q \xi_1 \xi_2 \cdots \xi_{n-1} \xi_n := \int_P^Q \xi_1(R_1) \int_P^{R_1} \xi_2(R_2) \cdots \int_P^{R_{n-2}} \xi_{n-1}(R_{n-1}) \int_P^{R_{n-1}} \xi_n,$$

for a collection of dummy parameters $R_1, \ldots, R_{n-1}$ and 1-forms $\xi_1, \ldots, \xi_n$.

We begin by recalling some key formal properties satisfied by iterated path integrals [6].

Proposition 2.1. *Let $\xi_1, \ldots, \xi_n$ be 1-forms, holomorphic at points P, Q on C. Then*:

(1) $\int_P^P \xi_1 \xi_2 \cdots \xi_n = 0$,

(2) $\sum_{\text{all permutations } \sigma} \int_P^Q \omega_{\sigma(i_1)} \omega_{\sigma(i_2)} \cdots \omega_{\sigma(i_n)} = \prod_{j=1}^n \int_P^Q \omega_{i_j}$,

(3) $\int_P^Q \omega_{i_1} \cdots \omega_{i_n} = (-1)^n \int_Q^P \omega_{i_n} \cdots \omega_{i_1}$.

As an easy corollary of Proposition 2.1(2), we have:

Corollary 2.2. *For a 1-form ω_i and points P, Q as before,*

$$\int_P^Q \omega_i\, \omega_i \cdots \omega_i = \frac{1}{n!} \left(\int_P^Q \omega_i \right)^n.$$

When possible, we will use this to write an iterated integral in terms of a single integral.

3. p-adic cohomology

We briefly recall some p-adic cohomology from [12], necessary for formulating the integration algorithms.

Let C' be the affine curve obtained by deleting the Weierstrass points from C, and let $A = K[x, y, z]/(y^2 - f(x), yz - 1)$ be the coordinate ring of C'. Let $A^\dagger$ denote the Monsky-Washnitzer weak completion of A; it is the ring consisting of infinite sums of the form

$$\sum_{i=-\infty}^{\infty} \frac{B_i(x)}{y^i}, \quad B_i(x) \in K[x], \quad \deg B_i \leq 2g,$$

further subject to the condition that $v_p(B_i(x))$ grows faster than a linear function of i as $i \to \pm\infty$. We make a ring out of these using the relation $y^2 = f(x)$.

These functions are holomorphic on the space over which we integrate, so we consider odd 1-forms written as

$$\omega = g(x, y)\,\frac{dx}{2y}, \quad g(x, y) \in A^\dagger.$$

Any such differential can be written as

$$\omega = dF + c_0\omega_0 + \cdots + c_{2g-1}\omega_{2g-1}, \tag{1}$$

with $F \in A^\dagger, c_i \in K$, and

$$\omega_i = x^i\,\frac{dx}{2y} \quad (i = 0, \ldots, 2g-1).$$

Namely, the set of differentials $\{\omega_i\}_{i=0}^{2g-1}$ forms a basis of the odd part of the de Rham cohomology of $A^\dagger$, which we denote as $H^1_{dR}(C')^-$.

One computes the p-power Frobenius action ϕ^* on $H^1_{dR}(C')^-$ as follows:

- Let ϕ_K denote the unique automorphism lifting Frobenius from $\mathbb{F}_q$ to K. Extend ϕ_K to $A^\dagger$ by setting

$$\begin{aligned}
\phi(x) &= x^p, \\
\phi(y) &= y^p\left(1 + \frac{\phi(f)(x^p) - f(x)^p}{f(x)^p}\right)^{\frac{1}{2}} \\
&= y^p \sum_{i=0}^{\infty} \binom{\frac{1}{2}}{i} \frac{(\phi(f)(x^p) - f(x)^p)^i}{y^{2pi}}.
\end{aligned}$$

- Use the relations

$$y^2 = f(x),$$
$$d(x^i y^j) = \left(2ix^{i-1}y^{j+1} + jx^i f'(x)y^{j-1}\right)\frac{dx}{2y}$$

to reduce large powers of x and large (in absolute value) powers of y to write $\phi^*(\omega)$ in the form (1).

This reduction process is known as *Kedlaya's algorithm* [12], and we will repeatedly use this algorithm to reduce iterated integrals involving $\omega \in A^\dagger \frac{dx}{2y}$ to iterated integrals in terms of basis elements ω_i.

4. Integrals: lemmas

Recall that we use Kedlaya's algorithm to compute single Coleman integrals as follows:

Algorithm 4.1 (Coleman integration in non-Weierstrass disks [1]).

Input: The basis differentials $(\omega_i)_{i=0}^{2g-1}$, points $P, Q \in C(\mathbb{C}_p)$ in non-Weierstrass residue disks, and a positive integer m such that the residue fields of P, Q are contained in $\mathbb{F}_{p^m}$.

Output: The integrals $\left(\int_P^Q \omega_i\right)_{i=0}^{2g-1}$.

1. Calculate the action of the m-th power of Frobenius on each basis element (see Remark 4.2):
$$(\phi^m)^*\omega_i = dh_i + \sum_{j=0}^{2g-1} M_{ij}\omega_j.$$
2. By a change of variables, we obtain
$$\sum_{j=0}^{2g-1}(M-I)_{ij}\int_P^Q \omega_j = h_i(P) - h_i(Q) - \int_P^{\phi^m(P)}\omega_i - \int_{\phi^m(Q)}^{Q}\omega_i \qquad (2)$$
(the *fundamental linear system*). Since the eigenvalues of the matrix M are algebraic integers of $\mathbb{C}$-norm $p^{m/2} \neq 1$ (see [12, §2]), the matrix $M - I$ is invertible, and we may solve (2) to obtain the integrals $\int_P^Q \omega_i$.

Remark 4.2. To compute the action of ϕ^m, first carry out Kedlaya's algorithm to write
$$\phi^*\omega_i = dg_i + \sum_{j=0}^{2g-1} B_{ij}\omega_j.$$

If we view h, g as column vectors and M, B as matrices, induction on m shows that

$$h = \phi^{m-1}(g) + B\phi^{m-2}(g) + \cdots + B\phi_K(B)\cdots\phi_K^{m-2}(B)g,$$
$$M = B\phi_K(B)\cdots\phi_K^{m-1}(B).$$

Note, however, that when points $P, Q \in C(\mathbb{C}_p)$ are in the same residue disk, the "tiny" Coleman integral between them can be computed using a local parametrization, just as in the case of a real-valued line integral. This is also true when the integrals are iterated (see Section 5).

However, to compute general iterated integrals, we will need to employ the analogue of "additivity in endpoints" to link integrals between different residue disks. First, let us consider the case where we are breaking up the path by one point.

Lemma 4.3. *Let P, P', Q be points on C such that a path is to be taken from P to Q via P'. Let $\xi_1, \ldots, \xi_n$ be a collection of 1-forms holomorphic at the points P, P', Q. Then*

$$\int_P^Q \xi_1 \cdots \xi_n = \sum_{i=0}^{n} \int_{P'}^{Q} \xi_1 \cdots \xi_i \int_P^{P'} \xi_{i+1} \cdots \xi_n.$$

Proof. We proceed by induction. The case $n = 1$ is clear. Let us suppose the statement holds for $n = k$. Then

$$\begin{aligned}\int_P^Q \xi_1 \cdots \xi_{k+1} &= \left(\int_P^Q \xi_1 \cdots \xi_k\right)(R) \int_P^R \xi_{k+1} \\ &= \left(\sum_{i=0}^{k} \int_{P'}^{Q} \xi_1 \cdots \xi_i \int_P^{P'} \xi_{i+1} \cdots \xi_k\right)(R) \int_P^R \xi_{k+1}.\end{aligned}$$

Observe that the summand with $i = k$ can be rewritten as

$$\left(\int_{P'}^Q \xi_1 \cdots \xi_k\right)(R) \int_P^R \xi_{k+1} = \left(\int_{P'}^Q \xi_1 \cdots \xi_k\right)(R) \left(\int_P^{P'} \xi_{k+1} + \int_{P'}^R \xi_{k+1}\right),$$

and that further, the terms with $i < k$ give us

$$\sum_{i=0}^{k-1} \int_{P'}^{Q} \xi_1 \cdots \xi_i \int_P^{P'} \xi_{i+1} \cdots \xi_{k+1}.$$

Thus we have

$$\begin{aligned}\int_P^Q \xi_1\cdots\xi_{k+1} &= \sum_{i=0}^{k-1}\int_{P'}^{Q}\xi_1\cdots\xi_i\int_P^{P'}\xi_{i+1}\cdots\xi_{k+1}\\ &\qquad + \Big(\int_{P'}^{Q}\xi_1\cdots\xi_k\Big)\Big(\int_P^{P'}\xi_{k+1}\Big) + \int_{P'}^{Q}\xi_1\cdots\xi_{k+1}\\ &= \sum_{i=0}^{k+1}\int_{P'}^{Q}\xi_1\cdots\xi_i\int_P^{P'}\xi_{i+1}\cdots\xi_{k+1},\end{aligned}$$

as desired. □

Applying Lemma 4.3 twice, we obtain a link between different residue disks:

Lemma 4.4 (Link lemma). *Let points P, P', Q', Q be on C such that a path is to be taken from P to P' to Q' to Q. Let $\xi_1,\dots,\xi_n$ be a collection of 1-forms holomorphic at the points P, P', Q, Q'. Then*

$$\int_P^Q \xi_1\cdots\xi_n = \sum_{i=0}^{n}\int_{Q'}^{Q}\xi_1\cdots\xi_i\Big(\sum_{j=i}^{n}\int_{P'}^{Q'}\xi_{i+1}\cdots\xi_j\int_P^{P'}\xi_{j+1}\cdots\xi_n\Big).$$

Below we record a specific case of the link lemma, which we shall use throughout this paper.

Example 4.5 (Link lemma for double integrals). Suppose we have two differentials ξ_0, ξ_1. Then

$$\int_P^Q \xi_0\xi_1 = \int_P^{P'}\xi_0\xi_1 + \int_{P'}^{Q'}\xi_0\xi_1 + \int_{Q'}^{Q}\xi_0\xi_1 + \int_P^{P'}\xi_1\int_{P'}^{Q}\xi_0 + \int_{P'}^{Q'}\xi_1\int_{Q'}^{Q}\xi_0.$$

5. Tiny iterated integrals

We begin with an algorithm to compute tiny iterated integrals.

Algorithm 5.1 (Tiny iterated integrals).

Input: Points $P, Q \in C(\mathbb{C}_p)$ in the same residue disk (neither equal to the point at infinity) and differentials $\xi_1,\dots,\xi_n$ without poles in the disk of P.

Output: The integral $\int_P^Q \xi_1\xi_2\cdots\xi_n$.

1. Compute a parametrization $\big(x(t), y(t)\big)$ at P in terms of a local coordinate t.
2. For each k, write $\xi_k(x, y)$ in terms of t: $\xi_k(t) := \xi_k\big(x(t), y(t)\big)$.
3. Let $I_{n+1}(t) := 1$.

4. Compute, for $k = n, \dots, 2$, in descending order,

$$I_k(t) = \int_P^{R_{k-1}} \xi_k I_{k+1} = \int_0^{t(R_{k-1})} \xi_k(u) I_{k+1}(u),$$

with R_{k-1} in the disk of P.

5. Upon computing $I_2(t)$, we arrive at the desired integral:

$$\int_P^Q \xi_1 \xi_2 \cdots \xi_n = I_1(t) = \int_0^{t(Q)} \xi_1(u) I_2(u).$$

We show how we carry out Algorithm 5.1 for double integrals on an elliptic curve.

Example 5.2 (A tiny double integral). Let C be the elliptic curve

$$y^2 = x(x-1)(x+9),$$

let $p = 7$, and consider the points $P = (9, 36)$, $Q = \phi(P)$, and

$$R = \big(a + x(P), \sqrt{f(a + x(P))}\big),$$

so that R is in the same disk as P and Q. Furthermore, let $\omega_0 = \frac{dx}{2y}$ and $\omega_1 = \frac{x\,dx}{2y}$. We compute the double integral $\int_P^Q \omega_0 \omega_1$.

First compute the local coordinates at P:

$$\begin{aligned} x(t) &= 9 + t + O(t^{20}) \\ y(t) &= 36 + \frac{21}{4}t + \frac{119}{1152}t^2 - \frac{65}{55296}t^3 + \frac{2219}{95551488}t^4 - \frac{7}{509607936}t^5 + O(t^6). \end{aligned}$$

Then setting $I_2 := \int x \frac{dx}{2y}$, and making it a definite integral, we have

$$\begin{aligned} I_2|_P^R &= \int_P^R x \frac{dx}{2y} \\ &= \int_0^a x(t) \frac{dx(t)}{2y(t)} \\ &= \frac{1}{8}a - \frac{5}{2304}a^2 + \frac{91}{995328}a^3 - \frac{1121}{191102976}a^4 + \frac{22129}{45864714240}a^5 \\ &\qquad - \frac{360185}{7925422620672}a^6 + \frac{36737231}{7988826001637376}a^7 + O(a^8), \end{aligned}$$

from which we arrive at

$$\begin{aligned} I &= \int_0^{x(Q)-x(P)} I_2(a) \frac{dx(R(a))}{2y(R(a))} \\ &= 4 \cdot 7^2 + 5 \cdot 7^3 + 2 \cdot 7^5 + 4 \cdot 7^6 + 2 \cdot 7^7 + O(7^8). \end{aligned}$$

6. Iterated integrals: linear system

As in the case of computing single integrals, to compute general iterated Coleman integrals, we use Kedlaya's algorithm to calculate the action of Frobenius on de Rham cohomology. This gives us a linear system that allows us to solve for all $(2g)^n$ n-fold iterated integrals on basis differentials.

Theorem 6.1. *Let $P, Q \in C(\mathbb{C}_p)$ be non-Weierstrass points such that the residue fields of P, Q are contained in $\mathbb{F}_{p^m}$. Let M be the matrix of the action of the m-th power of Frobenius on the basis differentials $\omega_0, \dots, \omega_{2g-1}$. For constants $c_{i_0,\dots,i_{n-1}}$ computable in terms of $(n-1)$-fold iterated integrals and n-fold tiny iterated integrals, the n-fold iterated Coleman integrals on basis differentials between P, Q can be computed via a linear system of the form*

$$\begin{pmatrix} \vdots \\ \int_P^Q \omega_{i_0} \cdots \omega_{i_{n-1}} \\ \vdots \end{pmatrix} = \left(I_{(2g)^n \times (2g)^n} - (M^t)^{\otimes n}\right)^{-1} \begin{pmatrix} \vdots \\ c_{i_0 \cdots i_{n-1}} \\ \vdots \end{pmatrix}.$$

Proof. By the link lemma (Lemma 4.4), we can reduce to the case where both P and Q are Teichmüller points (points fixed by some power of ϕ). Then we have

$$\begin{aligned} \int_P^Q \omega_{i_i} \cdots \omega_{i_n} &= \int_{\phi^m(P)}^{\phi^m(Q)} \omega_{i_i} \cdots \omega_{i_n} \\ &= \int_P^Q (\phi^m)^*(\omega_{i_i} \cdots \omega_{i_n}) \\ &= \int_P^Q (\phi^m)^*(\omega_{i_i}) \cdots (\phi^m)^*(\omega_{i_n}). \end{aligned} \tag{3}$$

Recall that given $\omega_0, \dots, \omega_{2g-1}$ a basis for $H^1_{dR}(C')^-$, we have

$$(\phi^m)^* \omega_{i_\ell} = df_{i_\ell} + \sum_{j=0}^{2g-1} M_{i_\ell j} \omega_j.$$

Substituting this expression in for each factor of (3) and expanding yields the linear system. □

To illustrate our methods, in the next section, we present a more explicit version of this theorem, accompanied by algorithms, in the case of double integrals. We show how these are used in Kim's nonabelian Chabauty method in Section 8.

7. Explicit double integrals

7A. ***The linear system for double integrals between Teichmüller points.*** In this subsection, we make explicit one aspect of Theorem 6.1: We give an algorithm to compute double integrals between Teichmüller points.

Algorithm 7.1 (Double Coleman integration between Teichmüller points).

Input: The basis differentials $(\omega_i)_{i=0}^{2g-1}$, Teichmüller points $P, Q \in C(\mathbb{C}_p)$ in non-Weierstrass residue disks, and a positive integer m such that the residue fields of P, Q are contained in $\mathbb{F}_{p^m}$.

Output: The double integrals $\big(\int_P^Q \omega_i\,\omega_j\big)_{i,j=0}^{2g-1}$.

1. Calculate the action of the m-th power of Frobenius on each basis element:
$$(\phi^m)^*\omega_i = df_i + \sum_{j=0}^{2g-1} M_{ij}\omega_j.$$
2. Use Algorithm 4.1 to compute the single Coleman integrals $\int_P^Q \omega_j$ on all basis differentials.
3. Use Step 2 and linearity to recover the other single Coleman integrals:
$$\int_P^Q df_i\, f_k, \int_P^Q \sum_{j=0}^{2g-1} M_{ij}\omega_j\, f_k$$
for each i, k.
4. Use the results of the above two steps to write down, for each i, k, the constant
$$\begin{aligned} c_{ik} = \int_P^Q df_i(R)(f_k(R)) - f_k(P)\big(f_i(Q) - f_i(P)\big) \\ + \int_P^Q \sum_{j=0}^{2g-1} M_{ij}\omega_j(R)\big(f_k(R) - f_k(P)\big) \\ + f_i(Q)\int_P^Q \sum_{j=0}^{2g-1} M_{kj}\omega_j - \int_P^Q f_i(R)\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j(R)\Big). \end{aligned}$$
5. Recover the double integrals (see Remark 7.2 below) via the linear system
$$\begin{pmatrix} \int_P^Q \omega_0\omega_0 \\ \int_P^Q \omega_0\omega_1 \\ \vdots \\ \int_P^Q \omega_{2g-1}\omega_{2g-1} \end{pmatrix} = (I_{4g^2\times 4g^2} - (M^t)^{\otimes 2})^{-1} \begin{pmatrix} c_{00} \\ c_{01} \\ \vdots \\ c_{2g-1,2g-1} \end{pmatrix}.$$

Remark 7.2. We obtain the linear system in the following manner. Since P, Q are Teichmüller, we have

$$\int_P^Q \omega_i\,\omega_k = \int_{\phi^m(P)}^{\phi^m(Q)} \omega_i\,\omega_k = \int_P^Q (\phi^m)^*(\omega_i\,\omega_k). \tag{4}$$

We begin by expanding the right side of (4).

Recall that given $\omega_0, \ldots, \omega_{2g-1}$ a basis for $H^1_{dR}(C')^-$, we have

$$(\phi^m)^*\omega_i = df_i + \sum_{j=0}^{2g-1} M_{ij}\omega_j.$$

Thus we have

$$\begin{aligned}
&\int_P^Q (\phi^m)^*(\omega_i\,\omega_k)\\
&= \int_P^Q (\phi^m)^*(\omega_i)(\phi^m)^*(\omega_k)\\
&= \int_P^Q \left(df_i + \sum_{j=0}^{2g-1} M_{ij}\omega_j\right)\left(df_k + \sum_{j=0}^{2g-1} M_{kj}\omega_j\right)\\
&= \int_P^Q df_i\,df_k + \left(\sum_{j=0}^{2g-1} M_{ij}\omega_j\right)df_k + df_i \sum_{j=0}^{2g-1} M_{kj}\omega_j + \sum_{j=0}^{2g-1} M_{ij}\omega_j \sum_{j=0}^{2g-1} M_{kj}\omega_j.
\end{aligned}$$

We expand the first three quantities separately. First, we have

$$\begin{aligned}
\int_P^Q df_i\,df_k &= \int_P^Q df_i(R) \int_P^R df_k\\
&= \int_P^Q df_i(R)\big(f_k(R) - f_k(P)\big)\\
&= \int_P^Q df_i(R)(f_k(R)) - f_k(P)\int_P^Q df_i(R)\\
&= \int_P^Q df_i(R)(f_k(R)) - f_k(P)\big(f_i(Q) - f_i(P)\big).
\end{aligned}$$

Next, we have

$$\begin{aligned}
\int_P^Q \left(\sum_{j=0}^{2g-1} M_{ij}\omega_j\right)df_k &= \int_P^Q \sum_{j=0}^{2g-1} M_{ij}\omega_j(R) \int_P^R df_k\\
&= \int_P^Q \sum_{j=0}^{2g-1} M_{ij}\omega_j(R)\big(f_k(R) - f_k(P)\big).
\end{aligned}$$

The third term (via integration by parts) is

$$\begin{aligned}\int_P^Q df_i\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j\Big)&\\ &= \int_P^Q df_i(R)\int_P^R\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j\Big)\\ &= f_i(R)\int_P^R\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j\Big)\Big|_{R=P}^{R=Q} - \int_P^Q f_i(R)\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j(R)\Big)\\ &= f_i(Q)\int_P^Q \sum_{j=0}^{2g-1} M_{kj}\omega_j - \int_P^Q f_i(R)\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j(R)\Big).\end{aligned}$$

Denote the sum of these terms by c_{ik}; in other words,

$$\begin{aligned}c_{ik} = \int_P^Q df_i(R)(f_k(R)) - f_k(P)\big(f_i(Q) - f_i(P)\big)&\\ + \int_P^Q \sum_{j=0}^{2g-1} M_{ij}\omega_j(R)\big(f_k(R) - f_k(P)\big)&\\ + f_i(Q)\int_P^Q \sum_{j=0}^{2g-1} M_{kj}\omega_j - \int_P^Q f_i(R)\Big(\sum_{j=0}^{2g-1} M_{kj}\omega_j(R)\Big).&\end{aligned}$$

Then rearranging terms, our linear system reads

$$\begin{pmatrix} \int_P^Q \omega_0\omega_0 \\ \int_P^Q \omega_0\omega_1 \\ \vdots \\ \int_P^Q \omega_{2g-1}\omega_{2g-1}\end{pmatrix} = \big(I_{4g^2\times 4g^2} - (M^t)^{\otimes 2}\big)^{-1}\begin{pmatrix} c_{00} \\ c_{01} \\ \vdots \\ c_{2g-1,2g-1}\end{pmatrix}.$$

7B. ***Linking double integrals.*** Let P' and Q' be in the disks of P and Q, respectively. Using the link lemma for double integrals (Example 4.5), we may link double integrals between different residue disks:

$$\begin{aligned}\int_P^Q \omega_i\,\omega_k&\\ &= \int_P^{P'} \omega_i\,\omega_k + \int_{P'}^{Q'} \omega_i\,\omega_k + \int_{Q'}^{Q} \omega_i\,\omega_k + \int_P^{P'} \omega_k \int_{P'}^{Q} \omega_i + \int_{P'}^{Q'} \omega_k \int_{Q'}^{Q} \omega_i.\end{aligned}$$

Algorithm 7.3 (Double Coleman integration using intermediary Teichmüller points).

Input: The basis differentials $(\omega_i)_{i=0}^{2g-1}$, points $P, Q \in C(\mathbb{C}_p)$ in non-Weierstrass residue disks.

Output: The double integrals

$$\left(\int_P^Q \omega_i\,\omega_j\right)_{i,j=0}^{2g-1}.$$

1. Compute Teichmüller points P', Q' in the disks of P, Q, respectively.
2. Use Algorithm 4.1 to compute the single integrals $\int_P^Q \omega_i, \int_{P'}^P \omega_i, \int_Q^{Q'} \omega_i$ for all i.
3. Use Algorithm 5.1 to compute the tiny double integrals $\int_{P'}^P \omega_i\,\omega_k, \int_{Q'}^Q \omega_i\,\omega_k$.
4. Use Algorithm 7.1 to compute the double integrals $\left\{\int_{P'}^{Q'} \omega_i\,\omega_j\right\}_{i,j=0}^{2g-1}$.
5. Correct endpoints using

$$\begin{aligned}\int_P^Q \omega_i\,\omega_k &= \int_P^{P'} \omega_i\,\omega_k + \int_{P'}^{Q'} \omega_i\,\omega_k + \int_{Q'}^{Q} \omega_i\,\omega_k + \int_P^{P'} \omega_k \int_{P'}^{Q} \omega_i + \int_{P'}^{Q'} \omega_k \int_{Q'}^{Q} \omega_i.\end{aligned}$$

7C.*Without Teichmüller points.* Alternatively, instead of finding Teichmüller points and correcting endpoints, we can directly compute double integrals using a slightly different linear system. Indeed, using the link lemma for double integrals, we take $\phi(P)$ and $\phi(Q)$ to be the points in the disks of P and Q, respectively, which gives

$$\begin{aligned}\int_P^Q \omega_i\,\omega_k = \int_P^{\phi(P)} \omega_i\,\omega_k + \int_{\phi(P)}^{\phi(Q)} \omega_i\,\omega_k + \int_{\phi(Q)}^{Q} \omega_i\,\omega_k \\ + \int_P^{\phi(P)} \omega_k \int_{\phi(P)}^{Q} \omega_i + \int_{\phi(P)}^{\phi(Q)} \omega_k \int_{\phi(Q)}^{Q} \omega_i. \quad (5)\end{aligned}$$

To write down a linear system without Teichmüller points, we begin as before, with

$$\int_{\phi(P)}^{\phi(Q)} \omega_i\,\omega_k = \int_P^Q \phi^*(\omega_i\,\omega_k) = c_{ik} + \int_P^Q \left(\sum_{j=0}^{2g-1} A_{ij}\omega_j\right)\left(\sum_{j=0}^{2g-1} A_{kj}\omega_j\right). \quad (6)$$

Putting together (5) and (6), we get

$$\begin{pmatrix} \vdots \\ \int_P^Q \omega_i\,\omega_k \\ \vdots \end{pmatrix} = \left(I_{4g^2\times 4g^2} - (M^t)^{\otimes 2}\right)^{-1} \cdot \begin{pmatrix} \vdots \\ c_{ik} - \int_{\phi(P)}^{P} \omega_i\,\omega_k - \left(\int_P^Q \omega_i\right)\left(\int_{\phi(P)}^{P} \omega_k\right) \\ \quad - \left(\int_Q^{\phi(Q)} \omega_i\right)\left(\int_{\phi(P)}^{\phi(Q)} \omega_k\right) + \int_{\phi(Q)}^{Q} \omega_i\,\omega_k \\ \vdots \end{pmatrix}. \quad (7)$$

This gives us the following alternative to Algorithm 7.1.

Algorithm 7.4 (Double Coleman integration).

Input: The basis differentials $(\omega_i)_{i=0}^{2g-1}$, points $P, Q \in C(\mathbb{Q}_p)$ in non-Weierstrass residue disks or in Weierstrass disks in the region of convergence.

Output: The double integrals $\left(\int_P^Q \omega_i\,\omega_j\right)_{i,j=0}^{2g-1}$.

1. Use Algorithm 4.1 to compute the single integrals $\int_P^Q \omega_i, \int_{\phi(P)}^{\phi(Q)} \omega_i$ for all i.
2. Use Algorithm 5.1 to compute $\int_{\phi(P)}^{P} \omega_i\,\omega_k, \int_{\phi(Q)}^{Q} \omega_i\,\omega_k$ for all i, k
3. As in Step 4 of Algorithm 7.1, compute the constants c_{ik} for all i, k.
4. Recover the double integrals using the linear system (7).

Example 7.5. Let C be the genus-2 curve $y^2 = x^5 - x^4 + x^3 + x^2 - 2x + 1$ and let $P = (1, -1)$, $Q = (-1, -1)$ and $p = 7$. We compute double integrals on basis differentials:

$$\begin{aligned}
\int_P^Q \omega_0\omega_0 &= 2\cdot 7^2 + 7^3 + 4\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_0\omega_1 &= 7^2 + 5\cdot 7^3 + 3\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_0\omega_2 &= 4\cdot 7 + 5\cdot 7^2 + 7^3 + O(7^4),\\
\int_P^Q \omega_0\omega_3 &= 7 + 5\cdot 7^2 + 3\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_1\omega_0 &= 7^2 + 6\cdot 7^3 + 5\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_1\omega_1 &= 4\cdot 7^2 + 3\cdot 7^3 + O(7^5),\\
\int_P^Q \omega_1\omega_2 &= 5\cdot 7 + 6\cdot 7^2 + 2\cdot 7^3 + 4\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_1\omega_3 &= 2 + 3\cdot 7 + 7^2 + 4\cdot 7^3 + O(7^4),\\
\int_P^Q \omega_2\omega_0 &= 7^2 + 4\cdot 7^3 + O(7^4),\\
\int_P^Q \omega_2\omega_1 &= 4\cdot 7 + 6\cdot 7^2 + 4\cdot 7^3 + 5\cdot 7^4 + O(7^5),\\
\int_P^Q \omega_2\omega_2 &= 2 + 5\cdot 7 + 3\cdot 7^2 + O(7^3),
\end{aligned}$$

$$\int_P^Q \omega_2\omega_3 = 5 + 2\cdot 7 + 3\cdot 7^2 + O(7^3),$$
$$\int_P^Q \omega_3\omega_0 = 3\cdot 7 + 2\cdot 7^2 + 5\cdot 7^3 + 5\cdot 7^4 + O(7^5),$$
$$\int_P^Q \omega_3\omega_1 = 5 + 5\cdot 7 + 7^2 + 6\cdot 7^3 + O(7^4),$$
$$\int_P^Q \omega_3\omega_2 = 6 + 7 + 5\cdot 7^2 + O(7^3),$$
$$\int_P^Q \omega_3\omega_3 = 2 + 6\cdot 7 + 5\cdot 7^2 + O(7^3).$$

Example 7.6. Using the previous example, we verify the Fubini identity

$$\int_P^Q \omega_j\,\omega_i + \int_P^Q \omega_i\,\omega_j = \left(\int_P^Q \omega_i\right)\left(\int_P^Q \omega_j\right).$$

We have

$$\int_P^Q \omega_0 = 5\cdot 7 + 2\cdot 7^2 + 5\cdot 7^3 + 7^4 + 4\cdot 7^5 + O(7^6),$$
$$\int_P^Q \omega_1 = 6\cdot 7 + 6\cdot 7^2 + 2\cdot 7^3 + 4\cdot 7^4 + 3\cdot 7^5 + O(7^6),$$
$$\int_P^Q \omega_2 = 5 + 5\cdot 7^3 + 6\cdot 7^4 + 2\cdot 7^5 + O(7^6),$$
$$\int_P^Q \omega_3 = 5 + 3\cdot 7 + 4\cdot 7^2 + 3\cdot 7^3 + 6\cdot 7^4 + 2\cdot 7^5 + O(7^6).$$

We see, for example, that

$$\int_P^Q \omega_0\omega_1 + \int_P^Q \omega_1\omega_0 = 2\cdot 7^2 + 4\cdot 7^3 + 2\cdot 7^4 + O(7^5) = \left(\int_P^Q \omega_0\right)\left(\int_P^Q \omega_1\right)$$
$$\int_P^Q \omega_2\omega_3 + \int_P^Q \omega_3\omega_2 = \qquad 4 + 4\cdot 7 + 7^2 + O(7^3) = \left(\int_P^Q \omega_2\right)\left(\int_P^Q \omega_3\right).$$

7D. ***Weierstrass points.*** Suppose one of P or Q is a finite Weierstrass point. Then directly using the linear system as above fails, since the f_i have essential singularities at finite Weierstrass points. We remedy this as follows:

Proposition 7.7. *Let Q be a non-Weierstrass point, P a finite Weierstrass point, and S be a point in the residue disk of P, near the boundary. Then the integral from P to Q can be computed as a sum of integrals*:

$$\int_P^Q \omega_i\,\omega_k = \int_P^S \omega_i\,\omega_k + \int_S^Q \omega_i\,\omega_k + \int_P^S \omega_k \int_S^Q \omega_i\,.$$

Proof. This follows from Lemma 4.3 in the case of $n = 2$, where $P' = S$. □

To compute tiny iterated integrals in a Weierstrass disk, we modify Algorithm 5.1 slightly:

Algorithm 7.8 (Tiny iterated integral in a Weierstrass disk).

Input: A Weierstrass point P, the degree d of a totally ramified extension, and basis differentials ω_i, ω_j.

Output: The integral

$$\int_P^S \omega_i \omega_j = \int_P^S \omega_i(R) \int_P^R \omega_j = \int_{t=0}^{t=1} \omega_i(R) \int_{u=0}^{u=t} \omega_j.$$

1. Compute local coordinates $(x(u), u)$ at P.
2. Let $a = p^{1/d}$. Rescale coordinates so that $y := au, x := x(au)$.
3. Compute $I_2(u) = \int x^j \frac{dx}{2y}$ as a power series in u.
4. Compute the appropriate definite integral using the step above:

$$\int_R^S x^j \frac{dx}{2y} = \int_0^t x(au) \frac{a\,du}{u} = I_2(t)$$

   (where $R = (x(t), t)$). Call this definite integral (now a power series in t) I_2.
5. Now since $R = (x(t), t)$, we have $\int_P^S \omega_i \omega_j = \int_0^1 x(t)^i I_2 \frac{dx(t)}{2t}$.

Suppose P is a finite Weierstrass point. While one could compute the integral $\int_P^Q \omega_i \omega_j$ directly using Algorithm 7.4 for all of the tiny double integrals (and Algorithm 7.8 for the other double integrals), in practice, that approach is expensive, as it requires the computation of several intermediate integrals with Frobenius of points that are defined over ramified extensions. This, in turn, makes the requisite degree d extension for convergence quite large.

Instead, the key idea is to compute a local parametrization at the finite Weierstrass point P and to use this to compute the indefinite integral $\int_P^* \omega_i$. Then to compute integrals involving "boundary points," one can simply evaluate this indefinite integral at the appropriate points, instead of directly computing parametrizations, and thus integrals, over a totally ramified extension of $\mathbb{Q}_p$. This idea is also used to evaluate double integrals involving boundary points.

Algorithm 7.9 (Intermediary integrals for double integrals with a Weierstrass endpoint).

Input: A finite Weierstrass point P, a non-Weierstrass point Q, the degree d of a totally ramified extension, the desired precision n of $\mathbb{Q}_p$, and basis differentials ω_i, ω_j.

Output: Necessary things for the eventual computation of $\int_P^Q \omega_i \omega_j$.

1. Compute $(x(t), t)$ local coordinates at P to precision nd.
2. Let $S = (x(a), a)$, where $a = p^{1/d}$.
3. Compute as a power series in t, $I_2(t) = \int x(t)^i \frac{dx(t)}{y(t)}$.

4. Compute the definite integral $\int_P^S \omega_i = I_2(a)$.
5. For all $i < j$, compute the definite integral $\int_P^S \omega_i \omega_j$ via Algorithm 5.1. Keep the intermediary indefinite integral.
6. For all $i = j$, use the fact that $\int_P^S \omega_i \omega_j = \frac{1}{2}\left(\int_P^S \omega_i\right)^2$ to compute the double integral in terms of the single integral.
7. For all $i > j$, use the fact that $\int_P^S \omega_i \omega_j = -\int_P^S \omega_j \omega_i + \int_P^S \omega_i \int_P^S \omega_j$ to compute $\int_P^S \omega_i \omega_j$ (instead of directly computing it as a double integral).
8. Compute $\int_S^{\phi(S)} \omega_i = \int_P^{\phi(S)} \omega_i - \int_P^S \omega_i$ by the indefinite integral in Step 3. Use this to deduce $\int_S^{\phi(S)} \omega_i \omega_j$ for $i = j$.
9. Use the indefinite integral in Step 5 to get $\int_S^{\phi(S)} \omega_i \omega_j$ for $i < j$.
10. Repeat the trick in Step 7 to get $\int_S^{\phi(S)} \omega_i \omega_j$ for $i > j$.
11. Compute $\int_Q^{\phi(Q)} \omega_i$ and use it to deduce $\int_Q^{\phi(Q)} \omega_i \omega_j$ for $i = j$.
12. Compute $\int_Q^{\phi(Q)} \omega_i \omega_j$ for $i < j$.
13. Repeat the trick in Step 7 to get $\int_Q^{\phi(Q)} \omega_i \omega_j$ for $i < j$.
14. Use $\int_S^Q \omega_i = \int_P^Q \omega_i - \int_P^S \omega_i$ to get $\int_S^Q \omega_i$.

Algorithm 7.10 (Double integrals from a Weierstrass endpoint).

Input: A finite Weierstrass point P, a non-Weierstrass point Q, and basis differentials ω_i, ω_j.

Output: The double integrals $\int_P^Q \omega_i \omega_j$.

1. Compute all of the integrals as in Algorithm 7.9.
2. Compute double integrals $\int_S^Q \omega_i \omega_j$ using the terms in Step 1 as appropriate in Algorithm 7.4. (See Remark 7.11 for an additional improvement to this step.)
3. Recover the double integrals $\int_P^Q \omega_i \omega_j = \int_P^S \omega_i \omega_j + \int_S^Q \omega_i \omega_j + \int_P^S \omega_j \int_S^Q \omega_i$ by using additivity.

Remark 7.11. In the case of $g = 1$, the linear system only yields *one* double integral not obtainable through single integrals. Indeed, for $0 \le i, j \le 1$, we have

$$\int_S^Q \omega_i \omega_i = \frac{1}{2}\left(\int_S^Q \omega_i\right)^2 \quad \text{and} \quad \int_S^Q \omega_i \omega_j = -\int_S^Q \omega_j \omega_i + \int_S^Q \omega_i \int_S^Q \omega_j.$$

So it suffices to compute $\int_S^Q \omega_0 \omega_1$. Thus, rather than computing all of the constants $c_{00}, c_{01}, c_{10}, c_{11}$ and their correction factors (see (7)), if we precompute the two double integrals that are expressible in terms of single integrals, as well as the product of single integrals that relates $\int_S^Q \omega_1 \omega_0$ to $\int_S^Q \omega_0 \omega_1$, it suffices to compute c_{01} (and its correction factor) to solve for the other three constants and $\int_S^Q \omega_0 \omega_1$.

In other words, the linear system in Algorithm 7.4 tells us that

$$\left(I_{4\times 4}-(M^t)^{\otimes 2}\right)\begin{pmatrix}\vdots\\ \int_P^Q \omega_i\,\omega_k\\ \vdots\end{pmatrix}=\begin{pmatrix}\vdots\\ c_{ik}-\int_{\phi(P)}^{P}\omega_i\,\omega_k-\left(\int_P^Q\omega_i\right)\left(\int_{\phi(P)}^{P}\omega_k\right)\\ \quad-\left(\int_Q^{\phi(Q)}\omega_i\right)\left(\int_{\phi(P)}^{\phi(Q)}\omega_k\right)+\int_{\phi(Q)}^{Q}\omega_i\,\omega_k\\ \vdots\end{pmatrix},$$

which we write as

$$A\begin{pmatrix}i_{00}\\ v_{01}\\ s_{01}-v_{01}\\ i_{11}\end{pmatrix}=\begin{pmatrix}x_{00}\\ \ell_{01}\\ x_{10}\\ x_{11}\end{pmatrix},$$

where the vector on the left consists of integrals (with $i_{00}=\int_S^Q\omega_0\,\omega_0$, $i_{11}=\int_S^Q\omega_1\,\omega_1$, $s_{01}=\int_S^Q\omega_0\int_S^Q\omega_1$ all computed), and the vector on the right consists of constants (with ℓ_{01} computed). So we solve for $v_{01}:=\int_S^Q\omega_0\,\omega_1, x_{00}, x_{10}, x_{11}$, since knowing v_{01} gives us the complete set of double integrals on basis differentials. While this only gives a constant speedup in terms of complexity, in practice, this helps when S is defined over a highly ramified extension of $\mathbb{Q}_p$.

As numerical checks, one may use the following corollaries of Proposition 7.7.

Corollary 7.12. *For P, Q Weierstrass points and S a third point, we have additivity in endpoints*: $\int_P^Q\omega_i\,\omega_j+\int_Q^S\omega_i\,\omega_j=\int_P^S\omega_i\,\omega_j$.

Corollary 7.13. *For P, Q Weierstrass points, we have*

$$\int_P^Q\omega_i\,\omega_j+\int_P^Q\omega_j\,\omega_i=0.$$

It is worth noting that in general, unlike in the case of a single Coleman integral, for P and Q both Weierstrass points, unless $i=k$, the double Coleman integral $\int_P^Q\omega_i\,\omega_k$ is not necessarily 0. However, in the case of $i=k$, the integral can be computed as $\int_P^Q\omega_i\,\omega_i=\frac{1}{2}\left(\int_P^Q\omega_i\right)^2=0$.

Example 7.14. Consider the curve $y^2=x(x-1)(x+9)$, over $\mathbb{Q}_7$, and the points $P_1=(1,0)$, $P_2=(0,0)$, and $Q=(-1,4)$. We have

$$\begin{pmatrix}\int_{P_1}^Q\omega_0\,\omega_0\\ \int_{P_1}^Q\omega_0\,\omega_1\\ \int_{P_1}^Q\omega_1\,\omega_0\\ \int_{P_1}^Q\omega_1\,\omega_1\end{pmatrix}=\begin{pmatrix}2\cdot 7^2+5\cdot 7^3+4\cdot 7^4+3\cdot 7^5+O(7^6)\\ 6\cdot 7+5\cdot 7^2+4\cdot 7^3+6\cdot 7^4+O(7^6)\\ 2\cdot 7^2+3\cdot 7^3+3\cdot 7^4+7^5+O(7^6)\\ 1+5\cdot 7+5\cdot 7^3+4\cdot 7^4+4\cdot 7^5+O(7^6)\end{pmatrix}$$

and

$$\begin{pmatrix} \int_{P_2}^{Q} \omega_0\omega_0 \\ \int_{P_2}^{Q} \omega_0\omega_1 \\ \int_{P_2}^{Q} \omega_1\omega_0 \\ \int_{P_2}^{Q} \omega_1\omega_1 \end{pmatrix} = \begin{pmatrix} 2\cdot 7^2 + 5\cdot 7^3 + 4\cdot 7^4 + 3\cdot 7^5 + O(7^6) \\ 2\cdot 7^2 + 7^3 + 6\cdot 7^4 + 5\cdot 7^5 + O(7^6) \\ 6\cdot 7 + 5\cdot 7^2 + 6\cdot 7^3 + 3\cdot 7^4 + 3\cdot 7^5 + O(7^6) \\ 1 + 5\cdot 7 + 5\cdot 7^3 + 4\cdot 7^4 + 4\cdot 7^5 + O(7^6) \end{pmatrix},$$

from which we see that $\int_{P_1}^{P_2} \omega_0\omega_1 \neq 0$ and likewise $\int_{P_1}^{P_2} \omega_1\omega_0 \neq 0$.

8. Kim's nonabelian Chabauty method

We now present the motivation for all of the algorithms thus far. Let $\mathscr{C}/\mathbb{Z}$ be the minimal regular model of an elliptic curve $C/\mathbb{Q}$ of analytic rank 1 with Tamagawa numbers all 1. Let $\mathscr{X} = \mathscr{C} - \{\infty\}$ and $\omega_0 = \frac{dx}{2y}, \omega_1 = \frac{x\,dx}{2y}$. Taking a tangential basepoint b at ∞ (or letting b be an integral 2-torsion point), we have the analytic functions

$$\log_{\omega_0}(z) = \int_b^z \omega_0, \quad D_2(z) = \int_b^z \omega_0\omega_1.$$

With this setup, we have:

Theorem 8.1 [2; 14]. *Suppose P is a point of infinite order in $\mathscr{C}(\mathbb{Z})$. Then $\mathscr{X}(\mathbb{Z}) \subset \mathscr{C}(\mathbb{Z}_p)$ is in the zero set of*

$$f(z) := \big(\log_{\omega_0}(P)\big)^2 D_2(z) - \big(\log_{\omega_0}(z)\big)^2 D_2(P).$$

Corollary 8.2 [2; 14]. *The expression*

$$\frac{D_2(P)}{\big(\log_{\omega_0}(P)\big)^2} \tag{8}$$

is independent of the point P of infinite order in $\mathscr{C}(\mathbb{Z})$.

Example 8.3. We revisit Example 1 in [2]. Let E be the rank-1 elliptic curve $y^2 = x^3 - 1323x + 3942$, with minimal model $\mathscr{E}$ having Cremona label 65a1. Consider the following points on E which are integral on $\mathscr{E}$: $b = (3, 0)$, $P = (39, 108)$, $Q = (-33, -108)$, $R = (147, 1728)$. Using Algorithm 7.10, we compute the integrals

$$\int_b^P \omega_0\omega_1 = 4\cdot 11 + 4\cdot 11^2 + 7\cdot 11^3 + 9\cdot 11^4 + 5\cdot 11^6 + O(11^7),$$
$$\int_b^P \omega_0 = 4\cdot 11 + 7\cdot 11^2 + 9\cdot 11^3 + 3\cdot 11^4 + 5\cdot 11^5 + 7\cdot 11^6 + O(11^7),$$

$$\int_b^Q \omega_0\omega_1 = 4\cdot 11 + 4\cdot 11^2 + 7\cdot 11^3 + 9\cdot 11^4 + 5\cdot 11^6 + O(11^7),$$
$$\int_b^Q \omega_0 = 7\cdot 11 + 3\cdot 11^2 + 11^3 + 7\cdot 11^4 + 5\cdot 11^5 + 3\cdot 11^6 + O(11^7),$$
$$\int_b^R \omega_0\omega_1 = 5\cdot 11 + 6\cdot 11^2 + 7\cdot 11^3 + 5\cdot 11^4 + 3\cdot 11^5 + 9\cdot 11^6 + O(11^7),$$
$$\int_b^R \omega_0 = 3\cdot 11 + 7\cdot 11^2 + 2\cdot 11^3 + 3\cdot 11^4 + 7\cdot 11^6 + O(11^7),$$

and we see that the ratio in Corollary 8.2 is constant on integral points:

$$\frac{D_2(P)}{\left(\log_{\omega_0}(P)\right)^2} = \frac{D_2(Q)}{\left(\log_{\omega_0}(Q)\right)^2} = \frac{D_2(R)}{\left(\log_{\omega_0}(R)\right)^2},$$
$$= 3\cdot 11^{-1} + 6 + 2\cdot 11 + 10\cdot 11^2 + 3\cdot 11^3 + 5\cdot 11^4 + O(11^5).$$

However, for $S = (103, 980)$, which is not integral on $\mathscr{E}$, we see that

$$\int_b^S \omega_0\omega_1 = 3\cdot 11 + 10\cdot 11^2 + 4\cdot 11^3 + 10\cdot 11^4 + 7\cdot 11^5 + 10\cdot 11^6 + O(11^7)$$
$$\int_b^S \omega_0 = 11 + 7\cdot 11^3 + 5\cdot 11^5 + O(11^7)$$
$$\frac{D_2(S)}{\left(\log_{\omega_0}(S)\right)^2} = 3\cdot 11^{-1} + 10 + 6\cdot 11 + 9\cdot 11^2 + 8\cdot 11^3 + 6\cdot 11^4 + O(11^5).$$

Example 8.4. We give a variation on Example 4 in [2]. Let E be the rank-1 elliptic curve $y^2 = x^3 - 16x + 16$, with minimal model $\mathscr{E}$ having Cremona label 37a1. Letting P, Q be two fixed integral points on E, we can use the link lemma to rewrite Theorem 8.1 so that the relevant double integral is no longer from a tangential basepoint. Indeed, integral points z occur in the zero set of

$$\left(\left(\int_b^z \omega_0\right)^2 - \left(\int_b^P \omega_0\right)^2\right)\frac{\int_P^Q \omega_0\omega_1 + \int_P^Q \omega_0 \int_b^P \omega_1}{\left(\int_b^Q \omega_0\right)^2 - \left(\int_b^P \omega_0\right)^2}$$
$$- \left(\int_P^z \omega_0\omega_1 + \int_P^z \omega_0 \int_b^P \omega_1\right).$$

Slightly modifying Algorithm 7.4 to take as endpoint a parameter z (see [3, §7.2.2] for more details), we can recover the integral points

$$\{(0, \pm 4), (4, \pm 4), (-4, \pm 4), (8, \pm 20), (24, \pm 116)\}.$$

Remark 8.5. Note that in the classical Chabauty method, one can use the Jacobian of the curve J to find the global constant of integration (see [5; 10]). In particular,

the points on J form a $\mathbb{Z}$-module and we have multiplication-by-n morphisms $[n]\colon J(\mathbb{Q}_p) \to J(\mathbb{Q}_p)$, which gives $n\int_P^Q \omega = \int_{[n](P)}^{[n](Q)} \omega$. By choosing n carefully, we can ensure that $[n]P$ and $[n]Q$ both lie in the residue disk of the identity, and pulling back to the curve, all integrals can be computed by tiny integrals. For iterated integrals, we do not have appropriate endomorphisms available.

Acknowledgments

The author thanks Kiran Kedlaya and Nils Bruin for several helpful conversations, William Stein for access to the computer `sage.math.washington.edu`, and the referees for useful suggestions. This work was done as part of the author's doctoral thesis at MIT, during which she was supported by an NSF Graduate Fellowship and an NDSEG Fellowship. This paper was prepared for submission while the author was supported by NSF grant DMS-1103831.

References

[1] Jennifer S. Balakrishnan, Robert W. Bradshaw, and Kiran S. Kedlaya, *Explicit Coleman integration for hyperelliptic curves*, in Hanrot et al. [11], 2010, pp. 16–31. MR 2012b:14048

[2] Jennifer S. Balakrishnan, Kiran S. Kedlaya, and Minhyong Kim, *Appendix and erratum to "Massey products for elliptic curves of rank 1"*, J. Amer. Math. Soc. **24** (2011), no. 1, 281–291. MR 2011m:11108

[3] Jennifer Sayaka Balakrishnan, *Coleman integration for hyperelliptic curves: algorithms and applications*, Ph.D. thesis, Department of Mathematics, Massachusetts Institute of Technology, 2011. http://hdl.handle.net/1721.1/67785

[4] Amnon Besser, *Coleman integration using the Tannakian formalism*, Math. Ann. **322** (2002), no. 1, 19–48. MR 2003d:11176

[5] Nils Bruin, *Chabauty methods using elliptic curves*, J. Reine Angew. Math. **562** (2003), 27–49. MR 2004j:11051

[6] Kuo-tsai Chen, *Algebras of iterated path integrals and fundamental groups*, Trans. Amer. Math. Soc. **156** (1971), 359–379. MR 43 #1069

[7] Robert Coleman and Ehud de Shalit, *p-adic regulators on curves and special values of p-adic L-functions*, Invent. Math. **93** (1988), no. 2, 239–266. MR 89k:11041

[8] Robert F. Coleman, *Dilogarithms, regulators and p-adic L-functions*, Invent. Math. **69** (1982), no. 2, 171–208. MR 84a:12021

[9] ———, *Torsion points on curves and p-adic abelian integrals*, Ann. of Math. (2) **121** (1985), no. 1, 111–168. MR 86j:14014

[10] E. V. Flynn, Bjorn Poonen, and Edward F. Schaefer, *Cycles of quadratic polynomials and rational points on a genus-2 curve*, Duke Math. J. **90** (1997), no. 3, 435–463. MR 98j:11048

[11] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory: Proceedings of the 9th Biennial International Symposium (ANTS-IX) held in Nancy, July 19–23, 2010*, Lecture Notes in Computer Science, no. 6197, Berlin, Springer, 2010. MR 2011g:11002

[12] Kiran S. Kedlaya, *Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology*, J. Ramanujan Math. Soc. **16** (2001), no. 4, 323–338, errata: [13]. MR 2002m:14019

[13] ———, *Errata for*: *"Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology"*, J. Ramanujan Math. Soc. **18** (2003), no. 4, 417–418. MR 2005c:14027

[14] Minhyong Kim, *Massey products for elliptic curves of rank* 1, J. Amer. Math. Soc. **23** (2010), no. 3, 725–747. MR 2012b:11091

[15] William McCallum and Bjorn Poonen, *The method of Chabauty and Coleman*, preprint, 2010. http://math.mit.edu/~poonen/papers/chabauty.pdf

JENNIFER S. BALAKRISHNAN: jen@math.harvard.edu
Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.63

msp

Finding ECM-friendly curves through a study of Galois properties

Razvan Barbulescu, Joppe W. Bos, Cyril Bouvier,
Thorsten Kleinjung, and Peter L. Montgomery

We prove some divisibility properties of the cardinality of elliptic curve groups modulo primes. These proofs explain the good behavior of certain parameters when using Montgomery or Edwards curves in the setting of the elliptic curve method (ECM) for integer factorization. The ideas behind the proofs help us to find new infinite families of elliptic curves with good division properties increasing the success probability of ECM.

1. Introduction

The elliptic curve method (ECM) for integer factorization [22] is the asymptotically fastest known method for finding relatively small factors p of large integers N. In practice, ECM is used, on the one hand, to factor large integers. For instance, the 2011 ECM record is a 241-bit factor of $2^{1181}-1$ [12]. On the other hand, ECM is used to factor many small (100- to 200-bit) integers as part of the number field sieve [26; 21; 4], the most efficient general purpose integer factorization method.

Traditionally, the elliptic curve arithmetic used in ECM is implemented using Montgomery curves [23] (for example, in the widely used GMP-ECM software [35]). Generalizing the work of Euler and Gauss, Edwards [15] introduced a new normal form for elliptic curves which results in a fast realization of the elliptic curve group operation in practice. These "Edwards curves" have been generalized by Bernstein and Lange [9] for use in cryptography. Bernstein et al. [8] explored the possibility of using these curves in the ECM setting. After Hisil et al. [18] published a coordinate system which results in the fastest known realization of

MSC2010: primary 14H52; secondary 11Y05.
Keywords: elliptic curve method (ECM), Edwards curves, Montgomery curves, torsion properties, Galois groups.

curve arithmetic, a follow-up paper by Bernstein et al. [7] discusses the use of the so-called "$a = -1$" twisted Edwards curves in ECM.

It is common to construct or search for curves which have favorable properties. The success of ECM depends on the smoothness of the cardinality of the curve considered modulo the unknown prime divisor p of N. This usually means constructing curves with large torsion group over $\mathbb{Q}$ or finding curves such that the order of the elliptic curve, when considered modulo a family of primes, is always divisible by an additional factor. Examples are the Suyama construction [32], the curves proposed by Atkin and Morain [3], a translation of these techniques to Edwards curves [8; 7], and a family of curves suitable for Cunningham numbers [13].

In this paper we study and prove divisibility properties of the cardinality of elliptic curves over prime fields. We do this by studying properties of Galois groups of torsion points using Chebotarev's theorem [24]. Furthermore, we investigate some elliptic curve parameters for which ECM finds exceptionally many primes in practice, but which do not fit in any of the known cases of good torsion properties. We prove this behavior and provide parametrizations for infinite families of elliptic curves with these properties.

2. Galois properties of torsion points of elliptic curves

In this section we give a systematic way to compute the probability that the order of a given elliptic curve reduced by an arbitrary prime is divisible by a certain prime power.

2A. *Torsion properties of elliptic curves.*

Definition 2.1. Let K be a finite Galois extension of $\mathbb{Q}$, let p be a prime, and let $\mathfrak{p}$ be a prime ideal of K above p with residue field $k_\mathfrak{p}$. The decomposition group $\mathrm{Dec}(\mathfrak{p})$ of $\mathfrak{p}$ is the subgroup of $\mathrm{Gal}(K/\mathbb{Q})$ that stabilizes $\mathfrak{p}$. Denote by $\alpha^{(\mathfrak{p})}$ the canonical morphism from $\mathrm{Dec}(\mathfrak{p})$ to $\mathrm{Gal}(k_\mathfrak{p}/\mathbb{F}_p)$ and let $\phi_\mathfrak{p}$ be the Frobenius automorphism on the field $k_\mathfrak{p}$. We define

$$\mathrm{Frobenius}(p) = \bigcup_{\mathfrak{p}\mid p} (\alpha^{(\mathfrak{p})})^{-1}(\phi_\mathfrak{p}).$$

We say that a set S of primes *admits a natural density equal to* δ, and we write $\mathsf{P}(S) = \delta$, if

$$\lim_{N\to\infty} \frac{\#(S \cap \Pi(N))}{\#\Pi(N)}$$

exists and equals δ, where $\Pi(N)$ is the set of primes up to N. If $\mathrm{event}(p)$ is a property which can be defined for all primes except a finite set, when we write $\mathsf{P}(\mathrm{event}(p))$ we tacitly exclude the primes where $\mathrm{event}(p)$ cannot be defined.

Theorem 2.2 (Chebotarev, [24]). *Let K be a finite Galois extension of $\mathbb{Q}$. Let $H \subset \mathrm{Gal}(K/\mathbb{Q})$ be a conjugacy class. Then*

$$\mathsf{P}(\mathrm{Frobenius}(p) = H) = \frac{\#H}{\#\mathrm{Gal}(K/\mathbb{Q})}.$$

Before applying Chebotarev's theorem to the case of elliptic curves we introduce some notation. For every elliptic curve E over a field F and for all integers $m \geq 2$, we let $F(E[m])$ denote the smallest extension of F over which all of the geometric m-torsion points of E are rational. The next result is classical, but we present its proof for the intuition it brings.

Proposition 2.3. *For every integer $m \geq 2$ and elliptic curve E over a perfect field F, the following hold*:

(1) *$F(E[m])/F$ is a Galois extension.*

(2) *There is an injective morphism $\iota_m : \mathrm{Gal}(F(E[m])/F) \hookrightarrow \mathrm{Aut}(E(\overline{F})[m])$.*

Proof. Since the addition law of E can be expressed by rational functions over F, there exist polynomials $f_m, g_m \in F[X, Y]$ such that the coordinates of the points in $E(\overline{F})[m]$ are the solutions of the system $(f_m = 0, g_m = 0)$. Therefore $F(E[m])$ is the splitting field of $\mathrm{Res}_X(f_m, g_m)$ and $\mathrm{Res}_Y(f_m, g_m)$ and in particular is Galois. This proves statement (1).

For each $\sigma \in \mathrm{Gal}(F(E[m])/F)$ we denote by $\iota_m(\sigma)$ the function that sends $(x, y) \in E(\overline{F})[m]$ to $(\sigma(x), \sigma(y))$. Thanks to the discussion above, $\iota_m(\sigma)$ sends points of $E(\overline{F})[m]$ to $E(\overline{F})[m]$. Since the addition law can be expressed by rational functions over F, for each σ we have $\iota_m(\sigma) \in \mathrm{Aut}(E(\overline{F})[m])$. One easily checks that ι_m is a group morphism and its kernel is the identity, proving statement (2). □

Notation. Let E be an elliptic curve over $\mathbb{Q}$ and let $m \geq 2$ be an integer. We fix generators for $E(\overline{\mathbb{Q}})[m]$, thereby inducing an isomorphism

$$\psi_m : \mathrm{Aut}(E(\overline{\mathbb{Q}})[m]) \to \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z}).$$

Let ι_m be the injection given by Proposition 2.3, and let $\rho_m : \mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ be the injective morphism $\psi_m \circ \iota_m$.

Let p be a prime such that E has good reduction at p and $p \nmid m$. If k is an extension field of $\mathbb{F}_p$, we write $E(k)$ for the group of k-rational points on the reduction of E modulo p. Let $\iota_m^{(p)}$ be the injection of $\mathrm{Gal}(\mathbb{F}_p(E[m])/\mathbb{F}_p)$ into $\mathrm{Aut}(E(\overline{\mathbb{F}}_p)[m])$ given by Proposition 2.3. By [29, Proposition VII.3.1] there is a canonical isomorphism $r_m^{(\mathfrak{p})}$ from $\mathrm{Aut}(E(\overline{\mathbb{Q}})[m])$ to $\mathrm{Aut}(E(\overline{\mathbb{F}}_p)[m])$ for each prime ideal $\mathfrak{p}$ over p.

Remark 2.4. Note that $\#\mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q})$ is bounded by $\#\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$. For every prime π we have $\#\mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z}) = (\pi - 1)^2(\pi + 1)\pi$, and for every integer $k \geq 1$ we have $\#\mathrm{GL}_2(\mathbb{Z}/\pi^{k+1}\mathbb{Z}) = \pi^4 \#\mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z})$.

Notation. For all $g \in \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ we put $\mathrm{Fix}(g) = \{v \in (\mathbb{Z}/m\mathbb{Z})^2 \mid g(v) = v\}$. If C is a conjugacy class of elements of $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$, we let $\mathrm{Fix}(C)$ denote the isomorphism class of the group $\mathrm{Fix}(g)$, for some $g \in C$; this isomorphism class does not depend on the choice of g. We use analogous notations for the fixed groups of elements of, and conjugacy classes in, the groups $\mathrm{Aut}(E(\overline{\mathbb{Q}})[m])$ and $\mathrm{Aut}(E(\overline{\mathbb{F}}_p)[m])$.

Theorem 2.5. *Let E be an elliptic curve over $\mathbb{Q}$ and let $m \geq 2$ be an integer. Put $K = \mathbb{Q}(E[m])$. Let T be a subgroup of $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$. Then*:

(1) $\mathsf{P}\big(E(\mathbb{F}_p)[m] \simeq T\big) = \dfrac{\#\{g \in \rho_m(\mathrm{Gal}(K/\mathbb{Q})) \mid \mathrm{Fix}(g) \simeq T\}}{\#\,\mathrm{Gal}(K/\mathbb{Q})}$.

(2) *Let a and n be positive integers such that $a \leq n$ and $\gcd(a, n) = 1$, and let ζ_n be a primitive n-th root of unity. Put*

$$G_a = \{\sigma \in \mathrm{Gal}(K(\zeta_n)/\mathbb{Q}) \mid \sigma(\zeta_n) = \zeta_n^a\}.$$

Then

$$\mathsf{P}\big(E(\mathbb{F}_p)[m] \simeq T \mid p \equiv a \bmod n\big) = \frac{\#\{\sigma \in G_a \mid \mathrm{Fix}(\rho_m(\sigma|_K)) \simeq T\}}{\#G_a}.$$

Proof. Let $p \nmid m$ be a prime for which E has good reduction and let $\mathfrak{p}$ be a prime ideal of K over p. Let H denote the set $\{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) \mid \mathrm{Fix}(\iota_m(\sigma)) \simeq T\}$. First note that $E(\mathbb{F}_p)[m] = \mathrm{Fix}(\iota_m^{(p)}(\phi_p))$ where ϕ_p is the Frobenius in $\mathrm{Gal}(\mathbb{F}_p(E[m])/\mathbb{F}_p)$. Since the diagram

$$\begin{array}{ccccc}
\mathrm{Dec}(\mathfrak{p}) & \hookrightarrow & \mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q}) & \xhookrightarrow{\iota_m} & \mathrm{Aut}(E(\overline{\mathbb{Q}})[m]) \\
\downarrow{\scriptstyle \alpha^{(\mathfrak{p})}} & & & & \downarrow{\scriptstyle r_m^{(\mathfrak{p})}} \\
\mathrm{Gal}(k_{\mathfrak{p}}/\mathbb{F}_p) & \xrightarrow{\sim} & \mathrm{Gal}(\mathbb{F}_p(E[m])/\mathbb{F}_p) & \xhookrightarrow{\iota_m^{(p)}} & \mathrm{Aut}(E(\overline{\mathbb{F}}_p)[m])
\end{array}$$

is commutative and since $\mathrm{Frobenius}(p) \subset \mathrm{Gal}(K/\mathbb{Q})$ is the conjugacy class generated by $(\alpha^{(\mathfrak{p})})^{-1}(\phi_{\mathfrak{p}})$ we have $E(\mathbb{F}_p)[m] \simeq \mathrm{Fix}(\iota_m(\mathrm{Frobenius}(p)))$.

Decompose H into a disjoint union of conjugacy classes $C_1, \ldots, C_N$. Then $\mathrm{Fix}(\iota_m(\mathrm{Frobenius}(p))) \simeq T$ if and only if $\mathrm{Frobenius}(p)$ is one of the C_i. Thanks to Theorem 2.2 we obtain

$$\begin{aligned}
\mathsf{P}\big(E(\mathbb{F}_p)[m] \simeq T\big) &= \sum_{i=1}^{N} \mathsf{P}\big(\mathrm{Frobenius}(p) = C_i\big) \\
&= \sum_{i=1}^{N} \frac{\#C_i}{\#\,\mathrm{Gal}(K/\mathbb{Q})} = \frac{\#H}{\#\,\mathrm{Gal}(K/\mathbb{Q})}.
\end{aligned}$$

This proves statement (1).

Using similar arguments, we see that to prove statement (2) we have to evaluate

$$\frac{\mathsf{P}\big(\text{Frobenius}(p) \in \{C_1, \ldots, C_N\},\, p \equiv a \bmod n\big)}{\mathsf{P}(p \equiv a \bmod n)}.$$

Let p be a prime and $\mathfrak{p}$ a prime ideal as in the first part of the proof, and let $\mathfrak{P}$ be a prime ideal of $K(\zeta_n)$ lying over $\mathfrak{p}$. Furthermore let $\widetilde{C}_1, \ldots, \widetilde{C}_{\widetilde{N}}$ be the conjugacy classes of $\mathrm{Gal}(K(\zeta_n)/\mathbb{Q})$ that are in the preimages of $C_1, \ldots, C_N$ and whose elements σ satisfy $\sigma(\zeta_n) = \zeta_n^a$. Since $\mathrm{Gal}(K(\zeta_n)/\mathbb{Q})$ maps ζ_n to primitive n-th roots of unity we have for $\sigma \in (\alpha^{(\mathfrak{P})})^{-1}(\phi_{\mathfrak{P}})$ that $\sigma(\zeta_n) = \zeta_n^b$ for some b. Together with $\sigma(x) \equiv x^p \bmod \mathfrak{P}$ this gives $\zeta_n^b \equiv \zeta_n^p \bmod \mathfrak{P}$. If we exclude the finitely many primes dividing the norms of $\zeta_n^c - 1$ for $c = 1, \ldots, n-1$ we obtain $b \equiv p \bmod n$. Since $\text{Frobenius}(K(\zeta_n), p)$, the Frobenius conjugacy class for $K(\zeta_n)$, is the preimage of $\text{Frobenius}(p)$, the argument above gives

$$\mathsf{P}\big(\text{Frobenius}(p) \in \{C_1, \ldots, C_N\},\, p \equiv a \bmod n\big) = \mathsf{P}\big(\text{Frobenius}(K(\zeta_n), p) \in \{\widetilde{C}_1, \ldots, \widetilde{C}_{\widetilde{N}}\}\big).$$

Considering the denominator $\mathsf{P}(p \equiv a \bmod n)$ similarly completes the proof. □

Remark 2.6. Put $K = \mathbb{Q}(E[m])$. If $[K(\zeta_n) : \mathbb{Q}(\zeta_n)] = [K : \mathbb{Q}]$, then one has

$$\mathsf{P}\big(E(\mathbb{F}_p)[m] \simeq T \mid p \equiv a \bmod n\big) = \mathsf{P}\big(E(\mathbb{F}_p)[m] \simeq T\big)$$

for a coprime to n. Indeed, according to Galois theory,

$$\mathrm{Gal}(K(\zeta_n)/\mathbb{Q})/\,\mathrm{Gal}(K(\zeta_n)/K) \simeq \mathrm{Gal}(K/\mathbb{Q})$$

through $\overline{\sigma} \mapsto \sigma|_K$. Since $[K(\zeta_n) : \mathbb{Q}(\zeta_n)] = [K : \mathbb{Q}]$, we have $[K(\zeta_n) : K] = \varphi(n)$ and therefore each element σ of $\mathrm{Gal}(K/\mathbb{Q})$ extends in exactly one way to an element of $\mathrm{Gal}(K(\zeta_n)/\mathbb{Q})$ which satisfies $\sigma(\zeta_n) = \zeta_n^a$. Note that for $n \in \{3, 4\}$ the condition is equivalent to $\zeta_n \notin K$.

The families constructed by Brier and Clavier [13], which were developed to help factor integers N such that the n-th cyclotomic polynomial has roots modulo all prime factors of N, modify $[K(\zeta_n) : \mathbb{Q}(\zeta_n)]$ by imposing a large torsion subgroup over $\mathbb{Q}(\zeta_n)$.

The following corollary is an important particular case of Theorem 2.5.

Corollary 2.7. *Let E be an elliptic curve over $\mathbb{Q}$ and let π be a prime number. Put $K = \mathbb{Q}(E[\pi])$. Then*

$$\mathsf{P}\big(E(\mathbb{F}_p)[\pi] \simeq \mathbb{Z}/\pi\mathbb{Z}\big) = \frac{\#\{g \in \rho_\pi(\mathrm{Gal}(K/\mathbb{Q})) \mid \det(g - \mathrm{Id}) = 0,\, g \neq \mathrm{Id}\}}{\#\,\mathrm{Gal}(K/\mathbb{Q})},$$

$$\mathsf{P}\big(E(\mathbb{F}_p)[\pi] \simeq \mathbb{Z}/\pi\mathbb{Z} \times \mathbb{Z}/\pi\mathbb{Z}\big) = \frac{1}{\#\,\mathrm{Gal}(K/\mathbb{Q})}.$$

π	T	d_1	$\mathsf{P}_{\text{theor}}(E_1,\pi,T)$ $\mathsf{P}_{\text{exper}}(E_1,\pi,T)$	d_2	$\mathsf{P}_{\text{theor}}(E_2,\pi,T)$ $\mathsf{P}_{\text{exper}}(E_2,\pi,T)$
3	$\mathbb{Z}/3\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}$	48	$\frac{1}{48}\approx 0.02083$ 0.02082	16	$\frac{1}{16}=0.06250$ 0.06245
3	$\mathbb{Z}/3\mathbb{Z}$	48	$\frac{20}{48}\approx 0.4167$ 0.4165	16	$\frac{4}{16}=0.2500$ 0.2501
5	$\mathbb{Z}/5\mathbb{Z}\times\mathbb{Z}/5\mathbb{Z}$	480	$\frac{1}{480}\approx 0.002083$ 0.002091	32	$\frac{1}{32}=0.03125$ 0.03123
5	$\mathbb{Z}/5\mathbb{Z}$	480	$\frac{114}{480}\approx 0.2375$ 0.2373	32	$\frac{10}{32}=0.3125$ 0.3125

Table 1. Theoretical and experimental values of $\mathsf{P}(E,\pi,T):=\mathsf{P}\big(E(\mathbb{F}_p)[\pi]\simeq T\big)$ for the elliptic curves E_1 and E_2, for several primes π and groups T. The theoretical values were obtained from Corollary 2.7, and the experimental values were computed using all primes less than 2^{25}. The columns labeled d_1 and d_2 give the degrees of the number fields $\mathbb{Q}(E_1[\pi])$ and $\mathbb{Q}(E_2[\pi])$, respectively.

Example 2.8. We compute these probabilities for the curves $E_1: y^2=x^3+5x+7$ and $E_2: y^2=x^3-11x+14$ and the primes $\pi=3$ and $\pi=5$. Here E_1 illustrates the generic case, whereas E_2 has special Galois groups. One checks with Sage [30] that $[\mathbb{Q}(E_1[3]):\mathbb{Q}]=48$. Since $\#\,\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})=48$, Proposition 2.3 tells us that $\rho_3(\mathrm{Gal}(\mathbb{Q}(E_1[3])/\mathbb{Q}))=\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$. The group $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ contains 20 nonidentity elements having 1 as an eigenvalue. From Corollary 2.7 we find

$$\mathsf{P}\big(E_1(\mathbb{F}_p)[3]\simeq\mathbb{Z}/3\mathbb{Z}\big)=\tfrac{20}{48},\quad \mathsf{P}\big(E_1(\mathbb{F}_p)[3]\simeq\mathbb{Z}/3\mathbb{Z}\times\mathbb{Z}/3\mathbb{Z}\big)=\tfrac{1}{48}.$$

We used the same method for all the probabilities displayed in Table 1, where we compare them to experimental values.

Note that the relative difference between theoretical and experimental values never exceeds 0.4%. It is interesting to observe that reducing the Galois group does not necessarily increase the probabilities, as it is shown for $\pi=3$.

2B. ***Effective computations of*** $\mathbb{Q}(E[m])$ ***and*** $\rho_m(\mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q}))$ ***for prime powers.*** The main tools we use to compute $\mathbb{Q}(E[m])$ and its Galois group are the division polynomials, as defined below.

Definition 2.9. Let $E: y^2=x^3+ax+b$ be an elliptic curve over $\mathbb{Q}$ and $m\geq 2$ an integer. The *m-division polynomial* P_m is the monic polynomial whose roots are the x-coordinates of all the affine m-torsion points of E. We also define P_m^{new} to be the monic polynomial whose roots are the x-coordinates of the affine points of order exactly m.

Proposition 2.10. *For all $m \geq 2$ the polynomials P_m and P_m^{new} lie in $\mathbb{Q}[X]$. Furthermore,* $\deg(P_m) = (m^2 + 2 - 3\eta)/2$, *where η is the remainder of m modulo 2.*

Proof. For a proof we refer to [29, Exercise III.3.7, pp. 105–106]. □

Note that one obtains different division polynomials for other shapes of elliptic curves (Weierstrass, Montgomery, Edwards, and so on). Nevertheless, the Galois group $\mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q})$ is independent of the model of E, and can be computed with the division polynomials of Definition 2.9 because, in characteristic different from 2 and 3, every curve can be written in short Weierstrass form.

One can compute $\mathbb{Q}(E[\pi])$ for any prime $\pi \geq 3$ using the following method.

1. Make a first extension of $\mathbb{Q}$ through an irreducible factor of P_π to obtain a number field F_1 where P_π has a root α_1.
2. Let $f_2(y) = y^2 - (\alpha_1^3 + a\alpha_1 + b) \in F_1[y]$ and F_2 be the splitting field of f_2. There is a π-torsion point M_1 of E defined over F_2. In F_2, P_π has $(\pi - 1)/2$ trivial roots representing the x coordinates of the multiples of M_1.
3. Let F_3 be the extension of F_2 defined by an irreducible factor of $P_\pi \in F_2[x]$ other than those corresponding to the trivial roots.
4. Let α_2 be a new root of P_π in F_3. Let $f_4(y) = y^2 - (\alpha_2^3 + a\alpha_2 + b) \in F_3[y]$ and let F_4 be the splitting field of f_4. Then F_4 contains all the π-torsion of E.

The case of prime powers π^k with $k \geq 2$ is handled recursively. Having computed $\mathbb{Q}(E[\pi^{k-1}])$, we obtain $\mathbb{Q}(E[\pi^k])$ by repeating the four steps above with $P_{\pi^k}^{\text{new}}$ instead of P_π and by defining trivial roots to be the x-coordinates of the points $\{P + M_1 \mid P \in E[\pi^{k-1}]\}$.

In practice, we observe that in general P_π, f_2, $P_\pi^{(F_2)}$ and f_4 are irreducible, where $P_\pi^{(F_2)}$ is P_π divided by the factors corresponding to the trivial roots. If this is the case, then using the formula $\deg(P_\pi) = (\pi^2 - 1)/2$ from Proposition 2.10, we find that the absolute degree of F_4 is

$$\frac{\pi^2 - 1}{2} \cdot 2 \cdot \frac{\pi^2 - \pi}{2} \cdot 2 = (\pi - 1)^2(\pi + 1)\pi.$$

By Remark 2.4, $\#\,\mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z})$ is also equal to $(\pi - 1)^2(\pi + 1)\pi$, so in general we expect $\rho_\pi(\mathrm{Gal}(\mathbb{Q}(E[\pi])/\mathbb{Q})) = \mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z})$. Also, we observed that in general the degree of the extension $\mathbb{Q}(E[\pi^k])/\mathbb{Q}(E[\pi^{k-1}])$ is π^4.

The next theorem shows that the observations above are almost always true. It is a restatement of items (1) and (6) from the introduction of [27].

Theorem 2.11 (Serre). *Let E be an elliptic curve without complex multiplication.*

(1) *For all primes π the sequence of indices*

$$[\mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z}) : \rho_{\pi^k}(\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q}))] \quad \textit{for } k \geq 1$$

is nondecreasing and bounded by a constant depending on E and π.

(2) *For all primes π outside a finite set depending on E and for all $k \geq 1$,*

$$\rho_{\pi^k}(\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q}) = \mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z}).$$

Definition 2.12. Put $I(E,\pi,k) = [\mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z}) : \rho_{\pi^k}(\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q}))]$. If E does not admit complex multiplication, we define *Serre's exponent* to be the integer

$$n(E,\pi) = \min\{n \in \mathbb{Z}_{>0} \mid \forall k \geq n : I(E,\pi,k+1) = I(E,\pi,k)\}.$$

In [28] Serre showed that in some cases one can prove that $I(E,\pi,k) = 1$ for all positive integers k. Indeed, Serre proved that the surjectivity of ρ_{π^k} (or the equivalent equality $I(E,\pi,k) = 1$) follows from the surjectivity of ρ_π (or the equivalent equality $I(E,\pi,1) = 1$) for all rational elliptic curves E without complex multiplication and for all primes $\pi \geq 5$. In order to have the same kind of results for $\pi = 2$ (respectively, $\pi = 3$) one has to suppose that ρ_2, ρ_4 and ρ_8 are surjective (respectively, ρ_3 and ρ_9 are surjective).

Serre also conjectured that only a finite number of primes, not depending on the curve E, can occur in the second point of Theorem 2.11. The current conjecture is that for all rational elliptic curves without complex multiplication and all primes $\pi \geq 37$, ρ_π is surjective. Zywina [36] describes an algorithm that computes, for a given E, the primes π for which ρ_π is not surjective; Zywina has checked the conjecture for all elliptic curves in Magma's database (currently this covers curves with conductor at most 140,000). For other recent progress on this conjecture of Serre, see [11] and [10].

Remark 2.13. One application of Serre's results is as follows. Experiments show that if E is an elliptic curve over $\mathbb{Q}$ without complex multiplication, then $E(\mathbb{F}_p)$ is close to a cyclic group for almost all primes p, regardless of the rank of E over $\mathbb{Q}$. For a given bound B, computing

$$\mathsf{P}(\exists \pi > B \mid \mathbb{Z}/\pi\mathbb{Z} \times \mathbb{Z}/\pi\mathbb{Z} \subset E(\mathbb{F}_p)) \tag{1}$$

goes beyond the scope of this paper. However, if π is a prime such that ρ_π is surjective, then Corollary 2.7 shows that

$$\mathsf{P}(\mathbb{Z}/\pi\mathbb{Z} \times \mathbb{Z}/\pi\mathbb{Z} \subset E(\mathbb{F}_p)) = \frac{1}{\pi(\pi+1)(\pi-1)^2}.$$

This suggests that the probability in expression (1) should be $O(1/B^3)$.

The method described above allows us to compute $\mathbb{Q}(E[m])$ as an extension tower. Then it is easy to obtain its absolute degree and a primitive element. Identifying $\rho_\pi(\mathrm{Gal}(\mathbb{Q}(E[m])/\mathbb{Q}))$ up to conjugacy is easy when there is only one subgroup (up to conjugacy) of $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ with the right order. When this is not the case

we use fixed generators for $E(\bar{\mathbb{Q}})[m]$ to check for each $g \in \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ whether g gives rise to an automorphism on $\mathbb{Q}(E[m])$. In practice, the bottleneck of this method is the factorization of polynomials with coefficients over number fields.

A faster probabilistic algorithm for computing $\mathrm{Gal}(\mathbb{Q}(E[\pi])/\mathbb{Q})$ was proposed by Sutherland [31]. This algorithm was not known by the authors at the time of writing and would have helped to accelerate the computation of the examples.

2C. ***Divisibility by a prime power.*** It is well-known that, for a given prime π, the cardinality of a randomly chosen elliptic curve over $\mathbb{F}_p$ has a larger probability of being divisible by π than a randomly chosen integer of size p (see [22, Proposition 1.14, p. 660]). In this subsection we shall consider the analogous problem, where instead of fixing p and varying E, we fix an $E/\mathbb{Q}$ and vary p.

Notation. Let π be a prime and let i, j, and k be nonnegative integers such that $i \le j$. We put

$$p_{\pi,k}(i,j) = \mathsf{P}\big(E(\mathbb{F}_p)[\pi^k] \simeq \mathbb{Z}/\pi^i\mathbb{Z} \times \mathbb{Z}/\pi^j\mathbb{Z}\big).$$

Let $\ell \le m$ be integers. When it is defined we write

$$p_{\pi,k}(\ell,m \mid i,j) = \mathsf{P}\big(E(\mathbb{F}_p)[\pi^{k+1}] \simeq \mathbb{Z}/\pi^\ell\mathbb{Z} \times \mathbb{Z}/\pi^m\mathbb{Z} \;\big|\; E(\mathbb{F}_p)[\pi^k] \simeq \mathbb{Z}/\pi^i\mathbb{Z} \times \mathbb{Z}/\pi^j\mathbb{Z}\big).$$

When it is clear from the context, π is omitted.

Remark 2.14. Since for every integer $m > 0$ and every prime p coprime to m we have $E(\mathbb{F}_p)[m] \subset \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$, it follows that $p_{\pi,k}(i,j) = 0$ for $j > k$. In the case $j < k$, if $p_{\pi,k}(\ell,m \mid i,j)$ is defined, it equals 1 if $(\ell,m) = (i,j)$ and equals 0 if $(\ell,m) \ne (i,j)$. Finally, for $j = k$, there are only three conditional probabilities which can be nonzero: $p_{\pi,k}(i,k \mid i,k)$, $p_{\pi,k}(i,k+1 \mid i,k)$, and $p_{\pi,k}(k+1,k+1 \mid k,k)$.

Theorem 2.15. *Let π be a prime and E an elliptic curve over $\mathbb{Q}$. If k is an integer such that $I(E,\pi,k+1) = I(E,\pi,k)$ (for example, if E has no complex multiplication and $k \ge n(E,\pi)$), then we have*

$$p_{\pi,k}(k+1,k+1 \mid k,k) = 1/\pi^4,$$
$$p_{\pi,k}(k,k+1 \mid k,k) = (\pi-1)(\pi+1)^2/\pi^4, \quad \textit{and}$$
$$p_{\pi,k}(i,k+1 \mid i,k) = 1/\pi \quad \textit{for } 0 \le i < k.$$

Proof. Let $M = (\mathbb{Z}/\pi^k\mathbb{Z})^2$. For all $g \in \mathrm{GL}_2(\pi M)$, we consider the set

$$\mathrm{Lift}(g) = \big\{h \in \mathrm{GL}_2(M) \;\big|\; h|_{\pi M} = g\big\} = \big\{g + \pi^{k-1}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \;\big|\; a,b,c,d \in \mathbb{Z}/\pi\mathbb{Z}\big\},$$

whose cardinality is π^4. Since $I(E,\pi,k+1) = I(E,\pi,k)$ we have

$$\frac{\#\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q})}{\#\mathrm{Gal}(\mathbb{Q}(E[\pi^{k+1}])/\mathbb{Q})} = \frac{\#\mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z})}{\#\mathrm{GL}_2(\mathbb{Z}/\pi^{k+1}\mathbb{Z})},$$

which equals $1/\pi^4$ by Remark 2.4. So for all $g \in \rho_{\pi^k}(\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q}))$, we have $\mathrm{Lift}(g) \subset \rho_{\pi^{k+1}}(\mathrm{Gal}(\mathbb{Q}(E[\pi^{k+1}])/\mathbb{Q}))$. Thanks to Theorem 2.5, the proof will follow if we count for each g the number of lifts with a given fixed group.

For $g = \mathrm{Id} \in \rho_{\pi^k}(\mathrm{Gal}(\mathbb{Q}(E[\pi^k])/\mathbb{Q}))$, there is only one element of $\mathrm{Lift}(g)$ fixing $(\mathbb{Z}/\pi^{k+1}\mathbb{Z})^2$, so $p_{\pi,k}(k+1,k+1 \mid k,k) = 1/\pi^4$.

The element $g = \mathrm{Id}$ can be lifted in exactly $\pi^4 - 1 - \#\mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z})$ ways to an element in $\mathrm{GL}_2(\mathbb{Z}/\pi^{k+1}\mathbb{Z})$ that fixes the π^k-torsion and a point of order π^{k+1}, but not all the π^{k+1}-torsion. Therefore $p_{\pi,k}(k,k+1 \mid k,k) = (\pi-1)(\pi+1)^2/\pi^4$.

Every element of $\mathrm{GL}_2(\mathbb{Z}/\pi^k\mathbb{Z})$ that fixes a line, but is not the identity, can be lifted in exactly π^3 ways to an element of $\mathrm{GL}_2(\mathbb{Z}/\pi^{k+1}\mathbb{Z})$ that fixes a line of $(\mathbb{Z}/\pi^{k+1}\mathbb{Z})^2$. This shows that $p_{\pi,k}(i,k+1 \mid i,k) = \pi^3/\pi^4 = 1/\pi$. □

The theorem below uses the information on $\mathrm{Gal}(\mathbb{Q}(E[\pi^{n(E,\pi)}])/\mathbb{Q})$ for a given prime π in order to compute the probabilities of divisibility by any power of π. It also gives a formula for the average π-adic valuation $\overline{v}_\pi$ of $\#E(\mathbb{F}_p)$, which we define as

$$\overline{v}_\pi = \sum_{k\geq 1} k\,\mathsf{P}\big(v_\pi(\#E(\mathbb{F}_p)) = k\big),$$

where v_π denotes π-adic valuation. We do not claim that $\overline{v}_\pi$ is equal to

$$\lim_{x\to\infty} \frac{1}{\#\Pi(x)} \sum_{p\leq x} v_\pi(\#E(\mathbb{F}_p)),$$

although we expect this to be true.

Notation. Let π be a prime. We set $\gamma_n(h) = \pi^n \sum_{\ell=0}^{h} \pi^\ell p_n(\ell,n)$, and we define

$$\delta(k) = \begin{cases} p_{i+1}(i+1,i+1) & \text{if } k = 2i+1, \\ 0 & \text{otherwise} \end{cases}$$

and

$$S_k(h) = \pi^k\left(\delta(k) + \sum_{\ell=h}^{\lfloor k/2\rfloor} p_{k-\ell}(\ell,k-\ell)\right).$$

Theorem 2.16. *Let π be a prime, let E an elliptic curve over $\mathbb{Q}$, and let n be a positive integer such that $I(E,\pi,k) = I(E,\pi,n)$ for all $k \geq n$ (for example, a curve without complex multiplication and $n \geq n(E,\pi)$). Then, for every $k \geq 1$,*

$$\mathsf{P}\big(\#E(\mathbb{F}_p) \equiv 0 \bmod \pi^k\big) = \frac{1}{\pi^k}\begin{cases} S_k(0) & \text{if } 1 \le k \le n, \\ \gamma_n(k-n-1) + S_k(k-n) & \text{if } n < k \le 2n, \\ \gamma_n(n) + p_n(n,n)\pi^{2n-1} - \pi^{4n-1-k}\, p_n(n,n) & \text{if } k > 2n. \end{cases}$$

Furthermore, $\overline{v}_\pi$ is finite, and we have

$$\overline{v}_\pi = 2\sum_{\ell=1}^{n-1} p_\ell(\ell,\ell) + \frac{\pi}{\pi-1}\sum_{\ell=0}^{n-1} p_n(\ell,n) + \sum_{\ell=0}^{n-2}\sum_{i=\ell+1}^{n-1} p_i(\ell,i) + \frac{\pi(2\pi+1)}{(\pi-1)(\pi+1)}\, p_n(n,n).$$

Proof. Let k be a positive integer. Using Figure 1, one checks that

$$\mathsf{P}\big(\#E(\mathbb{F}_p) \equiv 0 \bmod \pi^k\big) = \sum_{\ell=0}^{\lfloor k/2 \rfloor} p_{k-\ell}(\ell, k-\ell) + \delta(k). \tag{2}$$

Let $c_1 = 1/\pi^4$, $c_2 = (\pi-1)(\pi+1)^2/\pi^4$, and $c_3 = 1/\pi$. With these notations, the situation can be illustrated by Figure 1. For $j > n$ and $\ell < n$, the probability $p_j(\ell, j)$ is the product of the conditional probabilities of the unique path from (ℓ, j) to (ℓ, n) in the graph of Figure 1 times the probability $p_n(\ell, n)$. For $j > n$ and $\ell \ge n$, the probability $p_j(\ell, j)$ is the product of the conditional probabilities of the unique path from (ℓ, j) to (n, n) in the graph of Figure 1 times the probability $p_n(n, n)$.

There are three cases that are to be treated separately: $1 \le k \le n$, $n < k \le 2n$ and $k > 2n$. For $1 \le k \le n$, the result follows from (2). Let us give the computation

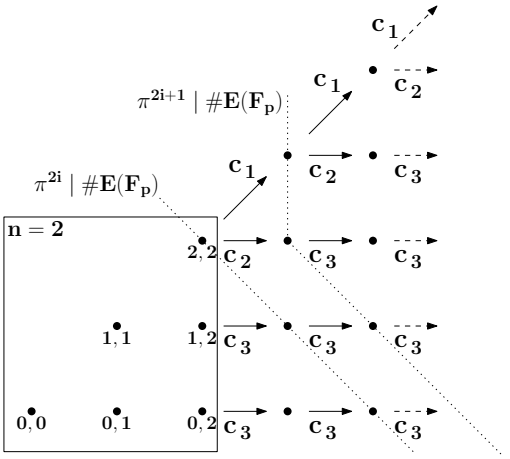


Figure 1. The node with coordinates (i, j) represents the event $\big(E(\mathbb{F}_p)[\pi^j] \simeq \mathbb{Z}/\pi^i\mathbb{Z} \times \mathbb{Z}/\pi^j\mathbb{Z}\big)$. The arrows represent the conditional probabilities of Theorem 2.15.

in more detail for the case for $k > 2n$, with $k = 2i$:

$$\begin{aligned}
&\mathsf{P}\big(\#E(\mathbb{F}_p) \equiv 0 \bmod \pi^{2i}\big) \\
&\quad = \sum_{\ell=0}^{i} p_{2i-\ell}(\ell, 2i-\ell) + \delta(2i) = \sum_{\ell=0}^{i} p_{2i-\ell}(\ell, 2i-\ell) \\
&\quad = \sum_{\ell=0}^{n-1} p_{2i-\ell}(\ell, 2i-\ell) + \sum_{\ell=n}^{i-1} p_{2i-\ell}(\ell, 2i-\ell) + p_i(i,i) \\
&\quad = \sum_{\ell=0}^{n-1} c_3^{2i-l-n} p_n(\ell, n) + \sum_{\ell=n}^{i-1} c_3^{2i-2l-1} c_2 c_1^{l-n} p_n(n,n) + c_1^{i-n} p_n(n,n).
\end{aligned}$$

This leads to the desired formula. The case $k > 2n$ odd and the case $n < k \le 2n$ are treated similarly.

To prove the statements about $\bar{v}_\pi$, we note that $\mathsf{P}(\#E(\mathbb{F}_p) \equiv 0 \bmod \pi^k)$ is $O(1/\pi^k)$ as $k \to \infty$. Thus, the sum defining $\bar{v}_\pi$ is absolutely convergent, and we are justified in rearranging terms to find

$$\bar{v}_\pi = \sum_{k \ge 1} k\, \mathsf{P}\big(v_\pi(\#E(\mathbb{F}_p)) = k\big) = \sum_{k \ge 1} \mathsf{P}\big(\#E(\mathbb{F}_p) \equiv 0 \bmod \pi^k\big).$$

Substituting in our formulas for the summands in the last expression, we obtain the formula for $\bar{v}_\pi$ given in the theorem. □

Example 2.17. Let us compare the theoretical and experimental average valuation of $\pi = 2$, $\pi = 3$ and $\pi = 5$ for the curves

$$E_1 : \ y^2 = x^3 + 5x + 7 \quad \text{and} \quad E_3 : \ y^2 = x^3 - 10875x + 526250,$$

which do not admit complex multiplication. (We exclude E_2 in this example because it does have complex multiplication.) For E_1, we apply Theorem 2.16 with $n = 1$ and compute the necessary probabilities with Corollary 2.7 knowing that the Galois groups are isomorphic to $\mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z})$. For E_3, we apply Theorem 2.16 with $n = 3$ for $\pi = 2$ and $n = 1$ for $\pi = 3$ and $\pi = 5$, and compute the necessary probabilities with Theorem 2.5 (when $n = 3$) and Corollary 2.7 (when $n = 1$). The results are shown in Table 2.

In order to apply Theorem 2.16, one has to show that $I(E, \pi, k) = I(E, \pi, n)$ for all $k \ge n$ (or $n \ge n(E, \pi)$ since E_1 and E_3 do not have complex multiplication). For E_1, we were able to prove that $n(E, \pi) = 1$ for $\pi = 2$, $\pi = 3$, and $\pi = 5$ by using the remarks at the end of Section 2B. For E_3, Andrew Sutherland computed for us the Galois groups up to the 2^5-, 3^3-, and 5^2-torsion. These computations lead us to believe that $n(E_3, 2) = 3$, $n(E_3, 3) = 1$, and $n(E_3, 5) = 1$, but we have been unable to prove that these values are correct; in particular, this means that the theoretical probabilities for E_3 given in Table 2 are conjectural.

π	$n(E_1,\pi)$	$\overline{v}_{\pi,\text{theor}}$ $\overline{v}_{\pi,\text{exper}}$	$n(E_3,\pi)$	$\overline{v}_{\pi,\text{theor}}$ $\overline{v}_{\pi,\text{exper}}$
2	1	$\frac{14}{9} \approx 1.556$ 1.555	3	$\frac{895}{576} \approx 1.554$ 1.554
3	1	$\frac{87}{128} \approx 0.680$ 0.679	1	$\frac{39}{32} \approx 1.219$ 1.218
5	1	$\frac{695}{2304} \approx 0.302$ 0.301	1	$\frac{155}{192} \approx 0.807$ 0.807

Table 2. Theoretical and experimental values of the average π-adic valuation of $\#E_1(\mathbb{F}_p)$ and $\#E_3(\mathbb{F}_p)$, for $\pi = 2, 3, 5$. The theoretical values come from Theorem 2.16, and the experimental values were computed using all primes less than 2^{25}. The values of $n(E_3,\pi)$ and those of $\overline{v}_{\pi,\text{theor}}$ for E_3 are conjectural.

3. Applications to some families of elliptic curves

As shown in the preceding section, changing the torsion properties is equivalent to modifying the Galois group. One can view the imposition of rational torsion points as a way of modifying the Galois group. In this section we change the Galois group either by splitting the division polynomials or by imposing some equations that directly modify the Galois group. With these ideas, we find new infinite ECM-friendly families and we explain the properties of some known curves.

3A. ***Preliminaries on Montgomery and twisted Edwards curves.*** Let K be a field whose characteristic is neither 2 nor 3.

Edwards curves. For $a, d \in K$, with $ad(a-d) \neq 0$, the twisted Edwards curve $ax^2 + y^2 = 1 + dx^2y^2$ is denoted by $E_{a,d}$. The "$a = -1$" twisted Edwards curves are denoted by E_d. In [8] completed twisted Edwards curves are defined by

$$\overline{E}_{a,d} = \big\{\big((X:Z),(Y:T)\big) \in \mathbb{P}^1 \times \mathbb{P}^1 \mid aX^2T^2 + Y^2Z^2 = Z^2T^2 + dX^2Y^2\big\}.$$

The completed points are the affine (x, y) embedded into $\mathbb{P}^1 \times \mathbb{P}^1$ by the map $(x, y) \mapsto ((x:1),(y:1))$; see [8] for more information. We denote $(1:0)$ by ∞.

Figure 2 gives an overview of all the 2- and 4-torsion, as well as some of the 8-torsion points, on $\overline{E}_{a,d}$, as specified in [8].

Montgomery curves and the Suyama family. Take $A, B \in K$ with $B(A^2-4) \neq 0$. The Montgomery curve $By^2 = x^3 + Ax^2 + x$ associated to (A, B) is denoted by $M_{A,B}$ (see [23]) and its completion in $\mathbb{P}^2$ by $\overline{M}_{A,B}$.

Remark 3.1. If $a, d, A, B \in K$ are such that $d = (A-2)/B$ and $a = (A+2)/B$, then there is a birational map between $\overline{E}_{a,d}$ and $\overline{M}_{A,B}$ given by

$$\big((x:z),(y:t)\big) \mapsto \big((t+y)x : (t+y)z : (t-y)x\big)$$

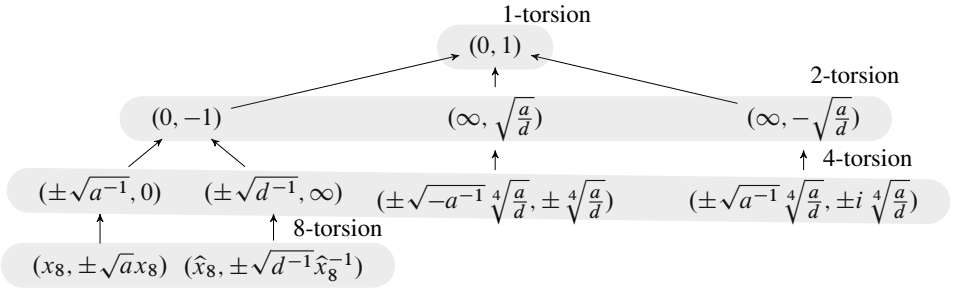


Figure 2. An overview of all 1-, 2-, and 4-torsion and some 8-torsion points on twisted Edwards curves. The x_8 and $\hat{x}_8$ in the 8-torsion points are such that $adx_8^4 - 2ax_8^2 + 1 = 0$ and $ad\hat{x}_8^4 - 2d\hat{x}_8^2 + 1 = 0$.

(see [6]). Therefore $\overline{M}_{A,B}$ and $\overline{E}_{a,d}$ have the same group structure over any field where they are both defined, and in particular they have the same torsion properties. Any statement in twisted Edwards language can be easily translated into Montgomery coordinates and vice versa.

A Montgomery curve for which there exist $x_3, y_3, k, x_\infty, y_\infty \in \mathbb{Q}$ such that

$$\begin{cases} P_3(x_3) = 0, \quad By_3^2 = x_3^3 + Ax_3^2 + x_3 & \text{(3-torsion point)}, \\ k = \dfrac{y_3}{y_\infty}, \quad k^2 = \dfrac{x_3^3 + Ax_3^2 + x_3}{x_\infty^3 + Ax_\infty^2 + x_\infty} & \text{(nontorsion point)}, \\ x_\infty = x_3^3 & \text{(Suyama equation)} \end{cases} \tag{3}$$

is called a Suyama curve. As described in [32; 34], the solutions of (3) can be parametrized by a rational value denoted σ. For all $\sigma \in \mathbb{Q}\setminus\{0, \pm 1, \pm 3, \pm 5, \pm\frac{5}{3}\}$, the associated Suyama curve has positive rank and a rational point of order 3.

Remark 3.2. In the following, when we say that a twisted Edwards curve $E_{a,d}$ (or a Montgomery curve $M_{A,B}$) has good reduction modulo a prime p, we also suppose that we have $v_p(a) = v_p(d) = v_p(a-d) = 0$ (respectively, $v_p(A-2) = v_p(A+2) = v_p(B) = 0$ for a Montgomery curve). In this case the reduction map is simply given by reducing the coefficients modulo p. The results below are also true for primes of good reduction which do not satisfy these conditions, by slightly modifying the statements and the proofs. Moreover, in ECM, if the conditions are not satisfied, we immediately find the factor p.

3B. ***The generic Galois group of a family of curves.*** In the following, when we talk about the *Galois group of the m-torsion of a family of curves*, we mean a group isomorphic to the Galois group of the m-torsion for all curves of the family except for a sparse set of curves (which can have a smaller Galois group).

For example, let us consider the Galois group of the 2-torsion for the family $\{\mathscr{E}_r : y^2 = x^3 + rx^2 + x \mid r \in \mathbb{Q} \setminus \{\pm 2\}\}$. The Galois group of the 2-torsion of the curve $\mathscr{E} : y^2 = x^3 + Ax^2 + x$ over $\mathbb{Q}(A)$ is $\mathbb{Z}/2\mathbb{Z}$. Hence, for most values of r the Galois group is $\mathbb{Z}/2\mathbb{Z}$ and for a sparse set of values the Galois group is the trivial group. So, we say that the Galois group of the 2-torsion of this family is $\mathbb{Z}/2\mathbb{Z}$.

To our best knowledge, there is no implementation of an algorithm computing Galois groups of polynomials with coefficients in a function field. Instead we can compute the Galois group for every curve of the family, so we can guess the Galois group of the family from a finite number of instantiations. In practice, we took a dozen random curves in the family; if the Galois groups of the m-torsion for these curves were all the same, we guessed that it was the Galois group of the m-torsion of the family of curves.

3C. ***Study of the 2^k-torsion of Montgomery and twisted Edwards curves.*** The rational torsion of a Montgomery/twisted Edwards curve is $\mathbb{Z}/2\mathbb{Z}$ but it is known that 4 divides the order of the curve when reduced modulo any prime p [32]. The following theorem gives more detail on the 2^k-torsion.

Theorem 3.3. *Let $E = E_{a,d}$ be a twisted Edwards curve (respectively, a Montgomery curve $M_{A,B}$) over $\mathbb{Q}$. Let p be a prime such that E has good reduction at p.*

(1) *Suppose $p \equiv 3 \pmod 4$. If a/d (respectively, $A^2 - 4$) is a quadratic residue modulo p, then $E(\mathbb{F}_p)[4] \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$.*

(2) *Suppose $p \equiv 1 \pmod 4$. If a (respectively, $(A+2)/B$) is a quadratic residue modulo p (in particular, if $a = \pm 1$) and a/d (respectively, $A^2 - 4$) is a quadratic residue modulo p, then $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z} \subset E(\mathbb{F}_p)[4]$.*

(3) *Suppose $p \equiv 1 \pmod 4$. If a/d (respectively, $A^2 - 4$) is a quadratic nonresidue modulo p and $a - d$ (respectively, B) is a quadratic residue modulo p, then $E(\mathbb{F}_p)[8] \simeq \mathbb{Z}/8\mathbb{Z}$.*

Proof. Using Remark 3.1, it is enough to prove the results in the Edwards language, which follow by some calculations using Figure 2. □

Theorem 3.3 suggests that by imposing equations on the parameters a and d we can improve the torsion properties. The case where a/d is a square has been studied in [8] for the family of Edwards curves with $a = 1$ and rational torsion group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$, and in [7] for the family with $a = -1$ and rational torsion group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$. Here we focus on two other equations:

$$\exists c \in \mathbb{Q},\ a = -c^2 \qquad (A + 2 = -Bc^2 \text{ for Montgomery curves}), \tag{4}$$

$$\exists c \in \mathbb{Q},\ a - d = c^2 \qquad (B = c^2 \text{ for Montgomery curves}). \tag{5}$$

The cardinality of the Galois group of the 4-torsion for generic Montgomery curves is 16; this is reduced to 8 for the family of curves satisfying (4). Using Theorem 2.5, we can compute the changes of probabilities due to this new Galois group. For all curves satisfying (4) and all primes $p \equiv 1 \pmod 4$, the probability of having $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ as the 4-torsion group becomes 0 instead of $\frac{1}{4}$; the probabilities of having $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/4\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ as the 4-torsion group become $\frac{1}{4}$ instead of $\frac{1}{8}$.

The Galois group of the 8-torsion of the family of curves satisfying (5) has cardinality 128, instead of 256 for generic Montgomery curves. Using Theorem 2.5, one can see that the probabilities of having an 8-torsion point are improved.

Using Theorem 2.16, one can show that for both families of curves — the family satisfying (4) and the one satisfying (5) — the probability that the cardinality is divisible by 8 increases from $\frac{5}{8}$ to $\frac{3}{4}$, and the average valuation of 2 increase from $\frac{10}{3}$ to $\frac{11}{3}$.

3D. ***Better twisted Edwards curves with torsion $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ using division polynomials.*** In this section we search for curves such that some of the factors of the division polynomials split; by doing so, we hope to change the Galois groups. As an example we consider the family of $a = -1$ twisted Edwards curves E_d with $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$-torsion; these curves are exactly the ones with $d = -e^4$ (see [7]). The technique might be used in any context.

Looking for subfamilies. For a generic d, the polynomial P_8^{new} splits into three irreducible factors: two of degree 4 and one of degree 16. If one takes $d = -e^4$, the polynomial of degree 16 splits into three factors: two of degree 4, called $P_{8,0}$ and $P_{8,1}$, and one of degree 8, called $P_{8,2}$. By trying to force one of these three polynomials to split, we found four families, as shown in Table 3.

In all these families the generic average valuation of 2 is increased by $\frac{1}{6}$ — rising from $\frac{14}{3}$ up to $\frac{29}{6}$ — except for the family $e = (g - g^{-1})/2$, for which it is increased by $\frac{2}{3}$, bringing it to the same valuation as for the family of twisted Edwards curves with $a = 1$ and torsion isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$. Note that these four families cover all the curves presented in the first three columns of [7, Table 3.1], except the two curves with $e = \frac{26}{7}$ and $e = \frac{19}{8}$, which have a generic Galois group for the 8-torsion.

The family $e = (g - g^{-1})/2$. In this section, we study in more detail the family $e = (g - g^{-1})/2$. Using Theorem 2.5 one can prove that the group order modulo all primes is divisible by 16. However, we give an alternative proof which is also of independent interest. We need the following theorem which computes the 8-torsion points that double to the 4-torsion points $(\pm\sqrt[4]{-d^{-1}}, \pm\sqrt[4]{-d^{-1}})$.

Special form of e	Degrees of factors of $P_{8,0}$	$P_{8,1}$	$P_{8,2}$	Avg. 2-adic val. over p that are 1 mod 4	3 mod 4	all p
none	4	4	8	16/3	4	14/3
g^2	4	4	4, 4	17/3	4	29/6
$(2g^2+2g+1)/(2g+1)$	4	4	4, 4	17/3	4	29/6
$g^2/2$	2, 2	4	8	17/3	4	29/6
$(g-g^{-1})/2$	2, 2	2, 2	8	17/3	5	16/3

Table 3. Averages, over different subsets of primes, of the 2-adic valuation of $\#E(\mathbb{F}_p)$, for E in one of several subfamilies of twisted Edwards curves E_d with torsion group isomorphic to $\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/4\mathbb{Z}$. The subfamilies all have $d=-e^4$, where e is further specialized according to the entries in the first column. The second through fourth columns give the degrees of the factors of the polynomials $P_{8,i}$ defined in the article. The fifth through seventh columns give the average 2-adic valuation of $\#E(\mathbb{F}_p)$ as p ranges through primes that are 1 modulo 4, primes that are 3 modulo 4, and all primes, respectively.

Theorem 3.4. *Let E_d be a twisted Edwards curve over $\mathbb{Q}$ with $d=-e^4$, where $e=(g-g^{-1})/2$ for some $g\in\mathbb{Q}\setminus\{-1,0,1\}$. Let $p>3$ be a prime of good reduction. If $t\in\{1,-1\}$ is such that $tg(g-1)(g+1)$ is a quadratic residue modulo p, then the points $(x,y)\in E_d(\mathbb{F}_p)$ for which there is a $w\in\{1,-1\}$ such that*

$$y=\pm\sqrt{\frac{4tg^{2-w}}{(g-tw)^3(g+tw)}}\quad\text{and}\quad x=\pm g^w y \tag{6}$$

have order 8, *and double to* $(\pm e^{-1}, te^{-1})$.

Proof. For all points (x,y) of order 8, neither x nor y is equal to 0 or ∞. Following Theorem 2.10 of [8] we find that a point (x,y) doubles to

$$\begin{aligned}&\big((2xy:1+dx^2y^2),(x^2+y^2:1-dx^2y^2)\big)\\&\qquad=\big((2xy:-x^2+y^2),(x^2+y^2:2-(-x^2+y^2))\big).\end{aligned}$$

Let $s,t\in\{1,-1\}$ be such that (x,y) doubles to (se^{-1},te^{-1}). Then

$$\frac{2xy}{-x^2+y^2}=\frac{s}{e}\quad\text{and}\quad\frac{x^2+y^2}{2-(-x^2+y^2)}=\frac{t}{e}.$$

From the first equality we obtain $(x/y)^2+2esx/y+e^2=1+e^2$. Write $e=(g-g^{-1})/2$, so that we obtain $(x/y+se)^2=((g+g^{-1})/2)^2$. It follows that $x/y\in\{\pm g,\pm 1/g\}$, depending on the sign s and the sign after taking the square root. This gives $x^2=G^2y^2$ with $G^2\in\{g^2,g^{-2}\}$.

From the second equality we obtain $(e-t)x^2+(e+t)y^2=2t$, and substituting $x^2=G^2y^2$ results in $((e-t)G^2+(e+t))y^2=2t$. This can be solved for y

when $2t((e-t)G^2+(e+t))$ is a quadratic residue modulo p. This is equivalent to checking if either of

$$2t\big((e-1)g^2+(e+1)\big) = \frac{t(g-1)^3(g+1)}{g}, \tag{7}$$

$$2t\big((e-1)+(e+1)g^2\big) = \frac{t(g-1)(g+1)^3}{g} \tag{8}$$

is a quadratic residue modulo p. By assumption, $tg(g-1)(g+1)$ is a quadratic residue modulo p. Hence, expressions (7) and (8) are both quadratic residues modulo p. Solving for y and keeping track of all the signs results in the formulas in (6). □

Corollary 3.5. *Let $E = E_d$ be a twisted Edwards curve over $\mathbb{Q}$ such that $d = -((g-g^{-1})/2)^4$ for some $g \in \mathbb{Q} \setminus \{-1, 0, 1\}$, and let $p > 3$ be a prime of good reduction. Then $E(\mathbb{Q})$ has torsion group isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$, and the group order of $E(\mathbb{F}_p)$ is divisible by* 16.

Proof. The proof depends on the congruence class of p modulo 4.

If $p \equiv 1 \pmod 4$ then -1 is a quadratic residue modulo p. Hence, the 4-torsion points $(\pm i, 0)$ exist (see Figure 2) and $16 \mid \#E(\mathbb{F}_p)$.

If $p \equiv 3 \pmod 4$ then -1 is a quadratic nonresidue modulo p. Then exactly one of $\{g(g-1)(g+1), -g(g-1)(g+1)\}$ is a quadratic residue modulo p. Using Theorem 3.4 it follows that the curve $E(\mathbb{F}_p)$ has rational points of order 8, and hence $16 \mid \#E(\mathbb{F}_p)$. □

Corollary 3.5 explains the good behavior of the curve with $d = -\left(\frac{77}{36}\right)^4$ and torsion group isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ found in [7]. This parameter can be expressed as $d = -\left(\frac{77}{36}\right)^4 = -((g-g^{-1})/2)^4$ for $g = \frac{9}{2}$ and, therefore, the group order is divisible by an additional factor of 2.

Corollary 3.6. *Let $g \in \mathbb{Q} \setminus \{-1, 0, 1\}$, let $d = -((g-g^{-1})/2)^4$, and let $p \equiv 1 \pmod 4$ be a prime of good reduction for the curve E_d. If $g(g-1)(g+1)$ is a quadratic residue modulo p, then the group order of $E_d(\mathbb{F}_p)$ is divisible by* 32.

Proof. All 16 of the 4-torsion points are in $E_d(\mathbb{F}_p)$ (see Figure 2). By Theorem 3.4 we have at least one 8-torsion point. Hence, $32 \mid \#E_d(\mathbb{F}_p)$. □

We generated different values $g \in \mathbb{Q}$ by setting $g = \frac{i}{j}$ with $1 \le i < j \le 200$ such that $\gcd(i, j) = 1$. This resulted in 12,231 possible values for g, and Sage [30] found 614 nontorsion points. As expected, we observed that they behave similarly to the good curve found in [7].

Parametrization. In [7] a "generating curve" is specified which parametrizes d and the coordinates of the nontorsion points. Arithmetic on this generating curve can be used to generate an infinite family of twisted Edwards curves with torsion group isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ and with a nontorsion point. Using ideas from [13] we found a parametrization that does not involve a generating curve, and hence requires no curve arithmetic.

Theorem 3.7. *Let $t \in \mathbb{Q} \setminus \{0, \pm 1, \pm 3, \pm 1/3\}$ and set*

$$e = \frac{3(t^2-1)}{8t}, \quad d = -e^4, \quad x_\infty = \frac{1}{4e^3+3e}, \quad y_\infty = \frac{9t^4-2t^2+9}{9t^4-9}.$$

Then the twisted Edwards curve $-x^2 + y^2 = 1 + dx^2y^2$ has torsion subgroup isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$, and (x_∞, y_∞) is a nontorsion point.

Proof. Since $t \neq 0$ and $t \neq \pm 1$, we see that e, d, x_∞ and y_∞ are nonzero rationals; further, $e \neq \pm 1$ because $t \neq \pm 3$ and $t \neq \pm 1/3$, so $d \neq -1$. Thus, the twisted Edwards curves E_d is nonsingular, and its torsion subgroup is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ because $d = -e^4$. A calculation shows that the point (x_∞, y_∞) is on the curve; it is a nontorsion point because $x_\infty \notin \{0, \infty, e^{-1}, -e^{-1}\}$. □

This rational parametrization allowed us to impose additional conditions on the parameter e. For the four families, except $e = g^2$ which is treated below, the parameter e is given by an elliptic curve of rank 0 over $\mathbb{Q}$.

Corollary 3.8. *Let $P = (x, y)$ be a nontorsion point on the rank-1 elliptic curve $y^2 = x^3 - 36x$ over $\mathbb{Q}$. Let $t = (x+6)/(x-6)$ and let e be as in Theorem 3.7. Then the curve E_{-e^4} belongs to the family $e = g^2$ and has positive rank over $\mathbb{Q}$.*

3E. ***Better Suyama curves by a direct change of the Galois group.*** In this section we will present two families that change the Galois group of the 4- and 8-torsion without modifying the factorization pattern of the 4- and 8-division polynomial.

Suyama-11. Kruppa observed in [19] that among the Suyama curves, the one corresponding to $\sigma = 11$ finds exceptionally many primes. Barbulescu [5] extended this single example to an infinite family which we present in detail here.

Experiments show that the $\sigma = 11$ curve differs from other Suyama curves only by its probabilities to have a given 2^k-torsion group when reduced modulo primes $p \equiv 1 \pmod 4$. The reason is that the $\sigma = 11$ curve satisfies (4). Section 3C illustrates the changes in probabilities of the $\sigma = 11$ curve when compared to curves which do not satisfy (4) and shows that (4) improves the average valuation of 2 from $\frac{10}{3}$ to $\frac{11}{3}$.

We will refer to the set of Suyama curves that satisfy (4) as *Suyama*-11. When solving the system formed by Suyama's system plus (4), we obtain an elliptic

parametrization for σ. Given a point (u, v) on the curve

$$E_{\sigma_{11}}: \ v^2 = u^3 - u^2 - 120u + 432,$$

the associated σ is obtained as $\sigma = 5 + 120/(u - 24)$. The group $E_{\sigma_{11}}(\mathbb{Q})$ is generated by the points $P_\infty = (-6, 30)$, $P_2 = (-12, 0)$, and $Q_2 = (4, 0)$ of orders ∞, 2, and 2, respectively. We exclude 0, $\pm P_\infty$, P_2, Q_2, $P_2 + Q_2$, and $Q_2 \pm P_\infty$, which are the points producing invalid values of σ. The points $\pm R$, $Q_2 \pm R$ lead to isomorphic curves. Note that the $\sigma = 11$ curve corresponds to the point $(44, 280) = P_\infty + P_2$.

Edwards $\mathbb{Z}/6\mathbb{Z}$: Suyama-11 in disguise. In [7, §5] it is shown that the $a = -1$ twisted Edwards curves with $\mathbb{Z}/6\mathbb{Z}$-torsion over $\mathbb{Q}$ are precisely the curves E_d with

$$d = -\frac{16u^3(u^2 - u + 1)}{(u - 1)^6(u + 1)^2} \tag{9}$$

where u is a rational parameter.[1] In particular, according to [7, §5.3] one can translate any Suyama curve into Edwards language and then impose the condition that $-a$ is a square to obtain curves of the $a = -1$ type. Finally, [7, §5.5] points out that this family has exceptional torsion properties.

In order to understand the properties of this family, we translate it back into Montgomery language using Remark 3.1. Thus, we are interested in Suyama curves that satisfy the equation $A + 2 = -Bc^2$ (the Montgomery equivalent for $-a$ being a square). This is the Suyama-11 family, so its torsion properties were explained on page 81. These two families have been discovered independently in [5] and [7].

Suyama-$\frac{9}{4}$. In experiments by Zimmermann, new Suyama curves with exceptional torsion properties were discovered, such as $\sigma = \frac{9}{4}$. Further experiments show that their special properties are related to the 2^k-torsion and exclusively concern primes $p \equiv 1 \pmod 4$. Indeed, the $\sigma = \frac{9}{4}$ curve with satisfies (5). Section 3C illustrates the changes in probabilities of that curve when compared to curves which do not satisfy (5), and shows that (5) improves the average valuation of 2 from $\frac{10}{3}$ to $\frac{11}{3}$.

We refer to the set of Suyama curves satisfying (5) as *Suyama-$\frac{9}{4}$*. When solving the system formed by Suyama's system together with (5), we obtain an elliptic parametrization for σ. Given a point (u, v) on the curve

$$E_{\sigma_{9/4}}: \ v^2 = u^3 - 5u,$$

the associated σ is obtained as $\sigma = u$. The group $E_{\sigma_{9/4}}(\mathbb{Q})$ is generated by the points $P_\infty = (-1, 2)$ and $P_2 = (0, 0)$ of orders ∞ and 2, respectively. We exclude

[1] In the proof of [7, Theorem 5.1], the fraction corresponding to (9) is missing a minus sign.

the points 0, $\pm P_\infty$, P_2, and $P_2 \pm P_\infty$, which produce invalid values of σ. If two points in $E_{\sigma_{9/4}}(\mathbb{Q})$ differ by P_2 they correspond to isomorphic curves. The curve associated to $\sigma = \frac{9}{4}$ is obtained from the point $\left(\frac{9}{4}, -\frac{3}{8}\right) = [2]P_\infty$.

3F. ***Comparison.*** Table 4 gives a summary of all the families discussed in this article. The theoretical average valuations were computed with Theorem 2.16, Theorem 2.5, and Corollary 2.7, under some assumptions on Serre's exponent (see Example 2.17 for more information).

Note that, when we impose torsion points over $\mathbb{Q}$, the average valuation does not simply increase by 1, as can be seen in Table 4 for the average valuation of 3.

Family	Curve	n_2	$\overline{v}_{2,\text{theor}}$ $\overline{v}_{2,\text{exper}}$	n_3	$\overline{v}_{3,\text{theor}}$ $\overline{v}_{3,\text{exper}}$
Suyama	$\sigma = 12$	2	$\frac{10}{3} \approx 3.333$ 3.331	1	$\frac{27}{16} \approx 1.688$ 1.689
Suyama-11	$\sigma = 11$	2	$\frac{11}{3} \approx 3.667$ 3.369	1	$\frac{27}{16} \approx 1.688$ 1.687
Suyama-$\frac{9}{4}$	$\sigma = \frac{9}{4}$	3	$\frac{11}{3} \approx 3.667$ 3.364	1	$\frac{27}{16} \approx 1.688$ 1.687
$\mathbb{Z}/2 \times \mathbb{Z}/4\mathbb{Z}$ (Twisted Edwards E_{-e^4})	$e = 11$	3	$\frac{14}{3} \approx 4.667$ 4.666	1^*	$\frac{87}{128} \approx 0.680$ 0.679
$e = (g - g^{-1})/2$	$g = \frac{9}{2}$	3	$\frac{16}{3} \approx 5.333$ 5.332	1^*	$\frac{87}{128} \approx 0.680$ 0.679
$e = g^2$	$g = 3$	3	$\frac{29}{6} \approx 4.833$ 4.833	1^*	$\frac{87}{128} \approx 0.680$ 0.680
$e = g^2/2$	$g = \frac{9}{2}$	3	$\frac{29}{6} \approx 4.833$ 4.831	1^*	$\frac{87}{128} \approx 0.680$ 0.679
$e = \frac{2g^2+2g+1}{2g+1}$	$g = 1$	3	$\frac{29}{6} \approx 4.833$ 4.833	1^*	$\frac{87}{128} \approx 0.680$ 0.679

Table 4. Theoretical and experimental values of $\overline{v}_2$ and $\overline{v}_3$ for sample curves from the families discussed in this paper. The theoretical values come from Theorem 2.16, and the experimental values were computed using all primes less than 2^{25}. The columns labeled n_2 and n_3 give the values of $n(E, 2)$ and $n(E, 3)$. The notation $n = 1^*$ means that the Galois group is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/\pi\mathbb{Z})$.

4. Conclusion and further work

We have used Galois theory in order to analyze the torsion properties of elliptic curves. We have determined the behavior of generic elliptic curves and explained the exceptional properties of some known curves (Edwards curves of torsion $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$). The new techniques suggested by the theoretical study have helped us to find infinite families of curves having exceptional torsion properties. We list some questions which were not addressed in this work:

- How does Serre's work relate to the independence of the m- and m'-torsion probabilities for coprime integers m and m'?
- Is there a model predicting the success probability of ECM from the probabilities given in Theorem 2.16?
- Is it possible to effectively use the resolvent method [14] in order to compute equations which improve the torsion properties?

Acknowledgments

We thank Andrew Sutherland for bringing the article of Zywina [36] to our attention. We are also indebted to Everett Howe and Kiran Kedlaya for their editing effort and for the corrections they brought to this paper. This work was supported by the Swiss National Science Foundation under grant number 200020-132160 and by a PHC Germaine de Staël grant.

References

[1] Michel Abdalla and Paulo S. L. M. Barreto (eds.), *Progress in cryptology — LATINCRYPT* 2010: *Proceedings of the* 1*st International Conference on Cryptology and Information Security in Latin America held in Puebla*, *August* 8–11, 2010, Lecture Notes in Computer Science, no. 6212, Berlin, Springer, 2010.

[2] Elisardo Antelo, David Hough, and Paolo Ienne (eds.), *Proceedings of the* 20*th IEEE Symposium on Computer Arithmetic*: *ARITH*-20, Los Alamitos, CA, Institute of Electrical and Electronics Engineers, IEEE Computer Society, 2011.

[3] A. O. L. Atkin and F. Morain, *Finding suitable curves for the elliptic curve method of factorization*, Math. Comp. **60** (1993), no. 201, 399–405. MR 93k:11115

[4] Shi Bai, Pierrick Gaudry, Alexander Kruppa, François Morain, Emmanuel Thomé, and Paul Zimmermann, *Crible algébrique*: *Distribution*, *optimisation — number field sieve* (*CADO-NFS*). http://cado-nfs.gforge.inria.fr/

[5] Razvan Barbulescu, *Familles de courbes adaptées à la factorisation des entiers*, research report 00419218, version 2, INRIA, 2009. http://hal.inria.fr/inria-00419218/en/

[6] Daniel J. Bernstein, Peter Birkner, Marc Joye, Tanja Lange, and Christiane Peters, *Twisted Edwards curves*, in Vaudenay [33], 2008, pp. 389–405. MR 2010e:11057

[7] Daniel J. Bernstein, Peter Birkner, and Tanja Lange, *Starfish on strike*, in Abdalla and Barreto [1], 2010, pp. 61–80.

[8] Daniel J. Bernstein, Peter Birkner, Tanja Lange, and Christiane Peters, *ECM using Edwards curves*, Cryptology ePrint Archive, report 2008/016, 2008. http://eprint.iacr.org/2008/016

[9] Daniel J. Bernstein and Tanja Lange, *Faster addition and doubling on elliptic curves*, in Kurosawa [20], 2007, pp. 29–50. MR 2011d:11125

[10] Yu. Bilu, P. Parent, and M. Rebolledo, *Rational points on $X_0^+(p^r)$*, 2011. arXiv 1104.4641 [math.NT]

[11] Yuri Bilu and Pierre Parent, *Serre's uniformity problem in the split Cartan case*, Ann. of Math. (2) **173** (2011), no. 1, 569–584. MR 2012a:11077

[12] J. W. Bos, T. Kleinjung, A. K. Lenstra, and P. L. Montgomery, *Efficient SIMD arithmetic modulo a Mersenne number*, in Antelo et al. [2], 2011, pp. 213–221.

[13] Éric Brier and Christophe Clavier, *New families of ECM curves for Cunningham numbers*, in Hanrot et al. [16], 2010, pp. 96–109. MR 2011m:11243

[14] Henri Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics, no. 138, Springer, Berlin, 1993. MR 94i:11105

[15] Harold M. Edwards, *A normal form for elliptic curves*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 3, 393–422. MR 2008b:14052

[16] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory*: *Proceedings of the* 9*th Biennial International Symposium* (*ANTS-IX*) *held in Nancy*, *July* 19–23, 2010, Lecture Notes in Computer Science, no. 6197, Berlin, Springer, 2010. MR 2011g:11002

[17] Florian Hess, Sebastian Pauli, and Michael Pohst (eds.), *Algorithmic number theory*: *Proceedings of the* 7*th International Symposium* (*ANTS-VII*) *held at the Technische Universität Berlin*, *Berlin*, *July* 23–28, 2006, Lecture Notes in Computer Science, no. 4076, Berlin, Springer, 2006. MR 2007h:11001

[18] Huseyin Hisil, Kenneth Koon-Ho Wong, Gary Carter, and Ed Dawson, *Twisted Edwards curves revisited*, in Pieprzyk [25], 2008, pp. 326–343. MR 2546103

[19] Alexander Kruppa, *Speeding up integer multiplication and factorization*, Ph.D. thesis, Université Henri Poincaré — Nancy I, 2010. http://tel.archives-ouvertes.fr/tel-00477005/en/

[20] Kaoru Kurosawa (ed.), *Advances in cryptology — ASIACRYPT* 2007: *Proceedings of the* 13*th International Conference on the Theory and Application of Cryptology and Information Security held in Kuching*, *December* 2–6, 2007, Lecture Notes in Computer Science, no. 4833, Berlin, Springer, 2007. MR 2010i:94001

[21] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The development of the number field sieve*, Lecture Notes in Mathematics, no. 1554, Springer, Berlin, 1993. MR 96m:11116

[22] H. W. Lenstra, Jr., *Factoring integers with elliptic curves*, Ann. of Math. (2) **126** (1987), no. 3, 649–673. MR 89g:11125

[23] Peter L. Montgomery, *Speeding the Pollard and elliptic curve methods of factorization*, Math. Comp. **48** (1987), no. 177, 243–264. MR 88e:11130

[24] Jürgen Neukirch, *Class field theory*, Grundlehren der mathematischen Wissenschaften, no. 280, Springer, Berlin, 1986. MR 87i:11005

[25] Josef Pieprzyk (ed.), *Advances in cryptology — ASIACRYPT* 2008: *Proceedings of the* 14*th International Conference on the Theory and Application of Cryptology and Information Security held in Melbourne*, *December* 7–11, 2008, Lecture Notes in Computer Science, no. 5350, Berlin, Springer, 2008. MR 2010j:94005

[26] J. M. Pollard, *The lattice sieve*, in Lenstra and Lenstra [21], 1993, pp. 43–49. MR 1321220

[27] Jean-Pierre Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Invent. Math. **15** (1972), no. 4, 259–331. MR 52 #8126

[28] ———, *Quelques applications du théorème de densité de Chebotarev*, Inst. Hautes Études Sci. Publ. Math. (1981), no. 54, 323–401. MR 83k:12011

[29] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics, no. 106, Springer, Dordrecht, 2009. MR 2010i:11005

[30] W. A. Stein et al., *Sage Mathematics Software* (*version* 4.7), The Sage Development Team, 2011. http://www.sagemath.org

[31] Andrew Sutherland, *Computing the image of Galois*, presentation at the 12th meeting of the Canadian Number Theory Association held in Lethbridge, June 17–22, 2012. http://math.mit.edu/~drew/CNTA12.pdf

[32] Hiromi Suyama, *Informal preliminary report* (8), personal communication to Richard Brent, 1985.

[33] Serge Vaudenay (ed.), *Progress in cryptology — AFRICACRYPT* 2008: *Proceedings of the* 1*st International Conference on Cryptology in Africa held in Casablanca*, *June* 11–14, 2008, Lecture Notes in Computer Science, no. 5023, Berlin, Springer, 2008. MR 2009m:94064

[34] Paul Zimmermann and Bruce Dodson, 20 *years of ECM*, in Hess et al. [17], 2006, pp. 525–542. MR 2007j:11172

[35] Paul Zimmermann et al., *GMP-ECM* (*Elliptic curve method for integer factorization*), 2010. https://gforge.inria.fr/projects/ecm/

[36] David Zywina, *On the surjectivity of mod ℓ representations associated to elliptic curves*, preprint, 2011. http://www.mast.queensu.ca/~zywina/papers/EffectiveModl.pdf

RAZVAN BARBULESCU: razvan.barbulescu@inria.fr
Université de Lorraine, LORIA - Bât. A, Équipe CARAMEL, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy, France

JOPPE W. BOS: jbos@microsoft.com
Microsoft Research, One Microsoft Way, Redmond, WA 98052, United States

CYRIL BOUVIER: cyril.bouvier@inria.fr
ENS Paris and Université de Lorraine, LORIA - Bât. A, Équipe CARAMEL, Campus Scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy, France

THORSTEN KLEINJUNG: thorsten.kleinjung@epfl.ch
EPFL, Laboratory for Cryptologic Algorithms, CH-1015 Lausanne, Switzerland

PETER L. MONTGOMERY: pmontgom@math.ucla.edu
Microsoft Research, One Microsoft Way, Redmond, WA 98052, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.87

msp

Two grumpy giants and a baby

Daniel J. Bernstein and Tanja Lange

Pollard's rho algorithm, along with parallelized, vectorized, and negating variants, is the standard method to compute discrete logarithms in generic prime-order groups. This paper presents two reasons that Pollard's rho algorithm is farther from optimality than generally believed. First, "higher-degree local anticollisions" make the rho walk less random than the predictions made by the conventional Brent-Pollard heuristic. Second, even a truly random walk is suboptimal, because it suffers from "global anticollisions" that can at least partially be avoided. For example, after $(1.5+o(1))\sqrt{\ell}$ additions in a group of order ℓ (without fast negation), the baby-step-giant-step method has probability $0.5625+o(1)$ of finding a uniform random discrete logarithm; a truly random walk would have probability $0.6753\ldots+o(1)$; and this paper's new two-grumpy-giants-and-a-baby method has probability $0.71875+o(1)$.

1. Introduction

Fix a prime ℓ. The discrete-logarithm problem for a group G of order ℓ is the problem of finding $\log_g h$, given a generator g of G and an element h of G. The notation $\log_g h$ means the unique $s \in \mathbb{Z}/\ell$ such that $h = g^s$, where G is written multiplicatively.

The difficulty of finding discrete logarithms depends on G. For example, if G is the additive group $\mathbb{Z}/\ell$ (encoded as bit strings representing $\{0, 1, \ldots, \ell-1\}$ in the usual way), then $\log_g h$ is simply h/g, which can be computed in polynomial time using the extended Euclidean algorithm. As a more difficult example, consider the case that $p = 2\ell+1$ is prime and G is the order-ℓ subgroup of the multiplicative group $\mathbb{F}_p^*$ (again encoded in the usual way); index-calculus attacks then run in time subexponential in p and thus in ℓ. However, if G is the order-ℓ subgroup of $\mathbb{F}_p^*$ where $p-1$ is a much larger multiple of ℓ, then index-calculus attacks become

MSC2010: 11Y16.
Keywords: Pollard rho, baby-step giant-step, discrete logarithms, complexity.

much slower in terms of ℓ; the standard algorithms are then the baby-step-giant-step method, using at most $(2+o(1))\sqrt{\ell}$ multiplications in G, and the rho method, which if tweaked carefully uses on average $(\sqrt{\pi/2}+o(1))\sqrt{\ell}$ multiplications in G.

This paper focuses on generic discrete-logarithm algorithms such as the baby-step-giant-step method and the rho method. "Generic" means that these algorithms work for any order-ℓ group G, using oracles to compute $1 \in G$ and to compute $a, b \mapsto ab$ for any $a, b \in G$. See Section 2 for a precise definition.

If G is an elliptic-curve group chosen according to standard criteria then the best discrete-logarithm algorithms available are variants of the baby-step-giant-step method and the rho method, taking advantage of the negligible cost of computing inverses in G. There is a standard "inverting" (or "negating") variant of the concept of a generic algorithm, also discussed in Section 2. This paper emphasizes the noninverting case, but all of the ideas can be adapted to the inverting case.

Measuring algorithm cost. The most fundamental metric for generic discrete-logarithm algorithms, and the metric used throughout this paper, is the probability of discovering a uniform random discrete logarithm within m multiplications. By appropriate integration over m one obtains the average number of multiplications to find a discrete logarithm, the variance, and so on. We caution the reader that comparing probabilities of two algorithms for one m can produce different results from comparing averages, maxima, and so forth; for example, the rho method is faster than baby-step-giant-step on average but much slower in the worst case.

One can interpret a uniform random discrete logarithm as $\log_g h$ for a uniform random pair (g, h), or as $\log_g h$ for a fixed g and a uniform random h. The following trivial "worst-case-to-average-case reduction" shows that a worst-case discrete logarithm is at most negligibly harder than a uniform random discrete logarithm: One computes $\log_g h$ as $\log_g h' - r$ where $h' = hg^r$ for a uniform random $r \in \mathbb{Z}/\ell$.

There are many reasons that simply counting multiplications, the number m above, does not adequately capture the cost of these algorithms:

- A multiplication count ignores overhead; that is, the costs of computations other than multiplications. For example, the ongoing ECC2K-130 computation uses a very restricted set of Frobenius powers, sacrificing approximately 2% in the number of multiplications, because this reduces the overhead enough to speed up the entire computation.
- A multiplication count ignores issues of memory usage. For some algorithms, such as the baby-step-giant-step method, memory usage grows with $\sqrt{\ell}$, while for others, such as the rho method, memory usage is constant (or near-constant).
- A multiplication count is blind to optimizations of the multiplication operation. The question here is not simply how fast multiplication can be, but how multiplication algorithms interact with higher-level choices in these algorithms.

For example, Cheon, Hong, and Kim in [10] showed how to look ahead one step in the rho method for $\mathbb{F}_p^*$ and combine two multiplications into one at the expense of very little overhead, although memory usage increases.

- A multiplication count ignores issues of parallelization. Pollard's original rho method is difficult to parallelize effectively, but "distinguished point" variants of the rho method are heavily parallelizable with little overhead.
- A multiplication count ignores issues of vectorization. Modern processors can operate on a vector of words in one clock cycle, but this requires that the operation be the same across the entire vector. This issue was raised in a recent discussion of whether the negation map on an elliptic curve can actually be used to speed up the rho method, rather than merely to save multiplications; see [6] and [3] for the two sides of the argument.

An improvement in multiplication counts does not necessarily indicate an improvement in more sophisticated cost metrics. It is nevertheless reasonable to begin with an analysis of multiplication counts, as is done in a large fraction of the literature; followup analyses can then ask whether improved multiplication counts are still achievable by algorithms optimized for other cost metrics.

Contents of this paper. Brent and Pollard in [7] identified a source of nonrandomness in the rho method, and quantified the loss of success probability produced by this nonrandomness, under plausible heuristic assumptions. The Brent-Pollard nonrandomness (with various simplifications and in various special cases) has been stated by many authors as the main deficiency in the rho method, and the rho method has been the workhorse of large-scale discrete-logarithm computations. There appears to be a widespread belief that, except for the Brent-Pollard nonrandomness, the rho method is the best conceivable generic discrete-logarithm algorithm. Of course, the rho method can take more than $2\sqrt{\ell}$ multiplications in the worst case while the baby-step-giant-step method is guaranteed to finish within $2\sqrt{\ell}$ multiplications, but the rho method is believed to be the best way to spend a significantly smaller number of multiplications.

This paper shows that there are actually at least two more steps separating the rho method from optimality. First, the rho method is actually less random and less successful than the Brent-Pollard prediction, because the rho method suffers from a tower of what we call "local anticollisions"; Brent and Pollard account only for "degree-1 local anticollisions". Second, and more importantly, the rho method would not be optimal even if it were perfectly random, because it continues to suffer from what we call "global anticollisions". We introduce a new "two grumpy giants and a baby" algorithm that avoids many of these global anticollisions.

This new algorithm, like the original baby-step-giant-step algorithm, has low overhead but high memory. We have not found a low-memory variant. This means

that, for the moment, the algorithm is useful only for discrete-logarithm problems small enough to fit into fast memory. The algorithm nevertheless challenges the idea that the rho method is optimal for larger problems. The same approach might also be useful for "implicit" discrete-logarithm problems in which rho-type iteration is inapplicable, such as stage 2 of the $p-1$ factorization method, but those problems involve many overheads not considered in this paper.

Section 2 describes the general concept of anticollisions. Section 3 reviews the Brent-Pollard nonrandomness. Section 4 discusses higher-degree anticollisions in the rho method. Section 5 reports computations of optimal discrete-logarithm algorithms for small ℓ. Section 6 presents our new algorithm.

2. Anticollisions

This section introduces the concept of anticollisions in generic discrete-logarithm algorithms. This section begins by reviewing one of the standard ways to define such algorithms; readers familiar with the definition should still skim it to see our notation.

Generic discrete-logarithm algorithms. The standard way to formalize the idea that a generic algorithm works for any order-ℓ group G is to give the algorithm access to an oracle that computes $1 \in G$ and an oracle that computes the function $a, b \mapsto ab$ from $G \times G$ to G. The elements of G are encoded as a size-ℓ set $\overline{G}$ of strings.

An m-multiplication generic algorithm is one that calls the $a, b \mapsto ab$ oracle m times. The algorithm obtains 1 for free, and has g and h as inputs, so overall it sees $m+3$ group elements. We write $w_0 = 1$, $w_1 = g$, $w_2 = h$, and w_i for $i \geq 3$ as the $(i-2)$nd output of the $a, b \mapsto ab$ oracle: In other words, $w_i = w_j w_k$ for some $j, k \in \{0, 1, \ldots, i-1\}$ computed by the algorithm as functions of $w_0, w_1, \ldots, w_{i-1}$. These functions can also flip coins (that is, take as an additional input a sequence $b_0, b_1, \ldots$ of uniform random bits that are independent of each other, of g, of h, and so on.), but cannot make oracle calls.

The standard way to formalize the idea that a generic algorithm does not take advantage of the structure of G is to hide this structure by randomizing it. For example, one can take G as the additive group $\mathbb{Z}/\ell$, and take $\overline{G}$ as the usual binary representation of $\{0, 1, \ldots, \ell-1\}$, but choose a uniform random injection from G to $\overline{G}$ rather than the usual encoding. One defines the *generic* success probability of a generic algorithm by averaging not only over $\log_g h$ but also over the choices of this injection.

To allow inverting algorithms one also allows free access to an oracle that computes $a \mapsto 1/a$. Equivalently, one allows the algorithm to compute w_i as either $w_j w_k$ or w_j / w_k, and one also provides $1/w_i$. Of course, one can simulate this inversion oracle using approximately $\log_2 \ell$ calls to the multiplication oracle, since

$1/a = a^{\ell-1}$; an algorithm that uses only a small number of inversions can thus be simulated at negligible cost without inversions.

Slopes. Each w_i can be written as $h^{x_i} g^{y_i}$ for a pair $(x_i, y_i) \in (\mathbb{Z}/\ell)^2$ trivially computable by the algorithm. Specifically, $w_0 = 1 = h^{x_0} g^{y_0}$ where $(x_0, y_0) = (0, 0)$; $w_1 = g = h^{x_1} g^{y_1}$ where $(x_1, y_1) = (0, 1)$; $w_2 = h = h^{x_2} g^{y_2}$ where $(x_2, y_2) = (1, 0)$; if w_i is computed as $w_j w_k$ then $w_i = h^{x_i} g^{y_i}$ where $(x_i, y_i) = (x_j, y_j) + (x_k, y_k)$; and if an inverting algorithm computes w_i as w_j / w_k then $w_i = h^{x_i} g^{y_i}$ where $(x_i, y_i) = (x_j, y_j) - (x_k, y_k)$.

Normally these algorithms find $\log_g h$ by finding collisions in the map

$$(x, y) \mapsto h^x g^y$$

from $(\mathbb{Z}/\ell)^2$ to G. A collision $h^{x_i} g^{y_i} = h^{x_j} g^{y_j}$ with $(x_i, y_i) \neq (x_j, y_j)$ must have $x_i \neq x_j$ (otherwise $g^{y_i} = g^{y_j}$ so $y_i = y_j$ since g generates G), so the negative of the slope $(y_j - y_i)/(x_j - x_i)$ is exactly $\log_g h$. The discrete logarithms found by $w_0, w_1, \ldots, w_{m+2}$ are thus exactly the negatives of the $(m+3)(m+2)/2$ slopes (excluding any infinite slopes) between the $m+3$ points $(x_0, y_0), \ldots, (x_{m+2}, y_{m+2})$ in $(\mathbb{Z}/\ell)^2$. The number of discrete logarithms found in this way is the number d of distinct non-infinite slopes. The generic chance of encountering such a collision is exactly d/ℓ.

In the remaining cases, occurring with probability $1 - d/\ell$, these algorithms simply guess $\log_g h$. The success chance of this guess is 0 if the guess matches one of the negated slopes discussed above; otherwise the conditional success chance of this guess is $1/(\ell - d)$, so the success chance of this guess is $1/\ell$. The overall generic success chance of the algorithm is thus between d/ℓ and $(d+1)/\ell$, depending on the strategy for this final guess. In the extreme case $d = \ell$ this guess does not exist and the generic success chance is 1.

(Similar comments apply to inverting algorithms, but the bound on d is doubled, because there are twice as many opportunities to find $-\log_g h$. Specifically, comparing w_j to w_i finds the slope $(y_j - y_i)/(x_j - x_i)$, while comparing w_j to $1/w_i$ finds $(y_j + y_i)/(x_j + x_i)$.)

A similar model for generic discrete-logarithm algorithms was introduced by Shoup in [23], along with the bound $O(m^2/\ell)$ on the generic success probability of m-multiplication algorithms. Nechaev in [15] three years earlier had proven the collision-probability bound $O(m^2/\ell)$ in a weaker model, where algorithms are permitted only to remotely manipulate group elements without inspecting strings representing the group elements. Nechaev's model is equivalent to Shoup's model when one measures algorithm cost as the number of multiplications, but is more restrictive than Shoup's model in more sophisticated cost metrics; for example, Nechaev's model is unable to express the rho algorithm.

Chateauneuf, Ling, and Stinson in [9] introduced the idea of counting distinct slopes. They pointed out that the success probability of the baby-step-giant-step method is a factor $2 + o(1)$ away from the obvious quantification of the Nechaev-Shoup bound: m multiplications allow only $m/2$ baby steps and $m/2$ giant steps (if m is even), producing $(m/2+2)(m/2+1) \approx m^2/4$ slopes, while one can imagine $m+3$ points in $(\mathbb{Z}/\ell)^2$ potentially having as many as $(m+3)(m+2)/2 \approx m^2/2$ distinct slopes.

Computer searches reported in [9, Section 3] found for each $\ell < 100$ a set of only marginally more than $\sqrt{2\ell}$ points with slopes covering $\mathbb{Z}/\ell$. However, these sets of points do not form addition chains, and as far as we can tell the shortest addition chains for all of the constructions in [9] are worse than the baby-step-giant-step method in the number of multiplications used. The cost model used in [9] allows $a, b \mapsto a^s b^t$ as a single oracle call for any (s, t); we view that cost model as excessively simplified, and are skeptical that algorithms optimized for that cost model will be of any use in practice.

Anticollisions. We use the word "anticollision" to refer to an appearance of a useless slope — a slope that cannot create a new collision because the same slope has appeared before. Formally, an anticollision is a pair (i, j) with $i > j$ such that either

- $x_i = x_j$ or
- $(y_j - y_i)/(x_j - x_i)$ equals $(y_{j'} - y_{i'})/(x_{j'} - x_{i'})$ for some pair (i', j') lexicographically smaller than (i, j) with $i' > j'$.

The number of anticollisions is exactly the gap $(m+3)(m+2)/2 - d$, where as above d is the number of distinct non-infinite slopes. Our objective in this paper is to understand *why* anticollisions occur in addition chains in $(\mathbb{Z}/\ell)^2$, and how these anticollisions can be avoided.

In Section 3 we review a standard heuristic by Brent and Pollard that can be viewed as identifying some anticollisions in the rho method, making the rho method somewhat less effective than a truly random walk would be. In Section 4 we identify a larger set of anticollisions in the rho method, making the rho method even less effective than predicted by Brent and Pollard. This difference is most noticeable for rho walks that use a very small number of steps, such as hardware-optimized walks or typical walks on equivalence classes modulo Frobenius on Koblitz curves.

It should be obvious that even a truly random walk produces a large number of anticollisions when m grows to the scale of $\sqrt{\ell}$. In Section 6 we show that at least a constant fraction of these anticollisions can be eliminated: We construct an explicit and efficient addition chain with significantly fewer anticollisions, and thus significantly higher success probability, than a truly random walk.

3. Review of the Brent-Pollard nonrandomness

This section reviews the nonrandomness that Brent and Pollard pointed out in the rho method. The literature contains three formulas for this nonrandomness, in three different levels of generality, backed by two different heuristic arguments. As discussed in Section 4, these heuristics account for "degree-1 local anticollisions" but do not account for "higher-degree local anticollisions".

The rho method. The rho method precomputes r distinct "steps"

$$s_1, s_2, \ldots, s_r \in G - \{1\}$$

(as some initial w's), and then moves from w_i to $w_{i+1} = w_i s_j$, where j is a function of w_i. Write p_j for the probability that step s_j is used.

We suppress standard details of efficient parallelization and collision detection here, since our emphasis is on the success probability achieved after m multiplications. Inserting each new group element into an appropriate data structure will immediately recognize the first collision without consuming any multiplications.

The $\sqrt{V}$ formula. Brent and Pollard in [7, Section 2] introduced the following heuristic argument, concluding that if the values $w_0, \ldots, w_m$ are distinct then w_{m+1} collides with one of those values with probability approximately mV/ℓ, where V is defined below. This implies that the total chance of a collision within m multiplications (that is, within $w_0, \ldots, w_{m+2}$) is approximately $1-(1-V/\ell)^{m^2/2}$, which in turn implies that the average number of multiplications for a collision is approximately $\sqrt{\pi/2}\sqrt{\ell}/\sqrt{V}$. For comparison, a truly random walk would have $V = 1$.

This argument applies to a more general form of the rho method, in which some function F is applied to w_i to produce w_{i+1}. The first collision might be unlucky enough to involve w_0, but otherwise it has the form $w_{i+1} = w_{j+1}$ with $w_i \neq w_j$, revealing a collision $F(w_i) = F(w_j)$ in the function F. Applications vary in how they construct F and in the use that they make of a collision.

Assume, heuristically, that the probability of w_i matching any particular value y is proportional to the number of preimages of y; in other words, assume that $\Pr[w_i = y] = \#F^{-1}(y)/\ell$, where $F^{-1}(y)$ means $\{x : F(x) = y\}$. This heuristic is obviously wrong for w_0, but this is a minor error in context; the heuristic seems plausible for $w_1, \ldots, w_m$, which are each generated as outputs of F.

Assume that $w_0, \ldots, w_m$ are distinct. Define X as the set of preimages of $w_1, \ldots, w_m$, so that X is the disjoint union of $F^{-1}(w_1), \ldots, F^{-1}(w_m)$. Then the expected size of X is

$$\sum_x \Pr[x \in X] = \sum_x \sum_i \Pr[F(x) = w_i] = \sum_x \sum_i \sum_y \Pr[F(x) = y \text{ and } w_i = y].$$

Assume, heuristically, that $F(x) = y$ and $w_i = y$ are independent events. Then

$$\begin{aligned}\sum_x \Pr[x \in X] &= \sum_i \sum_y \sum_x \Pr[F(x) = y] \Pr[w_i = y] \\ &= \sum_i \sum_y \#F^{-1}(y)^2/\ell \\ &= m \sum_y \#F^{-1}(y)^2/\ell.\end{aligned}$$

Define V as the variance over y of $\#F^{-1}(y)$. The average over y of $\#F^{-1}(y)$ is 1, so $V = \left(\sum_y \#F^{-1}(y)^2/\ell\right) - 1$, so the expected size of X is $mV + m$. There are m known elements $w_0, \ldots, w_{m-1}$ of X; the expected number of elements of X other than $w_0, \ldots, w_{m-1}$ is mV. By hypothesis w_m is none of $w_0, \ldots, w_{m-1}$; if w_m were uniformly distributed subject to this constraint then it would have probability $mV/(\ell - m) \approx mV/\ell$ of being in X and thus leading to a collision in the next step.

The $\sqrt{1 - \sum_i p_i^2}$ formula. As part of [1] we introduced the following streamlined heuristic argument, concluding that the collision probability for w_{m+1} is approximately $m(1 - \sum_i p_i^2)/\ell$. This implies that the average number of multiplications for a collision is approximately $\sqrt{\pi/2}\sqrt{\ell}/\sqrt{1 - \sum_i p_i^2}$.

Fix a group element v, and let w and w' be two independent uniform random elements. Consider the event that w and w' both map to v but $w \neq w'$. This event occurs if there are distinct i, j such that the following three conditions hold simultaneously:

- $v = s_i w = s_j w'$;
- s_i is chosen for w;
- s_j is chosen for w'.

These conditions have probability $1/\ell^2$, p_i, and p_j respectively. Summing over all (i, j) gives the overall probability

$$\left(\sum_{i \neq j} p_i p_j\right)/\ell^2 = \left(\sum_{i,j} p_i p_j - \sum_i p_i^2\right)/\ell^2 = \left(1 - \sum_i p_i^2\right)/\ell^2.$$

Hence the probability of an immediate collision from w and w' is $\left(1 - \sum_i p_i^2\right)/\ell$, where we added over the ℓ choices of v.

After $m + 3$ group elements one has approximately $m^2/2$ potentially colliding pairs. If the inputs to the iteration function were independent uniformly distributed random points then the probability of success would be $1 - \left(1 - \left(1 - \sum_i p_i^2\right)/\ell\right)^{m^2/2}$ and the average number of iterations before a collision would be approximately

$\sqrt{\pi/2}\sqrt{\ell}/\sqrt{1-\sum_i p_i^2}$. The inputs to the iteration function in Pollard's rho method are not actually independent, but this has no obvious effect on the average number of iterations.

Relating the two formulas. We originally obtained the formula $\sqrt{1-\sum_i p_i^2}$ by specializing and simplifying the Brent-Pollard $\sqrt{V}$ formula as follows.

The potential preimages of y are $y/s_1, y/s_2, \ldots, y/s_r$, which are actual preimages with probabilities $p_1, p_2, \ldots, p_r$ respectively. A subset I of $\{1, 2, \ldots, r\}$ matches the set of indices of preimages with probability $\big(\prod_{i\in I} p_i\big)\big(\prod_{i\notin I}(1-p_i)\big)$, so the average of $\#F^{-1}(y)^2$ is

$$\sum_I \#I^2\Big(\prod_{i\in I} p_i\Big)\Big(\prod_{i\notin I}(1-p_i)\Big).$$

It is easy to see that most monomials (for example, $p_1p_2p_3$) have coefficient 0 in this sum; the only exceptions are linear monomials p_i, which have coefficient 1, and quadratic monomials p_ip_j with $i<j$, which have coefficient 2. The sum therefore equals

$$\sum_i p_i + 2\sum_{i,j:i<j} p_ip_j = \sum_i p_i + \Big(\sum_i p_i\Big)^2 - \sum_i p_i^2 = 2 - \sum_i p_i^2.$$

Hence $V = 1-\sum_i p_i^2$.

The $\sqrt{1-1/r}$ formula. In traditional "adding walks" (credited to Lenstra in [20, p. 66]; see also [21, p. 295] and [25]), each p_i is $1/r$, and $\sqrt{1-\sum_i p_i^2}$ is $\sqrt{1-1/r}$. This $\sqrt{1-1/r}$ formula first appeared in [25], with credit to the subsequent paper [4] by Blackburn and Murphy. The heuristic argument in [4] is the same as the Brent-Pollard argument.

Case study: Koblitz curves. The $\sqrt{1-\sum_i p_i^2}$ formula was first used to optimize walks on Koblitz curves. These walks map a curve point W to $W+\varphi^i(W)$, where φ is the Frobenius map and i is chosen as a function of the Hamming weight of the normal-basis representation of the x-coordinate of W. The Hamming weight is not uniformly distributed, and any reasonable function of the Hamming weight is also not uniformly distributed, so the $\sqrt{1-1/r}$ formula does not apply. Note that these are "multiplying walks" rather than "adding walks" (if $W = x_iH + y_iG$ then $W+\varphi^i(W) = s_ix_iH + s_iy_iG$ for certain constants $s_i \in (\mathbb{Z}/\ell)^*$), but the heuristics in this section are trivially adapted to this setting.

As a concrete example we repeat from [1] the analysis of our ongoing attack on ECC2K-130. All Hamming weights of x-coordinates of group elements are even, and experiments show that the distribution of even-weighted words of length 131 is close to the distribution of x-coordinates of group elements. Any iteration

function defined in this way therefore inevitably introduces an extra factor to the running time of

$$1/\sqrt{1-\sum_i \binom{131}{2i}^2/2^{260}} \approx 1.053211,$$

even if all 66 weights use different scalars s_i. We extract just 3 bits of weight information, using only 8 different values for the scalars, to reduce the time per iteration. The values are determined by $\mathrm{HW}(x_{P_i})/2 \bmod 8$; the distribution of $\sum_i \binom{131}{16i+2j}$ for $0 \le j \le 7$ gives probabilities

$$0.1414,\ 0.1443,\ 0.1359,\ 0.1212,\ 0.1086,\ 0.1057,\ 0.1141,\ 0.1288,$$

giving a total increase of the number of iterations by a factor of 1.069993.

4. Higher-degree local anticollisions

Consider the rho method using r "steps" $s_1, s_2, \ldots, s_r \in G$, as in the previous section. The method multiplies w_i by one of these steps to obtain w_{i+1}, multiplies w_{i+1} by one of these steps to obtain w_{i+2}, and so on.

Assume that the step w_{i+1}/w_i is different from the step w_{i+2}/w_{i+1}, but that w_{i+1}/w_i is the same as an earlier step w_{j+2}/w_{j+1}, and that w_{i+2}/w_{i+1} is the same as the step w_{j+1}/w_j. There are anticollisions $(i+1, j+2)$ and $(i+2, j+1)$, exactly the phenomenon discussed in the previous section: For example, w_{i+1} cannot equal w_{j+2} unless w_i equals w_{j+1}. There is, however, also a local anticollision $(i+2, j+2)$ not discussed in the previous section: w_{i+2} cannot equal w_{j+2} unless w_i equals w_j. The point is that the ratio w_{i+2}/w_i is a product of two steps, and the ratio w_{j+2}/w_j is a product of the same two steps in the opposite order.

We compute the heuristic impact of these "degree-2 local anticollisions", together with the degree-1 local anticollisions of Section 3, as follows. Assume for simplicity that 1, s_1, s_2, …, s_r, s_1^2, s_1s_2, …, s_1s_r, s_2^2, …, s_2s_r, …, s_{r-1}^2, $s_{r-1}s_r$, s_r^2 are distinct. Write $F(w)$ for the group element that w maps to. Fix a group element v, and consider the event that two independent uniform random group elements w, w' have $F(F(w)) = v = F(F(w'))$ with no collisions among $w, w', F(w), F(w')$. This event occurs if there are i, i', j, j' with $s_j \ne s_{j'}$ and $s_j s_i \ne s_{j'} s_{i'}$ such that the following conditions hold simultaneously:

- $v = s_j s_i w = s_{j'} s_{i'} w'$;
- $F(w) = s_i w$;
- $F(s_i w) = s_j s_i w$;
- $F(w') = s_{i'} w'$;
- $F(s_{i'} w') = s_{j'} s_{i'} w'$.

These conditions have probability $1/\ell^2$, p_i, p_j, $p_{i'}$, and $p_{j'}$ respectively. Given the first condition, the remaining conditions are independent of each other, since $w = v/(s_j s_i)$, $s_i w = v/s_j$, $w' = v/(s_{j'} s_{i'})$, and $s_{i'} w' = v/s_{j'}$ are distinct. This event thus has probability $\sum p_i p_j p_{i'} p_{j'}/\ell^2$ where the sum is over all i, j, i', j' with $s_j \neq s_{j'}$ and $s_j s_i \neq s_{j'} s_{i'}$. The complement of the sum is over all i, j, i', j' with $s_j = s_{j'}$ or $s_j s_i = s_{j'} s_{i'}$ — that is, with $j = j'$ or with $i' = j \neq j' = i$. The complement is thus

$$\sum_j p_j^2 + \sum_{i,j:i\neq j} p_i^2 p_j^2 = \sum_j p_j^2 + \Big(\sum_j p_j^2\Big)^2 - \sum_j p_j^4,$$

and the original sum is $1 - \sum_j p_j^2 - \big(\sum_j p_j^2\big)^2 + \sum_j p_j^4$. Adding over all v gives probability $\big(1 - \sum_j p_j^2 - \big(\sum_j p_j^2\big)^2 + \sum_j p_j^4\big)/\ell$ of this type of two-step collision between w and w'.

For example, if $p_i = 1/r$ for all i, then the degree-1-and-2 nonrandomness factor is $1/\sqrt{1 - 1/r - 1/r^2 + 1/r^3}$, whereas the Brent-Pollard (degree-1) nonrandomness factor is $1/\sqrt{1 - 1/r}$. These factors are noticeably different if r is small.

Beyond degree 2. More generally, a "degree-k local anticollision" $(i + k,\ j + k)$ occurs when the product of k successive steps w_{i+1}/w_i, w_{i+2}/w_{i+1}, ... matches the product of k successive steps w_{j+1}/w_j, w_{j+2}/w_{j+1}, ..., without a lower-degree local anticollision occurring. We define a "degree-(k, k') local anticollision" $(i + k,\ j + k')$ similarly.

Given the vector $(s_1, s_2, \ldots, s_r)$, one can straightforwardly compute the overall heuristic effect of local anticollisions of degree at most k, by summing the products $p_{i_1} \cdots p_{i_k} p_{i'_1} \cdots p_{i'_k}$ for which $1, s_{i_1}, s_{i'_1}, s_{i_1} s_{i_2}, s_{i'_1} s_{i'_2}, \ldots$ are distinct. Experiments indicate that the largest contribution is usually from the smallest degrees.

We emphasize that the results depend on the vector $(s_1, s_2, \ldots, s_r)$, because generic commutative-group equations such as $s_1 s_2 = s_2 s_1$ are not the only multiplicative dependencies among $s_1, s_2, \ldots, s_r$. One can check that $s_1, s_2, \ldots, s_r$ have no nongeneric multiplicative dependencies of small degree (and modify them to avoid such dependencies), but they always have medium-degree nongeneric multiplicative dependencies, including mixed-degree nongeneric multiplicative dependencies.

If $s_1, s_2, \ldots, s_r$ have only generic dependencies of degree at most k then the sum described above is expressible as a polynomial in the easily computed quantities $I_2 = \sum_j p_j^2$, $I_4 = \sum_j p_j^4$, and so forth, by a simple inclusion-exclusion argument. For example, the degree-1 nonrandomness factor is $1/\sqrt{1 - I_2}$, as in Section 3; the degree-≤ 2 nonrandomness factor is $1/\sqrt{1 - I_2 - I_2^2 + I_4}$, as explained above; the

degree-≤ 3 nonrandomness factor is $1/\sqrt{1 - I_2 - I_2^2 + I_4 - 3I_2^3 + 7I_2I_4 - 4I_6}$; the degree-$\le 4$ nonrandomness factor is

$$1/\sqrt{1-I_2-I_2^2+I_4-3I_2^3+7I_2I_4-4I_6-13I_2^4+53I_2^2I_4-56I_2I_6-17I_4^2+33I_8};$$

and so on. In the uniform case these factors are

$$1/\sqrt{1-1/r},$$
$$1/\sqrt{1-1/r-1/r^2+1/r^3},$$
$$1/\sqrt{1-1/r-1/r^2-2/r^3+7/r^4-4/r^5},$$

and so on.

Case study: $r = 6$. Hildebrand showed in [13] that almost every r-adding walk (with $p_j = 1/r$) reaches a nearly uniform distribution in $\mathbb{Z}/\ell$ within $O(\ell^{2/(r-1)})$ steps; in particular, within $o(\sqrt{\ell})$ steps for $r \ge 6$. Implementors optimizing Pollard's rho method for hardware often want r to be as small as possible to minimize overhead (the storage required for precomputed steps and the cost of accessing that storage), and in light of Hildebrand's result can reasonably choose $r = 6$. This raises the question of how random a 6-adding walk is; perhaps it is better to take a larger value of r, increasing overhead but reducing nonrandomness.

For $r = 6$, with $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ and generic $s_1, \ldots, s_6$, the heuristic nonrandomness factors are given (to 6 decimal places) in Table 1. These factors converge to approximately 1.129162 as the degree increases; see Appendix A. Evidently the Brent-Pollard heuristic captures *most* of the impact of local anticollisions for $r = 6$, but not all of the impact.

We tried 2^{32} experiments for $\ell = 1009$. Each experiment generated 6 uniform random steps $s_1, s_2, \ldots, s_6$ (without enforcing distinctness, and without any constraints on higher-degree multiplicative dependencies), carried out a random walk using $s_1, s_2, \ldots, s_6$ with equal probability, and stopped at the first collision. The average walk length was approximately 1.150076 times $\sqrt{\pi/2}\sqrt{\ell}$; note that this

Degree	Factor	Degree	Factor	Degree	Factor
1	1.095445	≤ 6	1.123767	≤ 11	1.126654
≤ 2	1.110984	≤ 7	1.124696	≤ 12	1.126926
≤ 3	1.117208	≤ 8	1.125383	≤ 13	1.127151
≤ 4	1.120473	≤ 9	1.125909	≤ 14	1.127341
≤ 5	1.122452	≤ 10	1.126322	≤ 15	1.127503

Table 1. Approximate values of heuristic nonrandomness factors for the case $r = 6$, with $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ and generic $s_1, \ldots, s_6$.

ℓ	Factor	Experiments	ℓ	Factor	Experiments
1009	1.150076	2^{32}	100000007	1.131149	2^{32}
10007	1.147874	2^{32}	1000000007	1.130194	2^{32}
100003	1.141283	2^{32}	10000000019	1.129680	2^{32}
1000003	1.136122	2^{32}	100000000003	1.129395	2^{28}
10000019	1.132946	2^{32}	1000000000039	1.129326	2^{26}

Table 2. Observed average walk length until a collision, for a uniform random walk in $\mathbb{Z}/\ell$ using 6 uniform random adding steps. "Factor" is the observed average walk length divided by $\sqrt{\pi/2}\sqrt{\ell}$, rounded to 6 digits after the decimal point. "Experiments" is the number of experiments carried out for ℓ.

does not count the multiplications used to generate $s_1, s_2, \ldots, s_6$. We then tried several larger values of ℓ; the resulting nonrandomness factors are shown in Table 2. Our heuristics predict that these numbers will converge to approximately 1.129162 as $\ell \to \infty$, rather than 1.095445.

Note that for small ℓ there is a larger chance of low-degree dependencies among the steps s_i, so it is not a surprise that smaller values of ℓ have larger nonrandomness factors. We do not know whether a quantitative analysis of this phenomenon would predict the numbers shown in Table 2 for small ℓ, or whether other phenomena also play a role.

Case study: Koblitz curves, revisited. Consider again the ECC2K-130 walk introduced in [1]. Here $\ell = 68056473384187692693232012949340998512 9$.

For $0 \le j \le 7$ define φ as the Frobenius map on the ECC2K-130 curve, and define $s_j \in \mathbb{Z}/\ell$ as $1 + 196511074115861092422032515080945363956^{j+3}$. This walk moves from P to $P + \varphi^{j+3}(P) = s_j P$ if the Hamming weight of the x-coordinate of P is congruent to $2j$ modulo 16; this occurs with probability (almost exactly) $p_j = \sum_i \binom{131}{16i+2j}/2^{130}$.

The only small-degree multiplicative dependencies among $s_0, \ldots, s_7$ are generic commutative-group equations such as $s_1 s_2 = s_2 s_1$. We already reported this in [1, Section 2] to explain why the walk is highly unlikely to enter a short cycle. We point out here that this has a larger effect, namely minimizing small-degree anticollisions. We now analyze the impact of the small-degree anticollisions that remain, those that arise from the generic commutative-group equations.

For degree 1 the nonrandomness factor is $1/\sqrt{1 - I_2} \approx 1.069993$. For degree ≤ 2 the nonrandomness factor is $1/\sqrt{1 - I_2 - I_2^2 + I_4} \approx 1.078620$. For degree ≤ 3 it is $1/\sqrt{1 - I_2 - I_2^2 - 3I_2^3 + I_4 + 7I_2 I_4 - 4I_6} \approx 1.081370$. For degree ≤ 4 it is ≈ 1.082550.

Case study: Mixed walks. The same type of analysis also applies to "mixed walks" combining noncommuting steps such as $w \mapsto ws_1$, $w \mapsto ws_2$, and $w \mapsto w^2$.

Degree	Factor	Degree	Factor	Degree	Factor
1	1.224745	≤ 5	1.285444	≤ 8	1.293067
≤ 2	1.248075	≤ 6	1.288605	≤ 9	1.294325
≤ 3	1.269973	≤ 7	1.291514	≤ 10	1.295107
≤ 4	1.277533				

Table 3. Approximate values of heuristic nonrandomness factors for the mixed walk $w \mapsto ws_1$, $w \mapsto ws_2$, and $w \mapsto w^2$, with generic s_1 and s_2 and equiprobable steps.

ℓ	Factor	Experiments	ℓ	Factor	Experiments
1009	1.292381	2^{41}	10000019	1.297130	2^{36}
10007	1.298240	2^{41}	100000007	1.297071	2^{32}
100003	1.297896	2^{40}	1000000007	1.297020	2^{32}
1000003	1.297360	2^{37}	10000000019	1.297018	2^{32}

Table 4. Observed average walk length until a collision, for a uniform random walk in $\mathbb{Z}/\ell$ using 2 uniform random adding steps and 1 doubling step. Columns have the same meaning as in Table 2.

A sequence of such steps maps w to a monomial such as $w^4 s_1 s_2^3$; we sum the products $p_{i_1} \cdots p_{i_k} p_{i'_1} \cdots p_{i'_k}$ for which the monomials corresponding to $()$, (i_1), (i'_1), (i_1, i_2), (i'_1, i'_2), and so on, are distinct. The heuristic nonrandomness factor for degree $\leq k$ is the reciprocal of the square root of this sum.

For three equiprobable steps $w \mapsto ws_1$, $w \mapsto ws_2$, and $w \mapsto w^2$, with generic s_1 and s_2, the heuristic nonrandomness factors are given (to 6 decimal places) in Table 3.

We tried experiments analogous to the 6-adding experiments described above. Each experiment generated 2 uniform random group elements s_1, s_2, carried out a random walk using $w \mapsto ws_1$, $w \mapsto ws_2$, and $w \mapsto w^2$ starting from a uniform random group element, and stopped at the first collision. Table 4 shows the resulting average walk lengths for various values of ℓ. The dependence on ℓ is much smaller here than it was in Table 2. The numerical data seems consistent with the idea that the limit of the actual nonrandomness factors as $\ell \to \infty$ matches the limit of the degree-$\leq k$ heuristic nonrandomness factors as $k \to \infty$: somewhere between 1.295 and 1.298, very far from the traditional degree-1 nonrandomness factor $\sqrt{3/2} \approx 1.224745$.

For comparison, Teske in [25, Table 5] reported using $1.776\sqrt{\ell}$ multiplications on average for 2000 experiments with the same type of walk. Teske's cycle-detection method cost a factor of approximately 1.13 in the number of multiplications, according to [25, Section 2.2], so $1.776\sqrt{\ell}$ corresponds to an observed nonrandomness factor of $1.776/(1.13\sqrt{\pi/2}) \approx 1.254$. This might seem noticeably different not just from 1.224745 but also from our 1.297. However, since the

standard deviation of random-walk lengths is on the same scale as the average, it is statistically unremarkable to see differences of a few percent after only 2000 experiments.

Optimizing asymptotics. It is frequently stated that the rho method, like a truly random walk, finishes in $(\sqrt{\pi/2}+o(1))\sqrt{\ell}$ multiplications on average.

However, the experimental results by Sattler and Schnorr [20, p. 76] and by Teske [25] showed clearly that $\sqrt{\pi/2}+o(1)$ is not achieved by small values of r, and in particular by Pollard's original rho method. The Brent-Pollard nonrandomness, and in particular the $\sqrt{1-1/r}$ formula, indicates that $\sqrt{\pi/2}+o(1)$ is not achieved by any bounded r; one must have $1/r \in o(1)$, that is, $r \to \infty$ as $\ell \to \infty$. On the other hand, if r grows too quickly then the cost of setting up r steps is nonnegligible.

This analysis does not contradict $\sqrt{\pi/2}+o(1)$. However, it does indicate that some care is required in the algorithm details, and that $\sqrt{\pi/2}+o(1)$ can be replaced by $\sqrt{\pi/2}+O(\ell^{-1/4})$ but not by $\sqrt{\pi/2}+o(\ell^{-1/4})$.

To optimize the $o(1)$ one might try choosing steps that are particularly easy to compute. For example, one might take $s_3 = s_1 s_2$, $s_4 = s_2 s_3$, and so on, where s_1, s_2 are random. We point out, however, that such choices are particularly prone to higher-degree anticollisions. We recommend taking into account not just the number of steps and the number of multiplications required to precompute those steps, but also the impact of higher-degree anticollisions.

5. Searching for better chains for small primes

If ℓ is small then by simply enumerating addition chains one can find generic discrete-logarithm algorithms that use fewer multiplications than the rho method.

This section reports, for each small prime ℓ, the results of two different computer searches. One search greedily obtained as many slopes as it could after each multiplication, deferring anticollisions as long as possible. The other search minimized the number of multiplications required to find an average slope. Chains found by such searches are directly usable in discrete-logarithm computations for these values of ℓ; perhaps they also provide some indication of what one can hope to achieve for much larger values of ℓ. These searches also show that merely counting the size of a slope cover, as in [9, Section 3], underestimates the cost of discrete-logarithm algorithms, although one can hope that the gap becomes negligible as ℓ increases.

A continuing theme in this section is that the obvious quantification of the Nechaev-Shoup bound is not tight. The bound says that an m-addition chain has $\leq (m+3)(m+2)/2$ slopes; but there is actually a gap, increasing with m, between $(m+3)(m+2)/2$ and the maximum number of slopes in an m-addition chain.

This section explains part of this gap by identifying two types of anticollisions that addition chains cannot avoid and stating an improved bound that accounts for these anticollisions. However, the improved bound is still not tight for most of these values of ℓ, and for long chains the improved bound is only negligibly stronger than the Nechaev-Shoup bound.

Greedy slopes. Define d_i as the number of distinct finite slopes among the points $(x_0, y_0), (x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i)$ in $(\mathbb{Z}/\ell)^2$. For example, the chain

$$(0,0), (0,1), (1,0), (0,2), (1,2), (1,4)$$

in $(\mathbb{Z}/7)^2$ has $(d_0, d_1, d_2, d_3, d_4, d_5) = (0, 0, 2, 3, 5, 7)$: There are 2 distinct finite slopes among $(0,0)$, $(0,1)$, $(1,0)$; 3 distinct finite slopes among $(0,0)$, $(0,1)$, $(1,0)$, $(0,2)$; 5 distinct finite slopes among $(0,0)$, $(0,1)$, $(1,0)$, $(0,2)$, $(1,2)$; and 7 distinct finite slopes among $(0,0)$, $(0,1)$, $(1,0)$, $(0,2)$, $(1,2)$, $(1,4)$.

For each prime $\ell < 128$ we computed the lexicographically maximum sequence $(d_0, d_1, \ldots)$ for all infinite addition chains starting $(0,0), (0,1), (1,0)$ in $(\mathbb{Z}/\ell)^2$. These maxima, truncated to the first occurrence of ℓ, are displayed in Table 5. For example, Table 5 lists $(0, 0, 2, 3, 5, 7)$ for $\ell = 7$, indicating that the lexicographic maximum is $(0, 0, 2, 3, 5, 7, 7, 7, 7, 7, \ldots)$: One always has $d_0 = 0$, $d_1 = 0$, and $d_2 = 2$; the maximum possible d_3 is 3; given $d_3 = 3$, the maximum possible d_4 is 5; given $d_3 = 3$ and $d_4 = 5$, the maximum possible d_5 is 7.

This computation was not quite instantaneous, because it naturally ended up computing all finite chains achieving the truncated maximum (and, along the way, all chains achieving every prefix of the truncated maximum). There are, for example, 5420 length-21 chains that match the $(d_0, d_1, \ldots)$ shown in Table 5 for $\ell = 109$.

Minimal weight. We also computed ℓ-slope addition chains of minimal weight for each prime $\ell < 48$. Here "weight" means $\sum_{i \geq 1} i\,(d_i - d_{i-1})$. Dividing this weight by ℓ produces the average, over all $s \in \mathbb{Z}/\ell$, of the number of multiplications (plus 2 to account for the inputs g and h) used to find slope s. It might make more sense to compute $(\ell - 1)$-slope addition chains of minimal weight, since a generic discrete-logarithm algorithm that finds $\ell - 1$ slopes also recognizes the remaining slope by exclusion, but the gap becomes negligible as ℓ increases.

Lexicographically maximizing $(d_0, d_1, \ldots)$, as in Table 5, does not always produce minimal-weight ℓ-slope addition chains. For example, the chain

$$(0,0), (0,1), (1,0), (0,2), (0,3), (1,3), (1,6), (2,12), (2,14), (2,16), (3,17), (4,28)$$

for $\ell = 29$ has weight 210 with $(d_0, d_1, \ldots) = (0, 0, 2, 3, 4, 7, 10, 14, 19, 23, 27, 29)$, while chains achieving the lexicographic maximum in Table 5 have weight 211. We similarly found weight 299 (compared to 300) for $\ell = 37$, weight 372 (compared

ℓ	Weight	$d_0\, d_1\, \ldots$
2	4	0 0 2
3	7	0 0 2 3
5	15	0 0 2 3 5
7	25	0 0 2 3 5 7
11	50	0 0 2 3 5 7 10 11
13	64	0 0 2 3 5 7 10 13
17	96	0 0 2 3 5 7 10 14 16 17
19	113	0 0 2 3 5 7 10 14 17 19
23	148	0 0 2 3 5 7 10 14 19 22 23
29	211	0 0 2 3 5 7 10 14 19 23 26 28 29
31	230	0 0 2 3 5 7 10 14 19 23 28 31
37	300	0 0 2 3 5 7 10 14 19 23 29 33 36 37
41	347	0 0 2 3 5 7 10 14 19 24 29 34 39 41
43	375	0 0 2 3 5 7 10 14 19 24 29 34 38 42 43
47	425	0 0 2 3 5 7 10 14 19 24 30 35 40 44 47
53	510	0 0 2 3 5 7 10 14 19 24 30 36 41 45 50 52 53
59	596	0 0 2 3 5 7 10 14 19 24 30 36 42 48 52 57 58 59
61	631	0 0 2 3 5 7 10 14 19 24 30 35 42 48 52 56 59 61
67	727	0 0 2 3 5 7 10 14 19 24 30 36 41 47 53 59 63 66 67
71	788	0 0 2 3 5 7 10 14 19 24 30 36 42 48 54 60 66 70 71
73	815	0 0 2 3 5 7 10 14 19 24 30 36 43 50 56 62 67 71 73
79	919	0 0 2 3 5 7 10 14 19 24 30 37 43 49 57 64 69 73 76 79
83	978	0 0 2 3 5 7 10 14 19 24 30 37 44 51 59 65 72 77 80 83
89	1081	0 0 2 3 5 7 10 14 19 24 30 37 44 53 60 66 74 80 84 87 89
97	1224	0 0 2 3 5 7 10 14 19 24 30 37 44 51 61 69 78 83 88 92 96 97
101	1307	0 0 2 3 5 7 10 14 19 24 30 37 45 53 60 69 76 82 89 93 97 100 101
103	1351	0 0 2 3 5 7 10 14 19 24 30 37 45 52 60 67 74 83 89 94 98 102 103
107	1422	0 0 2 3 5 7 10 14 19 24 30 37 45 53 61 70 77 84 91 96 100 104 107
109	1466	0 0 2 3 5 7 10 14 19 24 30 37 44 52 60 68 77 84 91 98 102 106 108 109
113	1536	0 0 2 3 5 7 10 14 19 24 30 37 44 52 62 70 78 86 94 99 105 109 113
127	1806	0 0 2 3 5 7 10 14 19 24 30 37 45 53 63 73 84 92 98 105 112 118 122 126 127

Table 5. For each $\ell < 128$, the lexicographically maximum $(d_0, d_1, \ldots)$. "Weight" means $\sum_{i \geq 1} i\,(d_i - d_{i-1})$.

to 375) for $\ell = 43$, and weight 423 (compared to 425) for $\ell = 47$. It is not clear whether this gap becomes negligible as ℓ increases.

Some obstructions. We explain here two simple ways that anticollisions appear in addition chains. Every addition chain produces at least a linear number of anticollisions that follow these simple patterns.

First, doubling a point (x_j, y_j) produces two anticollisions: The slopes from $2(x_j, y_j)$ to (x_j, y_j) and to $(0, 0)$ are the same as the slope from (x_j, y_j) to $(0, 0)$. Doubling another point (x_k, y_k) produces three anticollisions: The slope from $2(x_k, y_k)$ to $2(x_j, y_j)$ is the same as the slope from (x_k, y_k) to (x_j, y_j). A third doubling produces four anticollisions, and so on; doubling n points produces a total of $n(n+3)/2$ anticollisions of this type.

Second, adding (x_i, y_i) to a distinct point (x_j, y_j) produces two anticollisions: The slopes from $(x_i, y_i) + (x_j, y_j)$ to (x_i, y_i) and to (x_j, y_j) are the same as the slopes from (x_j, y_j) and from (x_i, y_i) to $(0, 0)$. Subsequently adding the same (x_i, y_i) to another point (x_k, y_k) produces three anticollisions: The slope from $(x_i, y_i) + (x_k, y_k)$ to $(x_i, y_i) + (x_j, y_j)$ is the same as the slope from (x_k, y_k) to (x_j, y_j), exactly as in Section 3.

Applying these principles easily explains the initial pattern $0, 0, 2, 3, 5, 7$ that appears in Table 5. The first addition (whether or not a doubling) must produce at least two anticollisions, and therefore produces at most one new slope to the previous three points; this explains the 3. The second addition also produces at least two anticollisions, and therefore at most two new slopes to the previous four points; this explains the 5. One might think that the next step is 8, but having only two anticollisions in each of the first three additions would imply that those three additions include at most one doubling and no other reuse of summands, for a total of at least five summands, while there are only four nonzero summands available for the first three additions.

More generally, a chain of $m \geq 2$ nontrivial additions involves $2m$ inputs selected from $m+1$ nonzero points, so there must be at least $m-1$ repetitions of inputs. These repetitions produce at least $m-2$ occurrences of three anticollisions (one doubling is free), on top of m occurrences of two anticollisions and one anticollision for the infinite slope from $(0, 0)$ to $(0, 1)$, for a total of at least $3m-1$ anticollisions, and thus a total of at most $(m+3)(m+2)/2-(3m-1) = (m^2-m+8)/2$ slopes. This explains $5, 7, 10, 14, 19$ in Table 5 but does not explain 24.

6. Two grumpy giants and a baby

This section presents the algorithm featured in the title of this paper. This algorithm is, as the name suggests, a modification to the standard baby-step-giant-step method. The modification increases the number of different slopes produced within m multiplications, and for a typical range of m increases the number beyond the effectiveness of the rho method.

In the baby-step-giant-step algorithm the baby steps compute $h^{x_i} g^{y_i}$ for $(x_i, y_i) \in (0, 0) + \{0, 1, 2, \ldots, \lceil\sqrt{\ell}\rceil\}(0, 1)$ and the giant steps compute $h^{x_i} g^{y_i}$ for $(x_i, y_i) \in (1, 0) + \{0, 1, 2, \ldots, \lfloor\sqrt{\ell}\rfloor\}(0, \lceil\sqrt{\ell}\rceil)$. The first observation is that the slopes within

one type of step are constant; the second observation is that once all steps are done all ℓ slopes appear. Our idea is to make the lines of fixed slope shorter; that is, we introduce more players. Note that introducing a second baby is not useful: Lines between the points in $(x, y) + \{0, 1, 2, \ldots, \lceil\sqrt{\ell}\rceil\}(0, 1)$ and $(0, 0) + \{0, 1, 2, \ldots, \lceil\sqrt{\ell}\rceil\}(0, 1)$ repeat each slope $\approx \sqrt{\ell}$ times. We thus need to introduce more giants to make progress.

The two-grumpy-giants-and-a-baby method is parametrized by a positive integer n, normally proportional to $\sqrt{\ell}$; the reader should imagine n being approximately $0.5\sqrt{\ell}$. The number of multiplications in the method is approximately $3n$. Here is the set of points $(x_i, y_i) \in (\mathbb{Z}/\ell)^2$ produced by the method:

$$\begin{aligned} \text{Baby}: &\ (0,0) + \{0, \ldots, n-1\}(0,1) \\ \text{Giant1}: &\ (1,0) + \{1, \ldots, n\}(0,n) \\ \text{Giant2}: &\ (2,0) - \{1, \ldots, n\}(0,n+1) \end{aligned}$$

The initial negation $(0, -(n+1))$ for Giant2 has negligible cost, approximately $\log_2 \ell$ multiplications. Choosing n and $n+1$ for the steps in the y direction for the two giants gives a good coverage of slopes since n and $n+1$ are coprime. The grumpy giants make big steps (on the scale of $\sqrt{\ell}$) and quickly walk in opposite directions away from each other. Luckily they are not minding the baby.

We now analyze the slopes covered by this method. Again it is not interesting to look at the slopes among one type of points. The slope between a point $(0, i)$ in the Baby set and a point $(1, jn)$ in the Giant1 set is $jn - i$; this means that all slopes in $\{1, \ldots, n^2\}$ are covered. The slope between $(0, i)$ in the Baby set and $(2, -j(n+1))$ in the Giant2 set is $(-j(n+1) - i)/2 \in \{-n^2 - 2n + 1, \ldots, -n - 1\}/2$; there are n^2 distinct slopes here, almost exactly covering $\{-n^2 - 2n + 1, \ldots, -n - 1\}/2$. The slope between $(1, in)$ in the Giant1 set and $(2, -j(n+1))$ in the Giant2 set is $-j(n+1) - in \in \{-2n^2 - n, \ldots, -2n - 1\}$; there are another n^2 distinct slopes here, covering about half the elements of $\{-2n^2 - n, \ldots, -2n - 1\}$.

To summarize, there are three sets of n^2 distinct slopes here, all between $-2n^2 - n + 1$ and n^2. One can hope for a total of $3n^2$ distinct slopes if $\ell > 3n^2 + n$, but this hope runs into two obstacles. The first obstacle is that the "odd" elements of $\{-n^2 - 2n + 1, \ldots, -n - 1\}$ can bump into the other sets when computing $(2i + 1)/2 = i + (\ell + 1)/2$; but for $\ell \in 4n^2 + O(n)$ this effect loses only $O(n)$ elements. The second obstacle is that any Giant1–Giant2 slopes between $(-n^2 - 2n)/2$ and $(-n - 2)/2$ will bump into $\{-n^2 - 2n + 1, \ldots, -n - 1\}/2$ for the "even" elements of $\{-n^2 - 2n + 1, \ldots, -n - 1\}$. This is approximately the rightmost $1/4$ of the Giant1–Giant2 interval, but only $n^2/8 + O(n)$ of the Giant1–Giant2 slopes are in this interval. Overall there are $23n^2/8 + O(n)$ distinct slopes, that is, $(0.71875 + o(1))\ell$ distinct slopes.

For comparison, the same $(3+o(1))n$ multiplications allow the original baby-step-giant-step method to compute $(1.5+o(1))n$ baby steps and $(1.5+o(1))n$ giant steps, producing only $(2.25+o(1))n^2 = (0.5625+o(1))\ell$ distinct slopes. The same number of multiplications in the rho method (with $r \in 1/o(1)$ different steps, simulating a uniform random walk within a factor $1+o(1)$) produces $(9+o(1))n^2/2 = (1.125+o(1))\ell$ random slopes, and thus $(1-\exp(1.125)+o(1))\ell = (0.6753\ldots+o(1))\ell$ distinct slopes with overwhelming probability. We have performed computer experiments to check each of these numbers.

Weighing the giants. We repeat a warning from Section 1: One algorithm can be better than another after a particular number of multiplications but nevertheless have worse average-case performance.

For example, the baby-step-giant-step method has two standard variants, which we call the baby-steps-then-giant-steps method (introduced by Shanks in [22, pages 419–420]) and the interleaved-baby-step-giant-step method (introduced much later by Pollard in [17, p. 439, top]). Both variants (with giant steps chosen to be of size $(1+o(1))\sqrt{\ell}$) reach 100% success probability using $(2+o(1))\sqrt{\ell}$ multiplications, while the rho method has a lower success probability for that number of multiplications. Average-case performance tells a quite different story: The baby-steps-then-giant-steps method uses $(1.5+o(1))\sqrt{\ell}$ multiplications on average; the interleaved-baby-step-giant-step method is better, using $(4/3+o(1))\sqrt{\ell} = (1.3333\ldots+o(1))\sqrt{\ell}$ multiplications on average; the rho method (again with $1/r \in o(1)$) is best, using $(\sqrt{\pi/2}+o(1))\sqrt{\ell} = (1.2533\ldots+o(1))\sqrt{\ell}$ multiplications on average.

Our analysis above shows that the two-grumpy-giants-and-a-baby method is more effective than the rho method (and the baby-step-giant-step method) as a way to use $(1.5+o(1))\sqrt{\ell}$ multiplications. One might nevertheless guess that the rho method has better average-case performance; for example, an anonymous referee stated that the new method "presumably has worse average-case running time".

Our computer experiments indicate that the (interleaved-)two-grumpy-giants-and-a-baby method actually has better average-case running time than the rho method. For example, for $\ell = 65537$, we found a chain of weight $20644183 = (1.23046\ldots)\ell^{1.5}$ with the two-grumpy-giants-and-a-baby method. Here we chose $n = 146$, used (suboptimal) binary addition chains for $(0,n)$ and $(0,\ell-n-1)$, and then cycled between points $(0,i)$ and $(1,in)$ and $(2,-i(n+1))$ until we had ℓ different slopes. For $\ell = 1000003$ we found a chain of weight $1205458963 = (1.20545\ldots)\ell^{1.5}$ in the same way with $n = 558$.

Variants. We have been exploring many variants of this algorithm. We have found experimentally that a 4-giants algorithm (two in one direction, two in the other, with computer-optimized shifts of the initial positions) outperforms this 2-giants

algorithm for $m \approx \sqrt{\ell}$. We speculate that gradually increasing the number of giants will produce an algorithm with $(0.5 + o(1))m^2$ distinct slopes, the best possible result (automatically also optimizing the average number of multiplications, the maximum, and so on), but it is not clear how to choose the shift distances properly.

Acknowledgments

This work was supported by the National Science Foundation under grants 0716498 and 1018836 and by the European Commission under Contract ICT-2007-216676 ECRYPT II. Computations were carried out on the LISA cluster at the SARA supercomputer center, supported by NCF grant MP-230-11. We thank the anonymous referees for several useful comments and questions. No babies (or giants) were harmed in the preparation of this paper.

Appendix A. Computing limits of anticollision factors

This appendix shows, for each integer $r > 3$, a reasonably fast method to compute the limit of the sequence of generic uniform heuristic nonrandomness factors

$$1/\sqrt{1-1/r},$$
$$1/\sqrt{1-1/r-1/r^2+1/r^3},$$
$$1/\sqrt{1-1/r-1/r^2-2/r^3+7/r^4-4/r^5},$$
$$\dots$$

considered in Section 4. For example, these factors converge to approximately 1.129162 for $r = 6$.

We are indebted to Neil Sloane's Online Encyclopedia of Integer Sequences [24] for leading us to [5] (by a search for the integer 4229523740916 shown below), and to Armin Straub for explaining how to use [2] and [18] to compute the sum $\sum_k u_k/r^{2k}$ discussed here. Our contribution here is the connection described below between anticollision factors and sums of squares of multinomials.

Review of sums of squares of multinomials. Define $U = \sum_i \sum_j s_i/s_j$ in the r-variable function field $\mathbb{Q}(s_1, \dots, s_r)$, and define u_k as the constant coefficient of U^k. Consider the problem of computing $\sum_{k\ge 0} u_k/r^{2k}$.

Note that $U^k = \sum_{i_1,\dots,i_k} \sum_{j_1,\dots,j_k} s_{i_1}\cdots s_{i_k}/s_{j_1}\cdots s_{j_k}$, so u_k is the number of tuples $(i_1, \dots, i_k, j_1, \dots, j_k)$ such that $s_{i_1}\cdots s_{i_k}/s_{j_1}\cdots s_{j_k} = 1$; that is, such that $(i_1, \dots, i_k)$ is a permutation of $(j_1, \dots, j_k)$. The tuples counted here were named "abelian squares" by Erdős in 1961, according to [19]; u_k here is "$f_r(k)$" in the notation of [19].

For example, $u_0 = 1$; $u_1 = r$; and $u_2 = 2r^2 - r$, which one can partition into counting $2r^2 - 2r$ tuples (i_1, i_2, j_1, j_2) with $i_1 \neq i_2$ and $\{i_1, i_2\} = \{j_1, j_2\}$, and r tuples with $i_1 = i_2 = j_1 = j_2$. More generally, the number of ways for $s_{i_1} \cdots s_{i_k}$ to equal $s_1^{a_1} \cdots s_r^{a_r}$ is the multinomial coefficient $\binom{k}{a_1, a_2, \ldots, a_r}$, so

$$u_k = \sum_{\substack{a_1, a_2, \ldots, a_r: \\ a_1 + a_2 + \cdots + a_r = k}} \binom{k}{a_1, a_2, \ldots, a_r}^2$$

$$= \sum_{m \geq 0} \binom{r}{m} \sum_{\substack{a_1, a_2, \ldots, a_m: \\ a_1 + a_2 + \cdots + a_m = k, \\ a_1 > 0, a_2 > 0, \ldots, a_m > 0}} \binom{k}{a_1, a_2, \ldots, a_m}^2.$$

Richmond and Rousseau, proving a conjecture of Ruehr, showed in [18] that u_k is asymptotically $r^{2k+r/2}/(4\pi k)^{(r-1)/2}$ as $k \to \infty$. See also [19, Theorem 4] for another proof. We conclude that $\sum_k u_k/r^{2k}$ converges for $r > 3$ (and not for $r = 3$). For example, with $r = 6$, the ratio u_k/r^{2k} is asymptotically $6^3/(4\pi k)^{2.5}$, so $\sum_k u_k/r^{2k}$ converges, and the tail $\sum_{k>n} u_k/r^{2k}$ is $\Theta(1/n^{1.5})$.

This Θ is not an explicit bound; [18] and [19] are not stated constructively. However, inspecting examples strongly suggests that $(u_k/r^{2k})/(r^{r/2}/(4\pi k)^{(r-1)/2})$ converges upwards to 1 as $k \to \infty$, so it seems reasonably safe to hypothesize that u_k/r^{2k} is at most $2r^{r/2}/(4\pi k)^{(r-1)/2}$. This hypothesis implies that

$$\sum_{k>n} \frac{u_k}{r^{2k}} \leq \sum_{k>n} \frac{2r^{r/2}}{(4\pi k)^{(r-1)/2}}$$

$$< \int_n^\infty \frac{2r^{r/2}}{(4\pi k)^{(r-1)/2}}\, dk$$

$$= \frac{4r^{r/2}}{(4\pi)^{(r-1)/2}(r-3)n^{(r-3)/2}},$$

so to compute tight bounds on $\sum_k u_k/r^{2k}$ it suffices to compute $\sum_{0 \leq k \leq n} u_k/r^{2k}$ for a moderately large integer n.

One can easily use the multinomial formula above to compute, for example, that $u_{10} = 4229523740916$ for $r = 6$, but if k and r are not very small then it is much more efficient to compute u_k from the generating function $\sum_k u_k x^k/k!^2 = \left(\sum_k x^k/k!^2\right)^r$ in the power-series ring $\mathbb{Q}[\![x]\!]$. Barrucand in [2] pointed out this formula for u_k and explained how to use it to compute a recurrence for u_k. For $r = 6$ we simply computed the 6th power of $\sum_k x^k/k!^2$ in $\mathbb{Q}[x]/x^{5001}$, obtaining the exact values of u_k for $0 \leq k \leq 5000$ and concluding that $\sum_{0 \leq k \leq 5000} u_k/6^{2k} \approx 1.275007093$. This computation was fast enough that we did not bother to explore

optimizations such as computing $\left(\sum_k x^k/k!^2\right)^r$ modulo various small primes or analyzing the numerical stability of Barrucand's recurrence.

Anticollision factors via sums of squares of multinomials. Define h_k as the number of tuples $(i_1, i_2, \ldots, i_k, j_1, j_2, \ldots, j_k) \in \{1, \ldots, r\}^{2k}$ such that

$$s_{i_1} \neq s_{j_1}, \quad s_{i_1}s_{i_2} \neq s_{j_1}s_{j_2}, \quad \ldots, \quad \text{and } s_{i_1}s_{i_2}\cdots s_{i_k} \neq s_{j_1}s_{j_2}\cdots s_{j_k}$$

in the polynomial ring $\mathbb{Z}[s_1, \ldots, s_r]$. For example, $h_0 = 1$; $h_1 = r^2 - r$; and $h_2 = r^4 - r^3 - r^2 + r$.

The degree-$\leq k$ generic uniform heuristic nonrandomness factor is $1/\sqrt{h_k/r^{2k}}$. The goal of this appendix is to compute $\lim_{k\to\infty} 1/\sqrt{h_k/r^{2k}}$.

Define H_k as the sum of quotients $s_{i_1}\cdots s_{i_k}/s_{j_1}\cdots s_{j_k}$ over the same tuples $(i_1, \ldots, i_k, j_1, \ldots, j_k)$ counted by h_k. For $k \geq 1$ the product

$$H_{k-1}U = H_{k-1}\sum_{i_k}\sum_{j_k}\frac{s_{i_k}}{s_{j_k}}$$

is the sum of quotients $s_{i_1}\cdots s_{i_k}/s_{j_1}\cdots s_{j_k}$ over the tuples $(i_1, \ldots, i_k, j_1, \ldots, j_k)$ with

$$s_{i_1} \neq s_{j_1}, \quad s_{i_1}s_{i_2} \neq s_{j_1}s_{j_2}, \quad \ldots, \quad \text{and } s_{i_1}s_{i_2}\cdots s_{i_{k-1}} \neq s_{j_1}s_{j_2}\cdots s_{j_{k-1}}.$$

These are the same as the tuples contributing to H_k, except for tuples having $s_{i_1}s_{i_2}\cdots s_{i_k} = s_{j_1}s_{j_2}\cdots s_{j_k}$. The product $H_{k-1}U$ is therefore the same as H_k, except for its constant coefficient. The constant coefficient of H_k is 0, so $H_k = H_{k-1}U - c_k$ where c_k is the constant coefficient of $H_{k-1}U$.

By induction $H_k = U^k - c_1U^{k-1} - c_2U^{k-2} - \cdots - c_k$. Recall that the constant coefficient of U^k is u_k, so $0 = u_k - c_1u_{k-1} - c_2u_{k-2} - \cdots - c_k$. In other words, $(1 - c_1x - c_2x^2 - \cdots)(1 + u_1x + u_2x^2 + \cdots) = 1$ in the power-series ring $\mathbb{Z}[\![x]\!]$. For the same reason, the product $(1 - c_1x - \cdots - c_kx^k)(1 + u_1x + \cdots + u_kx^k)$ is $1 - (c_1u_k + \cdots + c_ku_1)x^{k+1} - \cdots - c_ku_kx^{2k}$, so

$$\left(1 - \frac{c_1}{r^2} - \cdots - \frac{c_k}{r^{2k}}\right)\left(1 + \frac{u_1}{r^2} + \cdots + \frac{u_k}{r^{2k}}\right) = 1 - \epsilon_k$$

where $\epsilon_k = (c_1u_k + \cdots + c_ku_1)/r^{2k+2} + \cdots + c_ku_k/r^{4k}$. The bounds

$$0 \leq \epsilon_k \leq \frac{u_{k+1}}{r^{2k+2}} + \frac{u_{k+2}}{r^{2k+4}} + \cdots$$

show that $\epsilon_k \to 0$ as $k \to \infty$, so

$$\left(1 - \frac{c_1}{r^2} - \frac{c_2}{r^4} - \cdots\right)\left(1 + \frac{u_1}{r^2} + \frac{u_2}{r^4} + \cdots\right) = 1.$$

Mapping $s_1 \mapsto 1$, $s_2 \mapsto 1$, ..., $s_r \mapsto 1$ takes H_k to h_k and takes U to r^2, so $h_k = h_{k-1}r^2 - c_k$; that is, $h_k/r^{2k} = h_{k-1}/r^{2k-2} - c_k/r^{2k}$. By induction,

$$\frac{h_k}{r^{2k}} = 1 - \frac{c_1}{r^2} - \frac{c_2}{r^4} - \cdots - \frac{c_k}{r^{2k}}.$$

Hence

$$\lim_{k\to\infty} \frac{h_k}{r^{2k}} = 1 - \frac{c_1}{r^2} - \frac{c_2}{r^4} - \cdots = \frac{1}{1 + u_1/r^2 + u_2/r^4 + \cdots}.$$

The desired value $\lim_{k\to\infty} 1/\sqrt{h_k/r^{2k}}$ is therefore the square root of the sum $\sum_k u_k/r^{2k}$ computed above. In particular, for $r = 6$ we find

$$\lim_{k\to\infty} \frac{1}{\sqrt{h_k/r^{2k}}} \approx 1.129162.$$

References

[1] Daniel V. Bailey, Lejla Batina, Daniel J. Bernstein, Peter Birkner, Joppe W. Bos, Hsieh-Chung Chen, Chen-Mou Cheng, Gauthier van Damme, Giacomo de Meulenaer, Luis Julian Dominguez Perez, Junfeng Fan, Tim Güneysu, Frank Gurkaynak, Thorsten Kleinjung, Tanja Lange, Nele Mentens, Ruben Niederhagen, Christof Paar, Francesco Regazzoni, Peter Schwabe, Leif Uhsadel, Anthony Van Herrewege, and Bo-Yin Yang, *Breaking ECC2K-130*, Cryptology ePrint Archive, Report 2009/541, 2009. http://eprint.iacr.org/2009/541

[2] Pierre Barrucand, *Sur la somme des puissances des coefficients multinomiaux et les puissances successives d'une fonction de Bessel*, C. R. Acad. Sci. Paris **258** (1964), 5318–5320. MR 29 #40

[3] Daniel J. Bernstein, Tanja Lange, and Peter Schwabe, *On the correct use of the negation map in the Pollard rho method*, in Catalano et al. [8], 2011, pp. 128–146, expanded version at http://eprint.iacr.org/2011/003. MR 2012h:94145

[4] Simon R. Blackburn and Sean Murphy, *The number of partitions in Pollard rho*, technical report RHUL-MA-2011-11, Department of Mathematics, Royal Holloway, University of London, 2011. http://www.ma.rhul.ac.uk/static/techrep/2011/RHUL-MA-2011-11.pdf

[5] Jonathan M. Borwein, Dirk Nuyens, Armin Straub, and James Wan, *Some arithmetic properties of short random walk integrals*, Ramanujan J. **26** (2011), no. 1, 109–132. MR 2012j:60114

[6] Joppe W. Bos, Thorsten Kleinjung, and Arjen K. Lenstra, *On the use of the negation map in the Pollard rho method*, in Hanrot et al. [12], 2010, pp. 66–82. MR 2011k:11175

[7] Richard P. Brent and John M. Pollard, *Factorization of the eighth Fermat number*, Math. Comp. **36** (1981), no. 154, 627–630. MR 83h:10014

[8] Dario Catalano, Nelly Fazio, Rosario Gennaro, and Antonio Nicolosi (eds.), *Public key cryptography — PKC* 2011: *Proceedings of the* 14*th International Conference on Practice and Theory in Public Key Cryptography held in Taormina*, *March* 6–9, 2011, Lecture Notes in Computer Science, no. 6571, Springer, Heidelberg, 2011. MR 2012h:94007

[9] M. Chateauneuf, A. C. H. Ling, and D. R. Stinson, *Slope packings and coverings*, *and generic algorithms for the discrete logarithm problem*, J. Combin. Des. **11** (2003), no. 1, 36–50, preprint version at http://eprint.iacr.org/2001/094. MR 2003j:05035

[10] Jung Hee Cheon, Jin Hong, and Minkyu Kim, *Speeding up the Pollard rho method on prime fields*, in Pieprzyk [16], 2008, pp. 471–488. MR 2546112

[11] Walter Fumy (ed.), *Advances in cryptology — EUROCRYPT '*97: *Proceedings of the* 16*th International Conference on the Theory and Application of Cryptographic Techniques held in Konstanz*, *May* 11–15, 1997, Lecture Notes in Computer Science, no. 1233, Springer, Berlin, 1997. MR 98i:94002

[12] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory*: *Proceedings of the* 9*th Biennial International Symposium* (*ANTS-IX*) *held in Nancy*, *July* 19–23, 2010, Lecture Notes in Computer Science, no. 6197, Springer, Berlin, 2010. MR 2011g:11002

[13] Martin Hildebrand, *Random walks supported on random points of* $\mathbb{Z}/n\mathbb{Z}$, Probab. Theory Related Fields **100** (1994), no. 2, 191–203. MR 95j:60015

[14] Donald J. Lewis (ed.), 1969 *Number Theory Institute*: *Proceedings of the* 1969 *Summer Institutes on Number Theory*: *Analytic Number Theory*, *Diophantine Problems*, *and Algebraic Number Theory*; *held at the State University of New York at Stony Brook*, *July* 7–*August* 1, 1969, Proceedings of Symposia in Pure Mathematics, no. 20, American Mathematical Society, Providence, R.I., 1971. MR 47 #3286

[15] V. I. Nechaev, *Complexity of a determinate algorithm for the discrete logarithm*, Math. Notes **55** (1994), no. 2, 165–172. MR 96a:11145

[16] Josef Pieprzyk (ed.), *Advances in cryptology — ASIACRYPT* 2008: *Proceedings of the* 14*th International Conference on the Theory and Application of Cryptology and Information Security held in Melbourne*, *December* 7–11, 2008, Lecture Notes in Computer Science, no. 5350, Springer, Berlin, 2008. MR 2010j:94005

[17] J. M. Pollard, *Kangaroos*, *Monopoly and discrete logarithms*, J. Cryptology **13** (2000), no. 4, 437–447. MR 2001i:94059

[18] Bruce Richmond and Cecil Rousseau, *A multinomial summation* (*Donald Richards and Stamatis Cambanis*), SIAM Rev. **31** (1989), no. 1, 122–125, comment on Problem 87-2, *SIAM Rev.* **30** (1988), pp. 128–130.

[19] L. B. Richmond and Jeffrey Shallit, *Counting abelian squares*, Electron. J. Combin. **16** (2009), no. 1, Research Paper 72, 9 pages. MR 2010j:05036

[20] J. Sattler and C.-P. Schnorr, *Generating random walks in groups*, Ann. Univ. Sci. Budapest. Sect. Comput. **6** (1985), 65–79. MR 89a:68108

[21] C.-P. Schnorr and H. W. Lenstra, Jr., *A Monte Carlo factoring algorithm with linear storage*, Math. Comp. **43** (1984), no. 167, 289–311. MR 85d:11106

[22] Daniel Shanks, *Class number*, *a theory of factorization*, *and genera*, in Lewis [14], 1971, pp. 415–440. MR 47 #4932

[23] Victor Shoup, *Lower bounds for discrete logarithms and related problems*, in Fumy [11], 1997, pp. 256–266. MR 98j:94023

[24] Neil J. A. Sloane, *The on-line encyclopedia of integer sequences*, 2012. http://oeis.org

[25] Edlyn Teske, *On random walks for Pollard's rho method*, Math. Comp. **70** (2001), no. 234, 809–825. MR 2001g:11194

DANIEL J. BERNSTEIN: djb@cr.yp.to
Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607-7053, United States

TANJA LANGE: tanja@hyperelliptic.org
Department of Mathematics and Computer Science, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.113

msp

Improved techniques for computing the ideal class group and a system of fundamental units in number fields

Jean-François Biasse and Claus Fieker

We describe improvements to the subexponential methods for computing the ideal class group, the regulator and a system of fundamental units in number fields under the generalized Riemann hypothesis. We use sieving techniques adapted from the number field sieve algorithm to derive relations between elements of the ideal class group, and p-adic approximations to manage the loss of precision during the computation of units. These improvements are particularly efficient for number fields of small degree for which a speedup of an order of magnitude is achieved with respect to the standard methods.

1. Introduction

Let $K = \mathbb{Q}(\theta)$ be a number field of degree n and discriminant Δ. In this paper, we present improved fast methods for computing the structure of the ideal class group of the maximal order $\mathbb{O}_K$ of K, along with the regulator and a system of fundamental units of $\mathbb{O}_K$.

Class group and unit group computation are two of the four principal tasks for computational algebraic number theory postulated by Zassenhaus (together with the computation of the ring of integers and the Galois group). In particular, they occur in the resolution of Diophantine equations. For example, the Pell equation

$$T^2 - \Delta U^2 = 1, \quad T, U \in \mathbb{Z},$$

boils down to finding the fundamental unit in a real quadratic number field of discriminant Δ (see [26]). In addition, the Schäffer equation

$$y^2 = 1^k + 2^k + \cdots + (x-1)^k, \quad k \geq 2,$$

MSC2000: primary 54C40, 14E20; secondary 46E25, 20C20.
Keywords: number fields, ideal class group, regulator, units, index calculus, subexponentiality.

can be solved using solutions to the Pell equation [24]. Unit computations are key ingredients in solving almost all Diophantine equations, for example when solving Thue equations [8]. On the other hand, the computation of the ideal class group $\mathrm{Cl}(\mathbb{O}_K)$ of a number field K allows in particular to provide numerical evidence in favor of unproven conjectures such as the heuristics of Cohen and Lenstra [14] on the ideal class group of a quadratic number field, Littlewood's bounds [32] on $L(1, \chi)$, or Bach's bound on the minimal bound B such that ideals of norm lower than B generate the ideal class group. The class group enters also into the computation of the Mordell-Weil group of elliptic curves with the descent method, or the Brauer group computations for representation theory [16].

In 1968, Shanks [41; 42] proposed an algorithm relying on the baby-step giant-step method to compute the structure of the class number and the regulator of a quadratic number field in time $O(|\Delta|^{1/4+\epsilon})$, or $O(|\Delta|^{1/5+\epsilon})$ under the extended Riemann hypothesis [30]. In 1985 Pohst and Zassenhaus [37] published an algorithm that could determine the class group of arbitrary number fields. Then, a subexponential strategy for the computation of the group structure of the class group of an imaginary quadratic field was described in 1989 by Hafner and McCurley [21]. The expected running time of this method is bounded by $L_\Delta(1/2, \sqrt{2} + o(1))$ where

$$L_\Delta(\alpha, \beta) := e^{\beta(\log|\Delta|)^\alpha(\log\log|\Delta|)^{1-\alpha}}.$$

Buchmann [11] generalized this result to the case of an arbitrary extension, the heuristic complexity being valid for fixed degree n and Δ tending to infinity. In a recent work [6], Biasse described an algorithm achieving the heuristic complexity $L_\Delta(1/3, O(1))$ for certain classes of number fields where both the discriminant and the degree tend to infinity.

In parallel with theoretical improvements, considerable efforts have been invested to make the implementations of the subexponential methods efficient. In the quadratic case, Jacobson [25] described an algorithm based on the quadratic sieve for deriving relations between elements of $\mathrm{Cl}(\mathbb{O}_K)$. He successfully used it for computing the class group and the fundamental unit of quadratic number fields. His implementation contained some of the practical improvements described in the context of factorization such as self-initialization and the single large prime variant. This strategy was later improved by Biasse [7] who used a double large prime variant and a dedicated Gaussian elimination technique. Attempts have been made to generalize sieving techniques to general number fields [12; 34]. A variant of the number field sieve was used for deriving relations in the class group of cubic fields. On special classes of cubic number fields for which the regulator can be precomputed, it allowed the computation of the ideal class group. Promising timings were presented in [12; 34], for sizes of factor base that do not (to the best

of our knowledge) certify the result under the generalized Riemann hypothesis. In particular, a significant speedup was obtained over the standard random ideal factorization method.

Our contribution. In this paper, we present an algorithm based on sieving techniques adapted from recent implementations of the number field sieve [28] for computing $\mathrm{Cl}(\mathbb{O}_K)$ under the generalized Riemann hypothesis (GRH) for an arbitrary number field K. We also describe a p-adic method for computing the regulator and a system of fundamental units. We show that these methods allow a significant improvement for number fields of low degree over the current state of the art based on enumeration techniques.

2. Generalities on number fields

Let K be a number field of degree d. It has $r_1 \leq d$ real embeddings $(\sigma_i)_{i\leq r_1}$ and $2r_2$ complex embeddings $(\sigma_i)_{r_1<i\leq d}$ coming as r_2 pairs of conjugates, which we number so that $\sigma_{i+r_2} = \overline{\sigma_i}$ for $r_1 < i \leq r_1 + r_2$. The field K is isomorphic to $\mathbb{O}_K \otimes \mathbb{Q}$ where $\mathbb{O}_K$ denotes the ring of integers of K. We can embed K in $K_\mathbb{R} := K \otimes \mathbb{R} \simeq \mathbb{R}^{r_1} \times \mathbb{C}^{r_2}$, and extend the σ_i to $K_\mathbb{R}$. Let T_2 be the Hermitian form on $K_\mathbb{R}$ defined by $T_2(x, x') := \sum_i \sigma_i(x)\overline{\sigma_i}(x')$, and let $\|x\| := \sqrt{T_2(x, x)}$ be the corresponding L_2-norm. Choose $(\alpha_i)_{i\leq d}$ such that $\mathbb{O}_K = \bigoplus_i \mathbb{Z}\alpha_i$; then the discriminant of K is given by $\Delta = \det^2(T_2(\alpha_i, \alpha_j))$. The norm of an element $x \in K$ is defined as $\mathcal{N}(x) = \prod_i \sigma_i(x)$.

Let $\mathcal{I}$ be the group of nonzero fractional ideals of K and $\mathcal{P} \subseteq \mathcal{I}$ is the subgroup of principal fractional ideals. The norm of integral ideals is given by $\mathcal{N}(I) := [\mathbb{O}_K : I]$, which extends to fractional ideals by $\mathcal{N}(I/J) := \mathcal{N}(I)/\mathcal{N}(J)$. The norm of a principal ideal agrees with the norm of its generator: $\mathcal{N}(x\mathbb{O}_K) = |\mathcal{N}(x)|$.

The ideal class group of $\mathbb{O}_K$ is defined by $\mathrm{Cl}(\mathbb{O}_K) := \mathcal{I}/\mathcal{P}$. We denote by $[\mathfrak{a}]$ the class of a fractional ideal $\mathfrak{a}$ in $\mathrm{Cl}(\mathbb{O}_K)$ and by h the cardinality of $\mathrm{Cl}(\mathbb{O}_K)$. Elements of $\mathcal{I}$ admit a unique decomposition as a power product of prime ideals of $\mathbb{O}_K$ (with possibly negative exponents). An element $x \in \mathbb{O}_K$ is said to be a unit if $(x)\mathbb{O}_K = \mathbb{O}_K$, or equivalently if $|\mathcal{N}(x)| = 1$. The units of $\mathbb{O}_K$ form a multiplicative group of the form

$$U = \mu \times \langle\gamma_1\rangle \times \cdots \times \langle\gamma_r\rangle,$$

where μ is the torsion subgroup of U, $r := r_1 + r_2 - 1$ and the generators γ_i of the nontorsion part are called a system of fundamental units. The regulator is an invariant of K which allows us to certify the calculation of $\mathrm{Cl}(\mathbb{O}_K)$ and U. It is defined as $R = \mathrm{Vol}(\Gamma)$ where Γ is the lattice generated by vectors of the form

$$(c_1 \log|\gamma_i|_1, \ldots, c_{r+1} \log|\gamma_i|_{r+1}),$$

with $|x|_i := |\sigma_i(x)|$ for $i \leq r + 1$, $c_1 = 1$ for $i \leq r_1$, $c_i = 2$ otherwise.

3. The subexponential strategy

The idea behind the algorithm of Buchmann [11] is to find a set of ideals $\mathcal{B} = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_N\}$ whose classes generate $\mathrm{Cl}(\mathbb{O}_K)$, and then consider the surjective morphism

$$\begin{array}{ccccc} \mathbb{Z}^n & \xrightarrow{\quad\varphi\quad} & I & \xrightarrow{\quad\pi\quad} & \mathrm{Cl}(\mathbb{O}_K) \\ (e_1, \ldots, e_N) & \longmapsto & \prod_i \mathfrak{p}_i^{e_i} & \longmapsto & \prod_i [\mathfrak{p}_i]^{e_i}. \end{array}$$

From the fundamental theorem of homomorphisms, the ideal class group satisfies $\mathrm{Cl}(\mathbb{O}_K) \simeq \mathbb{Z}^N / \ker(\pi \circ \varphi)$. Therefore, the knowledge of $\ker(\pi \circ \varphi)$, which has the structure of a $\mathbb{Z}$-lattice, enables us to derive $\mathrm{Cl}(\mathbb{O}_K)$. In the meantime, elements of $\ker(\varphi)$ give us units as power-products of relations. From these units, we hope to derive a system of fundamental units of $\mathbb{O}_K$. The subexponential strategy can be broken down into three essential tasks: collecting relations, calculating the class group and calculating the unit group. The subexponentiality is a consequence of a careful choice of B.

3.1. ***Relation collection.*** A preliminary step to the relation collection is the choice of a generating set $\mathcal{B} = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_N\}$ of $\mathrm{Cl}(\mathbb{O}_K)$. We choose the set of prime ideals of norm bounded by an integer B. The use of the Minkowski bound certifies the result unconditionally, but it causes the algorithm to take a time exponential in the size of Δ. To achieve subexponentiality, many authors chose the bound of Bach [2], who proved that under GRH, $\mathrm{Cl}(\mathbb{O}_K)$ was generated by the classes of the prime ideals $\mathfrak{p}$ satisfying $\mathcal{N}(\mathfrak{p}) \leq 12(\log|\Delta|)^2$. Although asymptotically better, in practice this bound can be larger than the one described by Belabas et al. [4] who stated that under GRH, the class group is generated by the classes of the prime ideals of norm bounded by B provided that

$$\sum_{(m,\mathfrak{p}):\mathcal{N}(\mathfrak{p}^m)\leq B} \frac{\log \mathcal{N}(\mathfrak{p})}{\mathcal{N}(\mathfrak{p}^{m/2})}\left(1 - \frac{\log \mathcal{N}(\mathfrak{p}^m)}{\log B}\right) > \frac{1}{2}\log|\Delta| - 1.9n - 0.785r_1 + \frac{2.468n + 1.832r_1}{\log B}.$$

In the rest of the paper, we assume that $\mathcal{B}$ is constructed with the bound of Belabas et al. Indeed, Bach's bound enlarges the dimensions of the matrices that are processed during the computation of $\mathrm{Cl}(\mathbb{O}_K)$, thus inducing a slow-down that is not compensated by the fact that the relations are found more rapidly.

During the relation collection phase, we collect relations of the form

$$(\phi_i) = \mathfrak{p}_1^{e_{i,1}} \cdots \mathfrak{p}_N^{e_{i,N}},$$

where $\phi_i \in K$. We progressively build the matrix $M := (e_{i,j}) \in \mathbb{Z}^{k \times N}$ where k is the number of relations collected so far. Let $\Lambda \subseteq \ker(\pi \circ \varphi)$ be the lattice generated

by the rows of M. Operations on the rows of M allow us to retrieve a basis for Λ and its determinant. To determine if Λ has rank N, we perform operations modulo a random wordsize prime p. In particular, the LU decomposition of M modulo p allows us to identify the prime ideals that do not contribute to the rank of Λ. Additional relations involving these primes increase the rank of M, whose rows eventually generate a finite index sublattice of $\ker(\pi)$. To find this index, we compute the Hermite normal form (HNF) of M, that is, we perform unimodular operations encoded by $U \in \mathrm{GL}_k(\mathbb{Z})$ such that

$$UM = \begin{pmatrix} h_{11} & 0 & \cdots & 0 \\ \vdots & h_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ * & * & \cdots & h_{NN} \\ \hline & & (0) & \end{pmatrix},$$

with $0 \le h_{ij} < h_{jj}$ whenever $j < i$ and $h_{ij} = 0$ whenever $j > i$. Once the HNF of M is computed, adding new rows can be done very efficiently. In the meantime, the product $\prod_i h_{i,i}$ gives us an indication on $[\Lambda : \ker(\pi \circ \varphi)]$, as we see in Section 3.3.

3.2. ***Class group computation.*** Given a matrix $A \in \mathbb{Z}^{N \times N}$ whose rows generate $\ker(\pi \circ \varphi)$, unimodular transformations on both rows and columns of A yield the structure of $\mathrm{Cl}(\mathbb{O}_K)$. More precisely, for every nonsingular matrix $A \in \mathbb{Z}^{N \times N}$, there exist unimodular matrices $U, V \in \mathbb{Z}^{N \times N}$ such that

$$S := UAV = \mathrm{diag}(d_1, \ldots, d_N),$$

where $d_{i+1} \mid d_i$ for all i with $1 \le i < N$. The matrix S is called the Smith normal form (SNF) of A.

Theorem 1. *If the rows of $A \in \mathbb{Z}^{N \times N}$ are a basis for* $\ker(\pi \circ \varphi)$ *and* $\mathrm{diag}(d_1, \ldots, d_N)$ *is the SNF of A, then*

$$\mathrm{Cl}(\mathbb{O}_K) \simeq \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_N\mathbb{Z}.$$

Once enough relations have been found, the rows of M generate $\ker(\pi \circ \varphi)$, and the N nonzero rows of the HNF of M are a matrix $A \in \mathbb{Z}^{N \times N}$ whose rows are a basis for $\ker(\pi \circ \varphi)$, and the SNF of A gives us $\mathrm{Cl}(\mathbb{O}_K)$. However, finding the structure of $\mathrm{Cl}(\mathbb{O}_K)$ can also be done by computing the SNF of a matrix which is in practice significantly smaller than A, namely the essential part of A. Indeed, for each matrix H in HNF, there exists an index l such that $h_{i,i} = 1$ for all $i > l$. The upper left $l \times l$ submatrix of H is called its essential part. As the classes of $\mathfrak{p}_i$ for $i > l$ are generated by those of the $\mathfrak{p}_j$, $j \le l$, the SNF of the essential part of A suffices to recover $\mathrm{Cl}(\mathbb{O}_K)$.

3.3. ***Regulator and fundamental units computation.*** Computing the regulator and a system of fundamental units of K consists of finding kernel vectors of M. Indeed, if $X = (x_1, \ldots, x_k)$ satisfies $XM = 0$, then we have

$$\left(\prod_i \phi_i^{x_i}\right)\mathbb{O}_K = \mathbb{O}_K.$$

In other words, $\gamma := \prod_i \phi_i^{x_i}$ is a unit. Every kernel vector X of M yields a unit, and we want to compute the group generated by all those elements as well as the regulator of this group, defined to be zero if the group is not of full rank. So far, finding of relations between units is mostly done using real linear algebra (LLL), the core problem here being the numerical instability of the matrices. This in itself is a consequence of the well-known fact that units are very large in general: Writing the fundamental unit of a real quadratic fields explicitly with the canonical basis needs exponentially many digits while it is always possible to find a product representation of size polynomial in $\log|\Delta|$ (see [13; 43]). At the end of the procedure, we verify that the assumption we made on the completeness of the lattice of relations is true. To this end, we use an approximation of the Euler product

$$hR = \frac{|\mu|\sqrt{|\Delta|}}{2^{r_1}(2\pi)^{r_2}} \lim_{s\to 1}((s-1)\zeta_K(s)),$$

where $\zeta_K(s) = \sum_{\mathfrak{a}} 1/\mathcal{N}(\mathfrak{a})^s$ is the usual ζ-function associated to K and $|\mu|$ is the cardinality of μ. Indeed, it allows us to derive a bound h^* in polynomial time under ERH that satisfies $h^* \leq hR < 2h^*$; see [3]. If the values $\det(\Gamma)$ and $\det(\Lambda)$ do not satisfy this inequality, then we need to collect more relations.

4. Sieving techniques

In this section, we describe sieving techniques to derive relations in $\mathrm{Cl}(\mathbb{O}_K)$ for general number fields. This is a generalization of Jacobson's results [25] for quadratic number fields. Similar ideas were suggested in [12; 34] but the corresponding algorithms were either not implemented or are no longer available for comparison. Here we provide numerical data illustrating the considerable impact of these techniques for class group and unit group computation in the case of low degree number fields.

Given a generating set $\mathcal{B} = \{\mathfrak{p}_1, \ldots \mathfrak{p}_N\}$ for $\mathrm{Cl}(\mathbb{O}_K)$, the usual method for deriving relations consists of computing random exponents $\vec{e} := (e_1, \ldots, e_N)$, $\alpha \in \mathbb{O}_K$ and a reduced ideal $I_{\vec{e}}$ such that

$$\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_N^{e_N} = (\alpha) I_{\vec{e}}.$$

Then, every time $I_{\vec{e}}$ is $\mathcal{B}$-smooth (that is, is a power product of elements of $\mathcal{B}$), we obtain a relation. As the arithmetic of ideals is rather expensive when $n > 2$, the relation search in the computer algebra software PARI [35] and versions $2.x$ for $x < 18$ of Magma [9] consists of enumerating short elements of $I_{\vec{e}}$ via the Fincke-Pohst method [18].

Our method consists of deriving relations from smooth values of polynomials, thus avoiding the cost of the ideal arithmetic and of the ideal reduction. Our method for finding smooth values is based on the recent development of the number field sieve algorithm [28]. The use of trivial methods such as trial division for finding smooth values of our polynomials would yield the same theoretical complexity, but would be impractical for large discriminants. The most efficient implementation of the enumeration-based strategy for finding relations is the one of PARI. Therefore, in the following, we assess the impact of our sieving method by comparing its performance with those of PARI.

4.1. ***Polynomial selection.*** Let $\mathfrak{a}$ be a $\mathcal{B}$-smooth ideal of $\mathbb{O}_K$. In this section, we show how to provide polynomials $P \in \mathbb{Z}[X, Y]$ of degree n derived from $\mathfrak{a}$ such that every $(x, y) \in \mathbb{Z}^2$ such that $P(x, y)$ is B-smooth yields a relation. Note that in theory, $\mathfrak{a}$ can be any ideal, however, we obtained the best results by choosing $\mathfrak{a} = \mathbb{O}_K$. Let α and β be two linearly independent elements of $\mathfrak{a}$. Then, we create by interpolation a $P_{\alpha,\beta} \in \mathbb{Z}[X, Y]$ such that

$$P_{\alpha,\beta}(x, y) = \mathcal{N}(x\alpha + y\beta) \quad \text{for all } x, y \in \mathbb{Z}^2.$$

Every time $\phi_{x,y} := x\alpha + y\beta$ has a smooth norm, we add the relation corresponding to the principal ideal $(\phi_{x,y})$ to the relation matrix. Before applying sieving algorithms to $P_{\alpha,\beta}$ to derive relations, we need to ensure that it is likely to yield enough smooth values. Polynomial selection is an important part of the number field sieve algorithm, and so it is in our algorithm. However, the specificities of our context prevent us from directly adapting the methods of NFS for selecting the sieving polynomial. First of all, we can afford to find relations with many different choices of α and β, whereas the choice of a sieving polynomial in the NFS algorithm is fixed. We require that our choices of α and β yield polynomials with small coefficients, and that we have a sufficient randomization at the infinite places to avoid drawing $\phi_{x,y}$ spanning the same subgroup of the unit group of $\mathbb{O}_K$.

To randomize the choice of α, β, we consider random coefficients $a_1, \ldots, a_n \in \mathbb{R}^n$ such that $\sum_{i \le n} a_i = 0$. For every such n-tuple $\vec{a}$, we define the embedding

$$\psi_{\vec{a}} : \mathfrak{a} \to \mathbb{R}^n, \quad \alpha \mapsto (a_1 \log|\alpha|_1, \ldots, a_n \log|\alpha|_n).$$

For every choice of $\vec{a}$, the set of elements of the form $\psi_{\vec{a}}(\alpha)$ for $\alpha \in \mathfrak{a}$ is a lattice $\Lambda_{\vec{a}}$ of $\mathbb{R}^n$ for which we can find an LLL reduced basis for the norm

$$T_2^{\vec{a}} : (x_1, \ldots, x_n) \mapsto e^{2a_1} x_1^2 + \cdots + e^{2a_n} x_n^2.$$

For every choice of $\vec{a}$, the first two vectors α, β of an LLL reduced basis of $\Lambda_{\vec{a}}$ are potential candidates for the creation of a polynomial yielding smooth values. Every time we draw such a pair of elements of $\mathfrak{a}$, we need to make sure that they do not generate the same $\mathbb{Z}$-module as another pair previously used. To prevent this from happening, every time we draw a pair α, β by the previous method, we express them in terms of the canonical $\mathbb{Z}$-basis of $\mathbb{O}_K$. Thus, to every pair α, β corresponds the matrix $M_{\alpha,\beta} \in \mathbb{Z}^{2\times n}$ of their coordinates. The HNF of $M_{\alpha,\beta}$ uniquely represents the $\mathbb{Z}$-module spanned by (α, β). Thus, to avoid duplicates, we store a hash of the HNF of $M_{\alpha,\beta}$ in a hash table every time we use a pair (α, β) to draw relations. We summarize the procedure of the selection of a sieving polynomial in Algorithm 1.

Algorithm 1 (Polynomial selection).

Input: $\mathfrak{a}$, $(A_1, \ldots, A_n)$, HashTable.
Output: Sieving polynomial $P_{\alpha,\beta}$ corresponding to $\alpha, \beta \in \mathfrak{a}$.

1: **while** a new α, β has not been found **do**
2: Draw $|a_1| \leq A_1, \ldots, |a_n| \leq A_n$ at random such that $a_1 + \cdots + a_n = 0$.
3: Let α and β be the first two elements of a LLL-reduced basis of $\Lambda_{\vec{a}}$ for $\vec{a} = (a_1, \ldots, a_n)$.
4: Compute the hash $h_{\alpha,\beta}$ of the HNF of $M_{\alpha,\beta}$.
5: **if** $h_{\alpha,\beta} \notin$ HashTable **then**
6: Compute by interpolation $P_{\alpha,\beta} \in \mathbb{Z}[X, Y]$ with $P_{\alpha,\beta}(x, y) = \mathcal{N}(x\alpha + y\beta)$.
7: **end if**
8: **end while**
9: **return** $\alpha, \beta, P_{\alpha,\beta}$.

4.2. ***Line sieving.*** The quadratic sieve algorithm [39] used to derive smooth values of a binary quadratic form generalizes to the case of polynomials of arbitrary degree. Its design follows from the observation that if $P \in \mathbb{Z}[X, Y]$ is a polynomial of degree n, then

$$p \mid P(r_p, y_0) \quad \text{for all } y_0 \in \mathbb{Z} \quad \Longrightarrow \quad p \mid P(r_p + ip, y_0) \quad \text{for all } i \in \mathbb{Z}. \quad (1)$$

Given $y_0 \in \mathbb{Z}$, we wish to find the $x \in [-I/2, I/2]$ such that $P(x, y_0)$ is B-smooth, where B is the bound on the norm of the prime ideals in the factor base. Instead of trying them all, we prefer to isolate a short list of good candidates that we test by trial division. If $p \mid P(x, y_0)$ for many $p \leq B$, then $P(x, y_0)$ is likely to be B-smooth. From (1), we know that once we have one root r_p of $P(X, y_0)$ mod p, then we can derive all the others by translation by $(p, 0)$. Line sieving consists of initializing to zero an array S of length I whose cells represent the $x \in$

$[-I/2, I/2]$. Then, for each $p \leq B$, we compute the smallest roots $x_p \in [-I/2, I/2]$ of $P(X, y_0) \bmod p$ and repeat

$$S[x_p] \leftarrow S[x_p] + \log p, \quad x_p \leftarrow x_p + p.$$

Then, whenever $S[x] \approx \log P(x, y_0)$ for $x \in [-I/2, I/2]$, the value $P(x, y_0)$ is likely to be B-smooth. We summarize this procedure in Algorithm 2.

Algorithm 2 (Line sieving).

Input: $P \in Z[X, Y]$, $I, B, y_0 \in \mathbb{Z}$.
Output: Smooth values of $P(X, y_0)$ in $[-I/2, I/2]$.
1: $L \leftarrow \emptyset$; $S[x] \leftarrow 0$ for all $x \in [-I/2, I/2]$.
2: **for** $p \leq B$ **do**
3: Let x_p be the smallest root of $P(X, y_0) \bmod p$ in $[-I/2, I/2]$.
4: **while** $r_p \leq I/2$ **do**
5: $S[x_p] \leftarrow S[x_p] + \log p$, $x_p \leftarrow x_p + p$.
6: **end while**
7: **end for**
8: **for** $x \in [-I/2, I/2]$ **do**
9: **if** $S[x] \approx \log P(x, y_0)$ **then**
10: If $P(x, y_0)$ is B-smooth, $L \leftarrow L \cup \{x\}$.
11: **end if**
12: **end for**
13: **return** L.

4.3. ***Lattice sieving.*** Let $P_{\alpha,\beta}(X, Y) \in \mathbb{Z}[X, Y]$ be the sieving polynomial described in Section 4.1, B the bound on the norm of the ideals in the factor base, and $I, J \in \mathbb{Z}_{>0}$. Every pair $(x, y) \in [-I/2, I/2[\times [1, J]$ such that $P_{\alpha,\beta}(x, y)$ is B-smooth yields a relations. Therefore, one can repeat the line sieving operation on $P_{\alpha,\beta}(X, y_0)$ for every $y_0 \in [1, J]$. This method is efficient when sieving with primes $p < I$. but when the primes are significantly larger than I, the root computation at Step 3 of Algorithm 2 is often performed for nothing since there is a good chance that none of the $x \in [-I/2, I/2[$ will be a root of $P_{\alpha,\beta}(X, y_0) \bmod p$. A way around that is to have an array S of length IJ representing $[-I/2, I/2[$ and to fill it by line sieving methods for the primes $p < I$ and by lattice sieving for the other primes.

The lattice sieve was first described by Pollard [38]. Since then, it has been extensively studied and improved in the past 15 years, and the most recent developments of this methods yielded the factorization of RSA768 (see [28]). This strategy relies on a one-time enumeration of roots of $P_{\alpha,\beta}(X, Y) \bmod p$ in $[-I/2, I/2[\times [1, J]$. The entry $x \leq IJ$ of the array S that we use to store the logarithmic contributions

corresponds to the pair $(i, j) \in [-I/2, I/2[\times [1, J]$ where

$$i = (x - I/2) \mod I, \quad j = (x - i - I/2)/I.$$

As in the line sieving case, every entry of S is initialized to zero, and for every $p \leq B$ and every $(i, j) \in [-I/2, I/2[\times [1, J]$ such that $p \mid P_{\alpha,\beta}(i, j)$, we want to perform the operation $S[x] \leftarrow S[x] + \log p$. Line sieving repeated on every line $j \leq J$ allows us to efficiently do this for $p < I$. For the others, we followed the approach of [19], as it is done in [28] for the factorization of RSA768. By [19, Proposition 1], we know that for every p such that we have a root r_p of $P_{\alpha,\beta}(X, 1)$ modulo p, there exists a basis $\{(a, b), (c, d)\}$ of the lattice spanned by $\{(r_p, 1), (p, 0)\}$ that satisfies

- $b > 0$ and $d > 0$;
- $-I < a \leq 0 \leq c < I$;
- $c - a \geq I$.

This basis is computed via an algorithm described in [19] that relies on the continued fraction expansion of r_p. It satisfies $p \mid P_{\alpha,\beta}(ia + jc, ib + jd)$ for all $(i, j) \in \mathbb{Z}^2$. To fill the array S, we start from $(i, j) = (0, 0)$ which is a common root modulo all primes. Then, by induction, we construct the next pair (i', j') from (i, j) by choosing

- $(i, j) + (a, b)$ if $i \geq -a$;
- $(i, j) + (c, d)$ if $i < I - c$;
- $(i, j) + (a, b) + (c, d)$ if $I - c \leq i < -a$.

4.4. ***Special-q.*** The sieving space $[-I/2, I/2[\times [1, J]$ only contains a limited number of pairs (i, j) yielding a smooth value. Enlarging I and J might cause its size to rapidly exceed single precision. For a fixed prime q, the special-q strategy consists of sieving with a polynomial P_q derived from the original sieving polynomial P such that

$$\forall (i, j) \in [-I/2, I/2[\times [1, J], \quad \exists (x, y) \in \mathbb{Z}^2, \quad P_q(i, j) = P(x, y),$$
$$\forall (i, j) \in [-I/2, I/2[\times [1, J], \quad q \mid P_q(i, j).$$

This strategy was used by Pollard in his original paper [38] to sieve on the rational side, but most current implementations use it on the algebraic side as well [28]. To create P_q for a given q, we need a root r_q of P modulo q. Then, we find a reduced basis $(a_0, b_0), (a_1, b_1)$ of the lattice spanned by the vectors $(q, 0), (r_q, 1)$. The polynomial P_q is then simply given by

$$P_q(i, j) = P(ia_0 + ja_1, ib_0 + jb_1).$$

The reduced basis is given by successive Gaussian reductions, as explained in [19]. Then, to sieve with a given polynomial P, we repeat the procedure described in Section 4.3 for many different polynomials of the form P_q. Fortunately, once the roots of $P \mod p$ for all $p \leq B$ have been computed, it is possible to use these values to compute the roots of $P_q \mod p$ for $p \leq B$. Indeed,

$$P(ia_0 + ja_1, ib_0 + jb_1) \equiv 0 \mod p$$

means that there is some root r_p of $P(X,1) \mod p$ such that $r_p \equiv \frac{ia_0+ja_1}{ib_0+jb_1} \mod p$. This implies that we have $P_q(r_p^q, 1) \equiv 0 \mod p$ for

$$r_p^q \equiv \frac{i}{j} \equiv -\frac{a_1 - r_p b_1}{a_0 - r_p b_0} \mod p,$$

which gives us a root of $P_q(X, 1) \mod p$ from $(a_0, b_0), (a_1, b_1)$ and a root of $P(X, 1) \mod p$. We summarize our procedure to derive relations from an ideal $\mathfrak{a} \subseteq \mathbb{O}_K$ in Algorithm 3.

Algorithm 3 (Sieving procedure).

Input: $\mathfrak{a} \subseteq \mathbb{O}_K$, $\mathcal{B} = \{\mathfrak{p} \mid \mathcal{N}(\mathfrak{p}) \leq B\}$, $I, J \in \mathbb{Z}_{>0}$.

Select $\alpha, \beta \in \mathbb{O}_K$ and a sieving polynomial $P_{\alpha,\beta}$ with Algorithm 1.
For all $p \leq B$, compute the roots of $P_{\alpha,\beta}(X, 1) \mod p$.
for $q \leq B$ **do**
  Compute P_q and its roots modulo the $p \leq B$ as in Section 4.4.
  Let S be an array of size IJ initialized to 0.
  for $p \leq I$ **do**
    Do $S[x] \leftarrow S[x] + \log p$ for each x representing $(i, j) \in [-I/2, I/2[\times [1, J]$ such that $p \mid P_q(i, j)$ by repeating Algorithm 2 for each line $j \leq J$.
  end for
  for $p > I$ **do**
    Calculate a basis $\{(a, b), (c, d)\}$ of the lattice of points in $[-I/2, I/2[\times [1, J]$ that are roots of $P_q(X, Y) \mod p$ with the method of Section 4.3.
    Do $S[x] \leftarrow S[x] + \log p$ for each x representing $(i, j) \in [-I/2, I/2[\times [1, J]$ such that $p \mid P_q(i, j)$ by using the method of Section 4.3.
  end for
end for
for $x \leq IJ$ **do**
  if $S[x] \approx \log P_q(i, j)$, where x represent $(i, j) \in [-I/2, I/2[\times [1, J]$ **then**
    If $\log P_q(i, j)$ is B-smooth, store the corresponding relation.
  end if
end for

4.5. ***Overall relation collection phase.*** A necessary condition to compute the class group and the unit group is to produce a full-rank relation matrix M. Our sieving methods allow us to derive relations in $\mathrm{Cl}(\mathbb{O}_K)$ very rapidly, but it is hard to force a given prime to occur in a relation. The best performance is obtained by sieving with the trivial ideal $\mathbb{O}_K$. If we want to see a given prime ideal $\mathfrak{p} \mid (p)$ occur in a relation, one can use the special-q with $q = p$, or sieve with the ideal $\mathfrak{p}$. However, even after using those methods, some prime ideals still do not contribute to the rank of M. Rather than sieving in random power-products involving missing primes, one might prefer to switch to enumeration-based methods to complete the relation search. To identify the primes that need to appear in a relation, we perform an LU decomposition of the relation matrix modulo a random wordsize prime. We try to produce enough relations with sieving so that the rank of M is 97% of $\#\mathcal{B}$. Then we find additional relations with enumeration. We summarize this procedure:

Algorithm 4 (Full rank relation matrix computation).

Input: K, B.
Output: A full-rank relation matrix for the primes of norm bounded by B.

1: $\mathcal{B} \leftarrow \{\mathfrak{p} \mid \mathcal{N}(\mathfrak{p}) \leq B\} = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_N\}$.
2: Derive N relations by repeating Algorithm 3 with $\mathfrak{a} = (1)$. Let M be the relation matrix.
3: Perform an LU decomposition of M and let EmptyList be the list of zero columns.
4: **for** $\mathfrak{p} \in$ EmptyList **do**
5: Sieve with $\mathfrak{p}$, update M.
6: **end for**
7: Update EmptyList by updating the LU decomposition of M.
8: **for** $\mathfrak{p} \in$ EmptyList **do**
9: Find a relation involving $\mathfrak{p}$ by enumerating short elements in random power-products.
10: **end for**
11: **return** M.

To assess the advantage of sieving over enumeration techniques, we need to isolate its contribution to the performances of the class group and unit group computation. To do this, we used a modified version of the function `bnfinit` of the computer algebra software PARI that accepts in input a list of precomputed relations. We interfaced via Sage this version of PARI with a development version of Magma containing a function creating relations with the sieving algorithm. The Magma function tries to create enough relations so that the rank of M is 97% of $\#\mathcal{B}$ and passes it to PARI which adds new relations with enumeration methods

and calculates the class group and the unit group. We compared the performance of this approach to the traditional `bnfinit` function of PARI. There are two main reasons for using a hybrid version. The first one is that PARI's implementation of enumeration techniques is the most efficient. As these are necessary to finish the creation of the relation matrix after calling the sieving algorithm, it is interesting to see how the two perform together. Another reason for this choice is the fact that many different algorithms contribute to the computation of the class group and the unit group. In particular, we use time-consuming linear algebra methods such as the HNF computation. Our methodology avoids the risk of seeing the influence of the quality of the implementation of other algorithms occurring in the class group and unit group computation.

We performed our computations on a 2.6 GHz Opteron with 4 GB of memory. We used a branch of the development version 2.6.0 of PARI provided by Loïc Grenié and the development version of Magma, interfaced via Sage 4.7.2. We allocated 3 GB of memory to the computation made with PARI. For each size d, we drew at random 10 number fields with discriminant satisfying $\log_2|\Delta| = d$. For each discriminant, we computed the class group and the unit group with `bnfinit`, which we refer to as the PARI method, and with the hybrid version which we refer to as the PARI+Sieving method. The average timings, in CPU sec (rounded to the nearest integer), are presented in Table 1. They illustrate the impact of sieving methods for small degree number fields. It is very strong for number fields of degree 3, 4, and 5, for which we often witness a speedup by a factor at least 10, while it is rather moderate for degree-6 number fields, and negligible for number fields of degree 7 and 8. Finding smooth values of a polynomial gets more difficult when we increase its degree, but it is not the only reason why the impact of sieving decreases with the degree. Indeed, for degree 6 number field, our sieving algorithm still derives relations at a competitive pace, but there are many linear dependencies

n	$\log_2\|\Delta\|$	PARI	PARI+Sieving	n	$\log_2\|\Delta\|$	PARI	PARI+Sieving
3	120	76	11	5	120	33	18
3	140	694	66	5	140	295	64
3	160	6828	333	5	160	3402	378
3	180	29807	2453	5	180	16048	2342
4	120	38	7	6	120	40	111
4	140	366	24	6	140	294	161
4	160	4266	175	6	160	1709	1012
4	180	31661	1201	6	180	14549	8413

Table 1. Impact of sieving on class group and unit group computation of small degree number fields. Timings in CPU-seconds.

n	Magma 2.18	PARI	PARI+Sieving
20	0.7	0.5	0.2
30	6	5	3
40	22	44	66
45	128	271	556
50	170	593	1562
54	1453	1085	9251

Table 2. Impact of the quadratic sieve on computations in fields generated by a root of $X^2 + 4(10^n + 1)$. Timings in CPU-seconds.

whereas enumeration allows a more targeted search, thus avoiding linear dependencies. To put these improvements into perspective, we show in Table 2 the impact of Jacobson's self-initializing quadratic sieve [25] which is implemented in Magma 2.18. The timings for PARI and PARI+Sieving are derived under the same setting as for Table 1. In addition, we added the performances of Magma 2.18 which uses different methods for linear algebra. Timings for the same series of number fields were reported by Jacobson in [25, Table A.3] on a 296-MHz Sun processor (for a fair comparison one has to take into account the verification time since the timings of Table 1 and Table 2 correspond to a certification under GRH).

5. Computing the unit group

Assume that we have created a relation matrix $(e_{i,j})$ corresponding to the relations

$$(\phi_i) = \mathfrak{p}_1^{e_{i,1}} \cdots \mathfrak{p}_N^{e_{i,N}}.$$

Every kernel vector allows us to derive a unit of $\mathbb{O}_K$. Let $\beta_1, \ldots, \beta_k$ be a generating set of the units created so far. We compute a new unit β', and we wish to find a new minimal generating set for $\langle \beta_1, \ldots, \beta_k, \beta \rangle$. Usually this is done by computing (real) logarithms of the units followed by some approximate linear algebra to find a (tentative) relation as well as the (tentative) new basis. This is then followed by some verification of the relation to guarantee correctness, by using real based computations. The difficulty comes from the fact that the entries in the real matrix differ vastly in size — by several orders of magnitude — thus making it necessary to work with a huge precision; in fact the precision is also subexponential in the discriminant for guaranteed results.

Here, we propose to use p-adic logarithms instead. The key advantage comes from the much better control of error propagation in the linear algebra: Unless division by nonunits happens, linear algebra does not increase errors. However, while the correctness is based on the unproven Leopoldt conjecture about the non-vanishing of the p-adic regulator, this is not a problem in practice: Any relation

found by the p-adic method can easily be verified unconditionally, thus a failure of the algorithm would provide a counterexample to Leopoldt's conjecture.

We start by choosing a prime p such that the p-adic splitting field K_p has moderate degree; here we allow at most degree 2. In practice, we search for the smallest prime $p > 10000$ such that the p-adic splitting field is unramified of degree ≤ 2. Then we have n embeddings $\phi_i : K \to K_p$, and we define a map $L_p : K^* \to K_p^n$ given by $x \mapsto (\log \phi_i(x))_i$, where ϕ_i is the usual p-adic logarithm extended to K_p. In order to estimate the necessary p-adic precision, we also need the usual real logarithmic embedding, denoted by $L : K^* \to \mathbb{R}^{r+1}$. We are looking for a (rational) solution $(x_i)_i \in \mathbb{Q}^{k+1}$ to $\sum x_i L_p(\beta_i) = L_p(\beta')$. Using p-adic linear algebra we will instead get a p-adic solution (or a proof that β' is independent). Using standard rational reconstruction techniques, we derive the rational solution from the p-adic one and then the integral relation between the units. In order to estimate the p-adic precision, we bound numerator and denominator using Cramer's rule and universal lower bounds on the logarithms of units. The rational solution then also satisfies $\sum x_i L(\beta_i) = L(\beta')$. Let $(\alpha_i)_i$ be a basis for $\langle \beta_1, \ldots, \beta_s, \beta' \rangle$. By Cramer's rule,

$$x_i = \det(L_p(\beta_1), \ldots, L_p(\beta'), \ldots, L_p(\beta_s))/\det(L_p(\beta_1), \ldots, L_p(\beta_s)).$$

Since the (unknown) (α_i) form a basis, we see that

$$\det(L_p(\beta_1), \ldots, L_p(\beta'), \ldots, L_p(\beta_s))/\det(L_p(\alpha_1), \ldots, L_p(\alpha_s))$$

is an integer and the same is true for L instead of L_p; thus we can write x_i as a quotient of integers. In either case, to make sense of the determinants, we will have to select an appropriate number of rows to make the matrices square. To bound the integers, we make use of the Hadamard bound for $\det(L(\beta_1), \ldots, L(\beta'), \ldots, L(\beta_s))$ and some universal lower bound for $\det(L(\alpha_i))_i$. For the lower bound we use lower bounds of logarithms of nontorsion units: $\|L(\alpha_i)\|_2 \geq \frac{21}{128}(\log d)/d^2$ (see [17, 3.5]), or, if the unit group has full rank, $s = r = r_1 + r_2 - 1$, we use lower regulator bounds, possibly coming from the Euler product. Having obtained bounds from the real logarithm (L) with low precision, we calculate the p-adic precision required to find x_i using p-adic linear algebra and rational reconstruction. In the course of the computation it can happen that the p-adic determinants (p-adic regulators) have nontrivial valuation. In this case we have to restart the computation with a correspondingly higher precision to account for the loss. Since the Leopoldt conjecture has not been proved, we also need to verify the solution by computing a low-precision estimate for $\left\| \sum x_i L(\beta_i) - L(\beta') \right\|$ to compare it to the lower bound used above.

From the relation x_i we can easily obtain a presentation of the new basis α_i in terms of the β_i, β'. For optimization, we then proceed to compute a new basis $\widetilde{\alpha}_i$ such that the real logarithms are (roughly) LLL-reduced. We note that we do not

rely on any LLL estimates here, so any heuristic algorithm that aims at reducing the apparent size will do. Since we do not have any LLL algorithm that will accept real input (as opposed to rational), it is important that this does not influence the correctness.

5.1. ***Advantages of the p-adic method.*** There are two core advantages of the p-adic logarithms over the ordinary, complex, ones: First, the linear algebra problems we need to solve in order to find dependencies or relations between units have a much simpler error analysis. In fact, contrary to the complex case, it is possible through the use of ring based operations to solve linear equations without any additional loss of precision. This is very important in the context of unit computation since the matrices representing the image of $L(\alpha)$ are very badly conditioned for classical numerical methods. The other advantage of the p-adic logarithms is more subtle: If we assume Leopoldt's conjecture to hold for the field(s) we are interested in, then instead of doing linear algebra over $\mathbb{R}$ with a precision of say q to find dependencies, it is sufficient to work with a real precision of $q/2$ and a p-adic precision of $q/2$ as well. Thus, assuming classical multiplication, we gain a factor of about 4 through the use of lower precision. Using fast multiplication (in high precision), the gain is smaller but still noticeable. But the most important advantage is the much easier precision control: Instead of complicated and very delicate estimates for linear algebra problems, all we need are upper bounds on linear combinations with integral coefficients — which are trivial to obtain.

We should also mention that one disadvantage of the p-adic method lies in the total lack of control over the real size of the units, thus it needs to be paired with a crude (and uncritical for correctness) size reduction algorithm. Also, it is (currently) not possible to avoid completely the use of complex (or real) logarithms, as the p-adic method is not capable to proving a unit to be torsion without knowledge of bounds on the real size.

5.2. ***Lower bound from Euler product.*** Suppose that, as in the class group algorithm, we are given an approximation of the Euler product; that is, we have a real number E such that $1/\sqrt{2} \leq hR/E \leq \sqrt{2}$. After the relation matrix has full rank, and assuming the factor base is large enough for correctness, we have an upper bound for the class number, thus a lower bound for R. This lower bound will be several orders of magnitude larger than the universal bounds available otherwise.

5.3. ***Saturation.*** After the initial steps of the algorithm, when the relation matrix has full rank, we have a tentative class number h and a tentative regulator R. Experimentally, at this point, hR does not approximate the Euler product very well — the product will be off by several orders of magnitude. However, after finding one or two more relations, the product has the same size as the Euler product;

it frequently even looks like only a factor of 2 is missing in either h or R. To find the last missing relation can easily take more time than the entire previous run, therefore we suggest using saturation methods instead. At this point in the algorithm the relations define a subgroup U of the S-unit group U_S where S is the factor basis. From the Euler product we know that the index $(U_S : U) =: b$ is small, let's say $b < B$. For any prime $p \mid b$ there is some $u \in U_S \setminus U$ such that $u^p \in U$. Let us fix the prime p. For any prime ideal $Q \notin S$ such that $p \mid \mathcal{N}(Q) - 1$ we can define the map $\phi_Q : U \to \mathbb{F}_Q^*/(\mathbb{F}_Q^*)^p$ mapping S-units into the multiplicative group of the residue class field modulo p-th powers. The Chebotarev theorem [44] guarantees that if $u \in U$ is not a p-th power, there will be some Q such that $\phi_Q(u)$ is nontrivial, that is, u is not a p-th power modulo Q. We now simply intersect $\ker \phi_Q$ for several Q until either the intersection is U^p or it does not change for five consecutive Q. We expect that any $u \in U/\bigcap \ker \phi_Q$ will have a p-th root in U_S but not in U. Therefore $v^p = u$ is a new relation that will change hR by p. Repeating this for all $p < B$ until we cannot enlarge U any more we find the missing relations. Similar techniques have been used a long time but were confined to the unit group [45; 36]. This appears to be the first time that saturation has been applied to the full relation lattice.

5.4. ***Representation.*** During the execution of the algorithm, all (S-)units are naturally represented as power products of the relations coming from the sieving (or the saturation). It is well known that the explicit representation of the units with respect to a fixed basis for the field can require exponentially large coefficients, so it is important to operate on the power products as much as possible. However, even the exponent vectors constructed for the basis of the unit group, or the saturation, will become huge, so we need to "size reduce" the power products. In particular, this happens even if the resulting element is not too large. Using ideas of [13] for compact representations and [22] for reduced divisors in function fields, we can find a representation for those elements that depends only on the logarithmic size (and the number field) rather than the execution path. For any prime p we can write any unit $u = \prod r_i^{e_i} = \prod a_i^{p^i}$ with elements such that the size of a_i depends on the discriminant and p only. The length of the product comes from $L(u)$. Furthermore, in this presentation it is easy to test for p-th powers as only a_0 needs to be tested and this is a small element.

5.5. ***Example.*** To illustrate the power of the p-adic method, we look at a totally real quartic field generated by a root of

$$x^4 + 17211x^3 + 5213x^2 - 176910463x - 4958.$$

The discriminant Δ of the maximal order has 38 digits. In the course of the computation, we found 534 relations involving prime ideals of norm up to 3000 =

$0.4 \log^2|\Delta|$ describing a trivial class group. We then searched for 5 further relations to obtain units u_i ($1 \leq i \leq 5$). As power products of the relations, the units are given via exponent vectors e_i with $\|e_i\|_\infty$ ranging between 10^{80} and 10^{160} and $20 < \|e_i\|_1/\|e_i\|_\infty < 92$. So, while not uniformly large, the exponents are nonsparse, involving huge integers. Using a decimal precision of 170 digits, we establish that the logarithms of the units are roughly $\|L(u_i)\|_\infty \approx 10^{160}$. The first three units are indeed independent, giving a basis for a subgroup of full rank, the fourth is then dependent. Choosing the prime $p = 10337$ we get $\mathbb{Q}_p$ as a splitting field. Using a p-adic precision of 245 digits (that is, working in $\mathbb{Z}_p \bmod p^{245}$), we compute the dependency for the fourth unit, involving exponents of around 10^{360}. The new unit group is then tentatively LLL reduced, producing a new basis where the $\|L(\tilde{u}_i)\|_\infty$ are bounded by 10^7 only. The last unit then involves a much smaller dependency, here the exponents are only around 10^{60}.

Unfortunately, looking at the Euler product, the unit group is not complete. However, the saturation technique outlined above takes 1 sec to determine that the product of the three basis elements is (probably) a square. Finding a better representation where the exponents are all powers of 2 takes less than 1 sec and then we can enlarge the unit group easily.

Due to the implementation, the p-adic precision used was actually higher: Changing (increasing) precision is very computationally expensive, so we try to avoid this and simply double the precision. We used a precision of 320 for the p-adics and a maximal precision of 1000 for the real precision. The computation of the log is the dominating part: We spent 50 sec or 90% of the total processing time here.

6. Conclusion

We introduced new techniques to enhance the performances of the subexponential methods for computing the class group and the unit group of a number field. In particular, sieving allows a speedup of an order of magnitude for number fields of small degree. These techniques could be developed even further. Indeed, we have not taken into account all the improvements to sieving techniques described in the context of the number field sieve algorithm, such as large prime variations or cache-friendly methods. It is also notable that fast techniques for deriving relations in the class group of a small degree number field have applications in evaluating isogenies between small genus curves via complex multiplication methods. Indeed, in that case, evaluating isogenies between genus g curves involves relations in the class group of a degree $2g$ number field.

Acknowledgments

The first author is particularly grateful to Loïc Grenié for providing a special branch of PARI allowing to start the class group computation from an existing relation

matrix. He also thanks David Roe for helping with the Sage interface between Magma and PARI. The research presented in this paper was carried out while both authors were working in Sydney with the Magma group.

References

[1] J. V. Armitage (ed.), *Journées Arithmétiques*, 1980: *Lectures from the Conference held at the University of Exeter*, *Exeter*, *April* 13–19, 1980, London Mathematical Society Lecture Note Series, no. 56, Cambridge University Press, 1982. MR 84c:10003

[2] Eric Bach, *Explicit bounds for primality testing and related problems*, Math. Comp. **55** (1990), no. 191, 355–380. MR 91m:11096

[3] ———, *Improved approximations for Euler products*, in Dilcher [15], 1995, pp. 13–28. MR 96i:11124

[4] Karim Belabas, Francisco Diaz y Diaz, and Eduardo Friedman, *Small generators of the ideal class group*, Math. Comp. **77** (2008), no. 262, 1185–1197. MR 2009c:11179

[5] T. Beth, N. Cot, and I. Ingemarsson (eds.), *Advances in cryptology: Proceedings of the EUROCRYPT 84 workshop on the theory and application of cryptographic techniques held in Paris, April 9–11, 1984*, Lecture Notes in Computer Science, no. 209, Berlin, Springer, 1985. MR 86m:94003

[6] J.-F. Biasse, *An $L(1/3)$ algorithm for ideal class group and regulator computation in certain number fields*, 2012, to appear in *Math. Comp.*

[7] Jean-François Biasse, *Improvements in the computation of ideal class groups of imaginary quadratic number fields*, Adv. Math. Commun. **4** (2010), no. 2, 141–154. MR 2011e:11192

[8] Yuri Bilu and Guillaume Hanrot, *Solving Thue equations of high degree*, J. Number Theory **60** (1996), no. 2, 373–392. MR 97k:11040

[9] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system, I: The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[10] Wieb Bosma and Alf van der Poorten (eds.), *Computational algebra and number theory*: *Papers from the CANT*2 *Meeting held at Sydney University*, *November* 1992, Mathematics and its Applications, no. 325, Kluwer Acad. Publ., Dordrecht, 1995. MR 96c:00019

[11] Johannes Buchmann, *A subexponential algorithm for the determination of class groups and regulators of algebraic number fields*, in Goldstein [20], 1990, pp. 27–41. MR 92g:11125

[12] Johannes Buchmann, Michael J. Jacobson, Jr., Stefan Neis, Patrick Theobald, and Damian Weber, *Sieving methods for class group computation*, in Matzat et al. [33], 1999, pp. 3–10. MR 2000a:11177

[13] Johannes Buchmann, Christoph Thiel, and Hugh Williams, *Short representation of quadratic integers*, in Bosma and van der Poorten [10], 1995, pp. 159–185. MR 96c:11144

[14] H. Cohen and H. W. Lenstra, Jr., *Heuristics on class groups of number fields*, in Jager [27], 1984, pp. 33–62. MR 85j:11144

[15] Karl Dilcher (ed.), *Number theory*: *Proceedings of the Fourth Conference of the Canadian Number Theory Association held at Dalhousie University*, *Halifax*, *July* 2–8, 1994, CMS Conference Proceedings, no. 15, Providence, RI, American Mathematical Society, 1995. MR 96c:11003

[16] Claus Fieker, *Minimizing representations over number fields, II: Computations in the Brauer group*, J. Algebra **322** (2009), no. 3, 752–765. MR 2010e:20016

[17] Claus Fieker and Michael E. Pohst, *Dependency of units in number fields*, Math. Comp. **75** (2006), no. 255, 1507–1518. MR 2007a:11168

[18] U. Fincke and M. Pohst, *A procedure for determining algebraic integers of given norm*, in van Hulzen [23], 1983, pp. 194–202. MR 86k:11078

[19] Jens Franke and Thorsten Kleinjung, *Continued fractions and the lattice sieving*, paper presented at the Special-purpose Hardware for Attacking Cryptographic Systems (SHARCS) conference, Paris, February 24–25, 2005. http://www.ruhr-uni-bochum.de/itsc/tanja/SHARCS/talks/FrankeKleinjung.pdf

[20] Catherine Goldstein (ed.), *Séminaire de théorie des nombres, Paris 1988–1989*, Progress in Mathematics, no. 91, Birkhäuser, Boston, 1990. MR 91k:11004

[21] James L. Hafner and Kevin S. McCurley, *A rigorous subexponential algorithm for computation of class groups*, J. Amer. Math. Soc. **2** (1989), no. 4, 837–850. MR 91f:11090

[22] Florian Heß, *Zur Divisorklassengruppenberechnung in globalen Funktionenkörpern*, Ph.D. thesis, Technische Universität Berlin, 1999. http://page.math.tu-berlin.de/~kant/publications/diss/diss_FH.ps.gz

[23] J. A. van Hulzen (ed.), *Computer algebra: Proceedings of the European computer algebra conference (EUROCAL) held in London, March 28–30, 1983*, Lecture Notes in Computer Science, no. 162, Berlin, Springer, 1983. MR 86f:68004

[24] M. J. Jacobson, Jr., Á. Pintér, and P. G. Walsh, *A computational approach for solving* $y^2 = 1^k + 2^k + \cdots + x^k$, Math. Comp. **72** (2003), no. 244, 2099–2110. MR 2004c:11241

[25] Michael J. Jacobson, Jr., *Subexponential class group computation in quadratic orders*, Ph.D. thesis, Technische Universität Darmstadt, Aachen, Germany, 1999. http://www.shaker.eu/shop/978-3-8265-6374-4

[26] Michael J. Jacobson, Jr. and Hugh C. Williams, *Solving the Pell equation*, Springer, New York, 2009. MR 2009i:11003

[27] H. Jager (ed.), *Number theory, Noordwijkerhout 1983: Proceedings of the thirteenth Journées Arithmétiques held at Noordwijkerhout, July 11–15, 1983*, Lecture Notes in Mathematics, no. 1068, Berlin, Springer, 1984. MR 85i:11001

[28] Thorsten Kleinjung, Kazumaro Aoki, Jens Franke, Arjen K. Lenstra, Emmanuel Thomé, Joppe W. Bos, Pierrick Gaudry, Alexander Kruppa, Peter L. Montgomery, Dag Arne Osvik, Herman te Riele, Andrey Timofeev, and Paul Zimmermann, *Factorization of a 768-bit RSA modulus*, in Rabin [40], 2010, pp. 333–350. MR 2725602

[29] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The development of the number field sieve*, Lecture Notes in Mathematics, no. 1554, Springer, Berlin, 1993. MR 96m:11116

[30] H. W. Lenstra, Jr., *On the calculation of regulators and class numbers of quadratic fields*, in Armitage [1], 1982, pp. 123–150. MR 86g:11080

[31] Donald J. Lewis (ed.), *Proceedings of the* 1969 *Summer Institutes on Number Theory*: *Analytic Number Theory*, *Diophantine Problems*, *and Algebraic Number Theory*; *held at the State University of New York at Stony Brook*, *July* 7–*August* 1, 1969, Proceedings of Symposia in Pure Mathematics, no. 20, American Mathematical Society, Providence, R.I., 1971. MR 47 #3286

[32] J. E. Littlewood, *On the class-number of the corpus* $P(\sqrt{-k})$, Proc. London Math. Soc. **S2-27** (1928), no. 1, 358. MR 1575396

[33] B. Heinrich Matzat, Gert-Martin Greuel, and Gerhard Hiss (eds.), *Algorithmic algebra and number theory: Selected papers from the conference held at the University of Heidelberg, October 1997*, Springer, Berlin, 1999. MR 99h:00020

[34] Stefan Neis, *Zur Berechnung von Klassengruppen*, Ph.D. thesis, Technische Universität Darmstadt, 2002. http://tuprints.ulb.tu-darmstadt.de/epda/000283/

[35] PARI Group, Bordeaux, France, *PARI/GP, version 2.5.0*, 2011. http://pari.math.u-bordeaux.fr/

[36] M. Pohst and H. Zassenhaus, *Algorithmic algebraic number theory*, Encyclopedia of Mathematics and its Applications, no. 30, Cambridge University Press, 1989. MR 92b:11074

[37] Michael Pohst and Hans Zassenhaus, *Über die Berechnung von Klassenzahlen und Klassengruppen algebraischer Zahlkörper*, J. Reine Angew. Math. **361** (1985), 50–72. MR 87g:11147

[38] J. M. Pollard, *The lattice sieve*, in Lenstra and Lenstra [29], 1993, pp. 43–49. MR 1321220

[39] Carl Pomerance, *The quadratic sieve factoring algorithm*, in Beth et al. [5], 1985, pp. 169–182. MR 87d:11098

[40] Tal Rabin (ed.), *Advances in cryptology—CRYPTO 2010: Proceedings of the 30th Annual International Conference held in Santa Barbara, CA, August 15–19, 2010*, Lecture Notes in Computer Science, no. 6223, Berlin, Springer, 2010. MR 2012c:94002

[41] Daniel Shanks, *Class number, a theory of factorization, and genera*, in Lewis [31], 1971, pp. 415–440. MR 47 #4932

[42] ———, *The infrastructure of a real quadratic field and its applications*, Proceedings of the Number Theory Conference (Boulder, CO), Univ. Colorado, 1972, pp. 217–224. MR 52 #10672

[43] Christoph Thiel, *On the complexity of some problems in algorithmic algebraic number theory*, Ph.D. thesis, Universität des Saarlandes, 1995. http://www.cdc.informatik.tu-darmstadt.de/reports/reports/Christoph_Thiel.diss.pdf

[44] N. Tschebotareff, *Die Bestimmung der Dichtigkeit einer Menge von Primzahlen, welche zu einer gegebenen Substitutionsklasse gehören*, Math. Ann. **95** (1926), 191–228. JFM 51.0149.04

[45] Klaus Wildanger, *Über Grundeinheitenberechnung in algebraischen Zahlkörpern*, master's thesis, Heinrich Heine Universität Düsseldorf, 1993. http://page.math.tu-berlin.de/~kant/publications/diplom/wildanger.pdf

JEAN-FRANÇOIS BIASSE: biasse@lix.polytechnique.fr
Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

CLAUS FIEKER: fieker@mathematik.uni-kl.de
Fachbereich Mathematik, Universität Kaiserslautern, Postfach 3049, D-67653 Kaiserslautern, Germany

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.135

msp

Conditionally bounding analytic ranks of elliptic curves

Jonathan W. Bober

We describe a method for bounding the rank of an elliptic curve under the assumptions of the Birch and Swinnerton-Dyer conjecture and the generalized Riemann hypothesis. As an example, we compute, under these conjectures, exact upper bounds for curves which are known to have rank at least as large as $20, 21, 22, 23$, and 24. For the known curve of rank at least 28, we get a bound of 30.

1. Introduction

Determining the rank of an elliptic curve is a difficult problem, and there is currently no known unconditional algorithm for determining the rank of a given curve. The basic method for rigorously determining the rank of a curve is to find an upper bound for the rank by computing the size of some Selmer groups and to find a lower bound for the rank by finding enough independent rational points. In theory, if one continues this process long enough, and the Shafarevich-Tate group of the curve is finite, the upper and lower bounds should eventually coincide and the rank will be determined exactly.

In practice, things are not so simple. Finding points on the curve is sometimes not too bad, but the upper bounds for the rank are more problematic. Even the computation of the 2-Selmer rank is difficult, and it becomes prohibitively time-consuming as the coefficients of the elliptic curve grow; it is easy to write down a curve for which the state-of-the-art program for computing the 2-Selmer group, John Cremona's `mwrank` [5], will effectively take "forever."

If one is willing to accept the Birch and Swinnerton-Dyer conjecture that the rank of an elliptic curve is the same as the order of vanishing of its L-function at the central point, then it is possible to use the L-function to get information

MSC2010: primary 11M41; secondary 14G10.
Keywords: elliptic curve, rank, L-function, explicit formula.

about the rank. In fact, when the order of vanishing is between 0 and 3, it can be possible to compute the L-function to enough precision and use some extra information about the curve to determine the analytic rank exactly, as is done in [3], for example. When the rank is larger than this, though, currently the best one can do is determine that the first r derivatives of the L-function are very close to 0 and the $(r+1)$-st is not, which will provide a very good guess for the rank and a rigorous upper bound, assuming BSD.

This approach has its own problems, as it is much easier to write down a curve of large conductor than it is to compute the L-function of such a curve. For example, the known curve of rank at least 28 [8], which we will write down later, has conductor $N \approx 3.5 \times 10^{141}$, and current methods (such as those described in [19]) typically require summing on the order of $\sqrt{N}$ terms to compute the central value of the L-function. (It would take a computer about 10^{53} cpu-years just to add 1 to itself 10^{70} times.)

We present here a third method which is rather effective at bounding the rank, especially when the rank is large compared to the conductor, as long as one is willing to assume both the Birch and Swinnerton-Dyer conjecture and the Riemann Hypothesis for the L-function of the curve. This method is not completely new. It is based on Mestre's method [14] for (conditionally) bounding the rank of an elliptic curve based only on its conductor, and it was used by Fermigier [9] to study ranks of elliptic curves in certain families. However, it does not seem to have gained much traction and does not seem to have been used much, if at all, since.

The idea, in brief, is as follows. Take $f(x)$ to be a function such that $f(0) = 1$ and $f(x) \geq 0$ for all real x. Then, assuming the Riemann hypothesis, the sum $\sum f(\gamma)$, where $1/2 + i\gamma$ runs over the nontrivial zeros of $L(s, E)$ (counted with multiplicity), will be an upper bound for the analytic rank of E. Moreover, for certain choices of $f(x)$ this sum may be efficiently evaluated using the explicit formula for the L-function attached to E.

This method has recently been implemented by the author, and is available as part of William Stein's PSAGE [21] add-ons to Sage [22]. As an example of what it can do, we will examine 6 curves known to have rather large rank. We denote these curves by E_n, where the index n, taking the values $20, 21, 22, 23, 24, 28$ represents a known lower bound for the rank. We will write down these curves later (they are all taken from A. Dujella's website [6], and at the time of discovery each held the record for the curve with largest number of known independent rational points). The exact rank is not known for any of these curves. However, conditionally we may claim:

Theorem 1.1. *Assuming BSD and GRH, E_n has rank exactly n for $n = 20, 21, 22,$ 23, and 24, while E_{28} has rank 28 or 30.*

Remark 1.2. Around the time that I was writing this paper, Andrew Booker and Jo Dwyer were able to exactly compute the rank of E_{28}, again assuming the Birch and Swinnerton-Dyer conjecture and the Riemann Hypothesis for $L(s, E_{28})$. They use the method described here, but by using the optimization procedure described in Section 3 of [1] they are able to select a better test function as input to the explicit formula, and they get a correspondingly better bound.

2. Bounding ranks

2A. ***The method.*** Let

$$L(s, E) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} = \prod_p L_p(s, E)^{-1}$$

be the L-function of an elliptic curve, normalized so that the completed L-function $\Lambda(s, E)$ satisfies the functional equation $\Lambda(s, E) = \epsilon\Lambda(1 - s, E)$, and let c_n be defined by

$$-\frac{L'(s, E)}{L(s, E)} = \sum_{n=1}^{\infty} \frac{c_n}{n^s}.$$

More explicitly, if we define $\alpha(p)$ and $\beta(p)$ by

$$L_p(s, E) = (1 - \alpha(p)p^{-s})(1 - \beta(p)p^{-s}),$$

(note that α and β are only well defined up to permutation, and that at least one of them will be 0 when p is a prime of bad reduction), then

$$c_{p^m} = \big(\alpha(p)^m + \beta(p)^m\big) \log p,$$

and $c_n = 0$ when n is not a prime power.

Our main tool will be the explicit formula for $L(s, E)$, which we state in a friendly form in the following lemma.

Lemma 2.1. *Suppose that $f(z)$ is an entire function with $f(x + iy) \ll x^{-(1+\delta)}$ for $|y| < 1 + \epsilon$, for some $\epsilon > 0$, and that the Fourier transform of f*

$$\hat{f}(y) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ixy}\mathrm{d}x$$

exists and is such that

$$\sum_{n=1}^{\infty} \frac{c_n}{n^{1/2}} \hat{f}\left(\frac{\log n}{2\pi}\right)$$

converges absolutely. Then

$$\sum_{\gamma} f(\gamma) = \hat{f}(0)\frac{\log N}{2\pi} - \hat{f}(0)\frac{\log 2\pi}{\pi} + \frac{1}{\pi}\Re\left\{\int_{-\infty}^{\infty} \frac{\Gamma'}{\Gamma}(1+it)f(t)\mathrm{d}t\right\} - \frac{1}{2\pi}\sum_{n=1}^{\infty}\frac{c(n)}{n^{1/2}}\left(\hat{f}\left(\frac{\log n}{2\pi}\right) + \hat{f}\left(-\frac{\log n}{2\pi}\right)\right), \quad (1)$$

where $1/2 + i\gamma$ runs over the nontrivial zeros of $L(s, E)$, where E is an elliptic curve with conductor N.

Proof. A proof of the explicit formula in this form, or in a similar form, can be found in various sources — for example, [11, Theorem 5.12] — so we give only a brief sketch. The idea is to integrate the function

$$F(s)\frac{L'(s, E)}{L(s, E)},$$

where $F(1/2 + is) = f(s)$, on a vertical line to the right of the critical strip and, in the reverse direction, on a vertical line to the left of the critical strip. By the residue theorem, this integral will be equal to $2\pi\sum_{\gamma} f(\gamma)$. One now applies the functional equation to write the integral in the left half-plane as an integral in the right half-plane.

The sum over the Fourier coefficients of f arises from shifting contours to the region of absolute convergence and using the Dirichlet series for $L'(s)/L(s)$, while the other terms arise from shifting the remaining integrals to the line $\Re(s) = 1/2$.

The conditions on $f(z)$ are exactly those needed to make sure that this process can go through without trouble. Of course, it is also important that $L(s, E)$ is entire and that it satisfies a functional equation [25; 24; 2]. □

A convenient function to use in an application of the explicit formula is

$$f(z) = f(z; \Delta) = \left(\frac{\sin(\Delta\pi z)}{\Delta\pi z}\right)^2,$$

which has the simple Fourier transform

$$\hat{f}(x; \Delta) = \left(\frac{1}{\Delta}\right)\left(1 - \left|\frac{x}{\Delta}\right|\right), \quad |x| < \Delta.$$

With this choice of f, Equation (1) takes the form

$$\sum_{\gamma} f(\gamma; \Delta) = \frac{\log N}{\Delta 2\pi} - \frac{\log 2\pi}{\Delta\pi} + \frac{1}{\pi}\Re\left\{\int_{-\infty}^{\infty}\frac{\Gamma'}{\Gamma}(1+it)f(t; \Delta)\mathrm{d}t\right\} - \frac{1}{\Delta\pi}\sum_{p \le \exp(2\pi\Delta)}\log p \sum_{k=1}^{\lfloor 2\pi\Delta/\log p\rfloor}\frac{1}{p^{k/2}}\left(\alpha(p)^k + \beta(p)^k\right)\left(1 - \frac{k\log p}{2\pi\Delta}\right). \quad (2)$$

Since $f(\gamma;\Delta)\geq 0$ as long as γ is real, and $f(0;\Delta)=1$, Equation (2) will give an upper bound for the order of vanishing of $L(s,E)$ at $s=1/2$, as long as the Riemann Hypothesis holds for $L(s,E)$. And if Δ is not too large, we can quickly evaluate the right-hand side of Equation (2) to calculate this upper bound. It is also worth noting that, assuming RH,

$$-\lim_{\Delta\to\infty}\frac{1}{\Delta\pi}\sum_{p\leq\exp(2\pi\Delta)}\log p\sum_{k=1}^{\lfloor 2\pi\Delta/\log p\rfloor}\frac{1}{p^{k/2}}\left(\alpha(p)^k+\beta(p)^k\right)\left(1-\frac{k\log p}{2\pi\Delta}\right)$$
$$=\operatorname{ord}_{s=1/2}L(s,E)$$

so that, in principle, we should be able to get as good a bound for the rank as we like through this method. However, as the length of the prime sum grows exponentially in Δ, this method quickly becomes infeasible once Δ gets a little larger than 4.

2B. ***Some curves.*** As an example, we examine 6 elliptic curves from Dujella's online tables. They are

$$E_{20}\colon y^2+xy=x^3-43109298076633367795836209589116 6x$$
$$+51562835553666436590356527998711769093915330881 96,$$

$$E_{21}\colon y^2+xy+y=x^3+x^2-21584377242244392201516995270215983 5x$$
$$-19474361277787151947255961435459054151501792241320535,$$

$$E_{22}\colon y^2+xy+y=x^3-94029951777639136290302312116586 4x$$
$$+10707363070719743033425295515449274534651125011362,$$

$$E_{23}\colon y^2+xy+y=x^3-19252966408674012828065964616418441723x$$
$$+32685500727716376257923347071452044295907443056345614006,$$

$$E_{24}\colon y^2+xy+y=x^3-12003982203699224530353461919116679637 4x$$
$$+50422499248491067001080179916808272675944375622291141 5116,$$

and

$$E_{28}\colon y^2+xy+y=x^3-x^2-\left(\begin{matrix}20067762415575526585033208\times 10^{30}\\ +\ 209338542750930230312178956502\end{matrix}\right)x$$
$$+\left(\begin{matrix}34481611795030556467032985690390720374855 94\times 10^{40}\\ +\ 43593191803612660082962919394487322434 29\end{matrix}\right).$$

Each E_n has n known independent rational points of infinite order, so has at least rank n. (See [16; 17; 10; 12; 13; 8], or [6] for quick reference.) Using

Curve	$\log N_E$	Δ	$\sum_\gamma f(\gamma;\Delta)$	$\frac{\log N_E}{2\pi\Delta}$
E_{20}	170.09	2.0	21.70	13.54
E_{21}	196.68	2.5	22.68	12.52
E_{22}	182.72	2.0	23.71	14.54
E_{23}	205.06	2.5	24.49	13.05
E_{24}	219.93	2.5	25.57	14.00
E_{28}	325.90	3.2	31.30	16.21

Table 1. Computed upper bounds for the ranks of some curves, along with a heuristic guess of what these bounds should for a typical elliptic curve. The sum over the zeros here is rounded up; other numbers are rounded to nearest.

the methods described above, we compute rank bounds for each of these curves. These are listed in Table 1. The global root number can be computed for each curve. (In Sage, `E.root_number()`, which uses PARI [18], will finish quickly for E_{20}, E_{21}, and E_{22} and within a few hours for E_{23} and E_{24}. For E_{28} it is best to see the mailing list discussion which gives the factorization of the discriminant [7].) In each case the root number agrees with the parity of the known number of independent points, so to get a tight upper bound for the rank we only need to get within 2 of the number of known independent points, and so the computation in Table 1 gives the proof of Theorem 1.1.

2C. ***Curves of small conductor.*** For further testing, this method was also run on all elliptic curves up with conductor below 180000 (from Cremona's tables [4]) using $\Delta = 2.0$, a computation which ran in under a day on a fast 8 core computer. In this range there are 790677 isogeny classes of elliptic curves, and for all but 9882 isogeny classes it turns out that

$$\left\lfloor \sum_\gamma f(\gamma; 2.0) \right\rfloor = \operatorname{rank}(E);$$

in the remaining cases,

$$\left\lfloor \sum_\gamma f(\gamma; 2.0) \right\rfloor = \operatorname{rank}(E) + 1,$$

so consideration of the root number of the curve gives the exact rank.

3. Further comments

3A. ***Some evidence towards BSD.*** There is a way in which these computations can be seen as giving mild evidence in support of the Birch and Swinnerton-Dyer conjecture. The upper bound computed for a curve E is the value of the sum

$\sum_\gamma f(\gamma;\Delta)$, and as $f(\gamma;\Delta)$ decays fairly rapidly as γ grows, one does not expect this sum to be very large for a typical elliptic curve.

To obtain a crude approximation to what we might expect the value of this sum to be, consider that the local zero density of a typical $L(s, E)$ near the central point is approximately $2\pi/\log N_E$. Then, if the zeros are spaced uniformly at random (an assumption that is not really correct, but is close enough to true for our crude purposes), we might expect that

$$\sum_\gamma f(\gamma,\Delta) \approx \frac{\log N_E}{2\pi}\int_{-\infty}^{\infty} f(t;\Delta)\mathrm{d}t = \frac{\log N_E}{2\pi\Delta},$$

possibly with a small adjustment to take into account the parity of the rank. (More precisely, we might expect that if we average this sum over all elliptic curves of conductor close to N_E, the answer will not be too far from this integral.) Thus, when this sum is significantly larger than this estimate, it indicates an extreme concentration of zeros near the central point. (It is also possible to arrive at more refined version of this heuristic by considering the explicit formula. In such a case, it is necessary to assume that the family of elliptic curves considered is large enough that $a_p(E)$ averages to zero for each p, and we notice that the integral of the Γ-factor plays a small role as well.)

As some further small evidence for this heuristic, we note that the average of

$$\frac{4\pi}{\log N}\sum_\gamma f(\gamma;2.0)$$

over all isogeny classes up to 180000 is approximately .9638. The small difference from 1 should be accounted for by the Γ-factor, which tends to push zeros away from the central point.

It should also be possible to refine this heuristic somewhat to make a guess as to what the sum should be for a high rank curve by making the assumption that a zero of high order at the central point will push other zeros away.

3B. ***Correctness tests.*** The method described here is simple enough that it is easy to implement, which reduces the likeliness of bugs. It is still important to test it where possible, however, in order to have more confidence in its correctness.

As described in Section 2C, this code was run on every isogeny class up to conductor 180000, and the fact that the computed upper bound for the rank was never too small gives some confidence that the computation was done correctly. As a further test, one can also compute many zeros for the L-function of an elliptic curve of small conductor, compute the sum over zeros directly, and verify that it agrees with our explicit formula implementation. Table 2 lists some example curves with small conductor for which this was done. The agreement there is

Δ	E	# zeros	Direct	Equation (2)	Difference
2.0	11a	200000	0.00270875	0.00269961	9.17×10^{-6}
	15a	200000	0.00483749	0.00482836	9.13×10^{-6}
	17a	200000	0.00559516	0.00558605	9.11×10^{-6}
	37a	200000	1.00369174	1.00368272	9.01×10^{-6}
	118a	200000	1.00636141	1.00635255	8.86×10^{-6}
	389a	159650	2.00947449	2.00946618	8.30×10^{-6}
	5077a	85520	3.01508240	3.01507647	5.92×10^{-6}
	11197a	70950	3.02102728	3.02102250	4.77×10^{-6}
2.5	11a	200000	0.00172459	0.00172653	1.94×10^{-6}
	15a	200000	0.00170962	0.00171159	1.96×10^{-6}
	17a	200000	0.00250017	0.00250215	1.97×10^{-6}
	37a	200000	1.00335149	1.00335352	2.03×10^{-6}
	118a	200000	2.00585774	2.00586023	2.49×10^{-6}
	389a	159650	3.00797500	3.00797902	4.02×10^{-6}
	5077a	85520	1.00543612	1.00543825	2.14×10^{-6}
	11197a	70950	3.01798029	3.01798504	4.75×10^{-6}

Table 2. Sum of $f(\gamma; 2.0)$ and $f(\gamma; 2.5)$ computed directly with many zeros and using our implementation of (2). The curve labels correspond to isogeny classes in Cremona's tables [4] and the zeros were computed using Rubinstein's `lcalc` [20].

between 10^{-5} and 10^{-6}, which is roughly the precision to which the integral in the explicit formula was calculated, and is in line with what should be expected using what is a fairly small number of zeros.

Acknowledgments

Most of the computations in this paper run in a short amount of time, and were done on the author's personal computer. Some longer computations were run on the sage cluster at the University of Washington, supported by NSF grant DMS-0821725, and the riemann cluster at the University of Waterloo, funded by the Canada Foundation for Innovation, the Ontario Innovation Trust, and SGI.

The source code for our implementation is available as part of PSAGE [21]. It uses Sage [22], and hence PARI [18], to compute a_p for bad primes, and uses Andrew Sutherland's `smalljac` [23] to compute all other values of a_p.

Parts of this work began while the author was in residence at the Mathematical Sciences Research Institute during the Arithmetic Statistics program, Spring 2011, during which time the author was partially supported by NSF grant DMS-0441170,

administered by MSRI. Discussions during the informal "explicit formula seminar," especially with David Farmer and Michael Rubinstein, were influential in encouraging this work.

Currently the author is supported by NSF grant DMS-0757627, administered by the American Institute of Mathematics.

The author would also like to thank Allan MacLeod for pointing out a small but important typo in an earlier version of this paper.

References

[1] Andrew R. Booker, *Artin's conjecture*, *Turing's method*, *and the Riemann hypothesis*, Experiment. Math. **15** (2006), no. 4, 385–407. MR 2007k:11084

[2] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* $\mathbb{Q}$: *Wild* 3*-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939. MR 2002d:11058

[3] Joe P. Buhler, Benedict H. Gross, and Don B. Zagier, *On the conjecture of Birch and Swinnerton-Dyer for an elliptic curve of rank* 3, Math. Comp. **44** (1985), no. 170, 473–481. MR 86g:11037

[4] John Cremona, *Elliptic curve data*, 2012. http://homepages.warwick.ac.uk/staff/J.E.Cremona/ftp/data/INDEX.html

[5] ———, `mwrank`, 2012. http://homepages.warwick.ac.uk/∼masgaj/mwrank/

[6] Andrej Dujella, *History of elliptic curve rank records*, 2012. http://web.math.hr/∼duje/tors/rankhist.html

[7] Noam Elkies, John Cremona, Bruce Dodson, Koh-ichi Nagao, and Bjorn Poonen, *Z^28 in E(Q), etc.*, NMBRTHRY listserv, 2006. http://tinyurl.com/ElkiesEtAlZ28

[8] Noam D. Elkies, *Elliptic curves and surfaces of high rank*, *I*, *II*, *III*, Tech. Report 34/2007, Mathematisches Forschungsinstitut Oberwolfach, 2007, expanded version at arXiv:0709.2908 [math.NT]. http://www.mfo.de/document/0729/OWR_2007_34.pdf

[9] Stéfane Fermigier, *Étude expérimentale du rang de familles de courbes elliptiques sur* $\mathbb{Q}$, Experiment. Math. **5** (1996), no. 2, 119–130. MR 98g:11061

[10] ———, *Une courbe elliptique définie sur* $\mathbb{Q}$ *de rang* ≥ 22, Acta Arith. **82** (1997), no. 4, 359–363. MR 98j:11041

[11] Henryk Iwaniec and Emmanuel Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, no. 53, American Mathematical Society, Providence, RI, 2004. MR 2005h:11005

[12] Roland Martin and William McMillen, *An Elliptic Curve/Q of rank* 23, NMBRTHRY listserv, 16 March 1998. http://tinyurl.com/MartinMcMillen1

[13] ———, *An Elliptic Curve over Q with Rank at least* 24, NMBRTHRY listserv, 2 May 2000. http://tinyurl.com/MartinMcMillen2

[14] Jean-François Mestre, *Formules explicites et minorations de conducteurs de variétés algébriques*, Compositio Math. **58** (1986), no. 2, 209–232. MR 87j:11059

[15] F. Mezzadri and N. C. Snaith (eds.), *Recent perspectives in random matrix theory and number theory*, London Mathematical Society Lecture Note Series, vol. 322, Cambridge University Press, Cambridge, 2005. MR 2006c:11002

[16] Koh-ichi Nagao, *An example of elliptic curve over* $\mathbb{Q}$ *with rank* ≥ 20, Proc. Japan Acad. Ser. A Math. Sci. **69** (1993), no. 8, 291–293. MR 95a:11052

[17] Koh-ichi Nagao and Tomonori Kouya, *An example of elliptic curve over $\mathbb{Q}$ with rank ≥ 21*, Proc. Japan Acad. Ser. A Math. Sci. **70** (1994), no. 4, 104–105. MR 95e:11063

[18] The PARI Group, *PARI/GP* (*version* 2.4.3), 2011. http://pari.math.u-bordeaux.fr/

[19] Michael Rubinstein, *Computational methods and experiments in analytic number theory*, in Mezzadri and Snaith [15], 2005, pp. 425–506. MR 2006d:11153

[20] Michael O. Rubinstein, `lcalc`, 2012. http://code.google.com/p/l-calc/

[21] W. A. Stein et al., *Purple SAGE*, 2011. http://purple.sagemath.org

[22] ———, *Sage Mathematics Software* (*version* 4.7.2), 2011. http://www.sagemath.org

[23] Andrew Sutherland, `smalljac`, 2012. http://www-math.mit.edu/∼drew/

[24] Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), no. 3, 553–572. MR 96d:11072

[25] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551. MR 96d:11071

JONATHAN W. BOBER: `jwbober@gmail.com`
Department of Mathematics, University of Washington, Seattle, WA 98195-4350, United States
Current address: *Howard House, University of Bristol, Queens Avenue, Bristol BS8 1SN United Kingdom*

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.145

msp

A database of elliptic curves over $\mathbb{Q}(\sqrt{5})$: a first report

Jonathan Bober, Alyson Deines, Ariah Klages-Mundt,
Benjamin LeVeque, R. Andrew Ohana,
Ashwath Rabindranath, Paul Sharaba, and William Stein

We describe a tabulation of (conjecturally) modular elliptic curves over the field $\mathbb{Q}(\sqrt{5})$ up to the first elliptic curve of rank 2. Using an efficient implementation of an algorithm of Lassina Dembélé, we computed tables of Hilbert modular forms of weight $(2, 2)$ over $\mathbb{Q}(\sqrt{5})$, and via a variety of methods we constructed corresponding elliptic curves, including (again, conjecturally) all elliptic curves over $\mathbb{Q}(\sqrt{5})$ that have conductor with norm less than or equal to 1831.

1. Introduction

1A. ***Elliptic curves over*** $\mathbb{Q}$**.** Tables of elliptic curves over $\mathbb{Q}$ have been of great value in mathematical research. Some of the first such tables were those in Antwerp IV [4], which included all elliptic curves over $\mathbb{Q}$ of conductor up to 200, and also a table of all elliptic curves with bad reduction only at 2 and 3.

Cremona's book [10] gives a detailed description of algorithms that together output a list of all elliptic curves over $\mathbb{Q}$ of any given conductor, along with extensive data about each curve. The proof that his algorithm outputs *all* curves of given conductor had to wait for the proof of the full modularity theorem in [8]. Cremona has subsequently computed tables [12] of all elliptic curves over $\mathbb{Q}$ of conductor up to 300,000, including Mordell-Weil groups and other extensive data about each curve.

In another direction, Stein and Watkins (see [33; 1]) created a table of 136,832,795 elliptic curves over $\mathbb{Q}$ of conductor $\leq 10^8$, and a table of 11,378,911 elliptic curves over $\mathbb{Q}$ of prime conductor $\leq 10^{10}$. There are many curves of large discriminant

MSC2010: primary 11-04; secondary 11G05.
Keywords: elliptic curves, totally real number fields, Hilbert modular forms, tables, sage.

missing from the Stein-Watkins tables, since these tables are made by enumerating curves with relatively small defining equations, and discarding those of large conductor, rather than systematically finding all curves of given conductor no matter how large the defining equation.

1B. ***Why*** $\mathbb{Q}(\sqrt{5})$***?*** Like $\mathbb{Q}$, the field $F = \mathbb{Q}(\sqrt{5})$ is a totally real field, and many of the theorems and ideas about elliptic curves over $\mathbb{Q}$ have been generalized to totally real fields. As is the case over $\mathbb{Q}$, there is a notion of modularity of elliptic curves over F, and work of Zhang [36] has extended many results of Gross and Zagier [20] and Kolyvagin [24] to the context of elliptic curves over totally real fields.

If we order totally real number fields K by the absolute value of their discriminant, then $F = \mathbb{Q}(\sqrt{5})$ comes next after $\mathbb{Q}$ (the Minkowski bound implies that $|D_K| \geq (n^n/n!)^2$, where $n = [K : \mathbb{Q}]$, so if $n \geq 3$ then $|D_K| > 20$). That 5 divides $\mathrm{disc}(F) = 5$ thwarts attempts to easily generalize the method of Taylor and Wiles to elliptic curves over F, which makes $\mathbb{Q}(\sqrt{5})$ even more interesting. Furthermore F is a PID and elliptic curves over F admit global minimal models and have well-defined notions of minimal discriminants. The field F also has 31 CM j-invariants, which is far more than any other quadratic field (see Section 5). Letting $\varphi = \frac{1+\sqrt{5}}{2}$, we have that the group of units $\{\pm 1\} \times \langle \varphi \rangle$ of the ring $R = \mathcal{O}_F = \mathbb{Z}[\varphi]$ of integers of F is infinite, leading to additional complications. Finally, F has even degree, which makes certain computations more difficult, as the cohomological techniques of [19] are not available.

1C. ***Modularity conjecture.*** The following conjecture is open:

Conjecture 1.1 (Modularity). *The set of L-functions of elliptic curves over F equals the set of L-functions associated to cuspidal Hilbert modular newforms over F of weight* $(2, 2)$ *with rational Hecke eigenvalues.*

Given the progress on modularity theorems initiated by [35], we are optimistic that Conjecture 1.1 will be proved. *We assume Conjecture* 1.1 *for the rest of this paper.*

In Section 2 we sketch how to compute Hilbert modular forms using arithmetic in quaternion algebras. Section 3 gives numerous methods for finding an elliptic curve corresponding to a Hilbert modular form. It should be noted that these are the methods *originally* used to make the tables – in hindsight, it was discovered that some of the elliptic curves found using the more specific techniques could be found using a better implementation of the sieved enumeration of Section 3B. Section 4 addresses how to find all curves that are isogenous to a given curve. In Section 5 we enumerate the CM j-invariants in F. We discuss some projects for future work in Section 6. Finally, Section 7 contains tables that summarize various information about our dataset [5].

2. Computing Hilbert modular forms over F

In Section 2A we sketch Dembélé's approach to computing Hilbert modular forms over F, then in Section 2B we make some remarks about our fast implementation.

2A. ***Hilbert modular forms and quaternion algebras.*** Dembélé [14] introduced an algebraic approach via the Jacquet-Langlands correspondence to computing Hilbert modular forms of weight $(2,2)$ over F. The Hamiltonian quaternion algebra $F[i,j,k]$ over F is ramified exactly at the two infinite places, and contains the maximal order

$$S = R\big[\tfrac{1}{2}(1-\bar{\varphi}i+\varphi j),\ \tfrac{1}{2}(-\bar{\varphi}i+j+\varphi k),\ \tfrac{1}{2}(\varphi i-\bar{\varphi}j+k),\ \tfrac{1}{2}(i+\varphi j-\bar{\varphi}k)\big].$$

For any nonzero ideal $\mathfrak{n}$ in $R=\mathcal{O}_F$, let $\mathbb{P}^1(R/\mathfrak{n})$ be the set of equivalence classes of column vectors with two coprime entries $a,b\in R/\mathfrak{n}$ modulo the action of $(R/\mathfrak{n})^*$. We use the notation $[a:b]$ to denote the equivalence class of $\left(\begin{smallmatrix}a\\b\end{smallmatrix}\right)$. For each prime $\mathfrak{p}\mid\mathfrak{n}$, we fix a choice of isomorphism $F[i,j,k]\otimes F_\mathfrak{p}\approx M_2(F_\mathfrak{p})$, which induces a left action of S^* on $\mathbb{P}^1(R/\mathfrak{n})$. The action of $T_\mathfrak{p}$, for $p\nmid\mathfrak{n}$, is $T_\mathfrak{p}([x])=\sum[\alpha x]$, where the sum is over the classes $[\alpha]\in S/S^*$ with $N_{\mathrm{red}}(\alpha)=\pi_\mathfrak{p}$ (reduced quaternion norm), where $\pi_\mathfrak{p}$ is a fixed choice of totally positive generator of $\mathfrak{p}$. The Jacquet-Langlands correspondence implies that the space of Hilbert modular forms of level $\mathfrak{n}$ and weight $(2,2)$ is noncanonically isomorphic as a module over the Hecke algebra

$$\mathbb{T}=\mathbb{Z}[\,T_\mathfrak{p}:\mathfrak{p}\text{ nonzero prime ideal of }R\,]$$

to the finite dimensional complex vector space $V=\mathbb{C}[S^*\backslash\mathbb{P}^1(R/\mathfrak{n})]$.

2B. ***Remarks on computing with*** $\mathbb{P}^1(R/\mathfrak{n})$**.** In order to implement the algorithm sketched in Section 2A, it is critical that we can compute with $\mathbb{P}^1(R/\mathfrak{n})$ very, very quickly. For example, to apply the method of Section 3G below, in some cases we have to compute tens of thousands of Hecke operators. Thus in this section we make some additional remarks about this fast implementation.

When $\mathfrak{n}=\mathfrak{p}^e$ is a prime power, it is straightforward to efficiently enumerate representative elements of $\mathbb{P}^1(R/\mathfrak{p}^e)$, since each element $[x:y]$ of $\mathbb{P}^1(R/\mathfrak{p}^e)$ has a unique representative of the form $[1:b]$ or $[a:1]$ with a divisible by $\mathfrak{p}$, and these are all distinct. It is easy to put any $[x:y]$ in this canonical form and enumerate the elements of $\mathbb{P}^1(R/\mathfrak{p}^e)$, after choosing a way to enumerate the elements of $R/\mathfrak{p}^e$. An enumeration of $R/\mathfrak{p}^e$ is easy to give once we decide on how to represent $R/\mathfrak{p}^e$.

In general, consider the factorization $\mathfrak{n}=\prod_{i=1}^m\mathfrak{p}_i^{e_i}$. We have a bijection between $\mathbb{P}^1(R/\mathfrak{n})$ and $\prod_{i=1}^m\mathbb{P}^1(R/\mathfrak{p}_i^{e_i})$, which allows us to reduce to the prime power case, at the expense of having to compute the bijection $R/\mathfrak{n}\cong\prod R/\mathfrak{p}_i^{e_i}$. To this end, we *represent elements* of $R/\mathfrak{n}$ as m-tuples in $\prod R/\mathfrak{p}_i^{e_i}$, thus making computation of the bijection trivial.

To minimize dynamic memory allocation, thus speeding up the code by an order of magnitude, in the implementation we make some arbitrary bounds; this is not a serious constraint, since the linear algebra needed to isolate eigenforms for levels beyond this bound is prohibitive. We assume $m \leq 16$ and each individual $p_i^{e_i} \leq 2^{31}$, where p_i is the residue characteristic of $\mathfrak{p}_i$. In all cases, we represent an element of $R/\mathfrak{p}_i^{e_i}$ as a pair of 64-bit integers, and represent an element of $R/\mathfrak{n}$ as an array of 16 pairs of 64-bit integers. We use this representation in all cases, even if $\mathfrak{n}$ is divisible by less than 16 primes; the gain in speed coming from avoiding dynamic memory allocation more than compensates for the wasted memory.

Let $\mathfrak{p}^e$ be one of the prime power factors of $\mathfrak{n}$, and let p be the residue characteristic of $\mathfrak{p}$. We have one of the following cases:

- $\mathfrak{p}$ splits in R; then $R/\mathfrak{p} \cong \mathbb{Z}/p\mathbb{Z}$ and we represent elements of $R/\mathfrak{p}^e$ as pairs $(a, 0) \bmod p^e$ with the usual addition and multiplication in the first factor.
- $\mathfrak{p}$ is inert in R; then $R/\mathfrak{p}^e \cong (\mathbb{Z}/p^e\mathbb{Z})[x]/(x^2 - x - 1)$, and we represent elements by pairs $(a, b) \in \mathbb{Z}/p^e\mathbb{Z}$ with multiplication
$$(a, b)(c, d) = (ac + bd, ad + bd + bc) \mod p^e.$$
- $\mathfrak{p}$ is ramified and $e = 2f$ is even; this is exactly the same as the case when $\mathfrak{p}$ is inert but with e replaced by f, since $R/\mathfrak{p}^e R \cong (\mathbb{Z}/p^f\mathbb{Z})[x]/(x^2 - x - 1)$.
- $\mathfrak{p}$ is ramified (so $p = 5$) and $e = 2f - 1$ is odd; the ring $A = R/\mathfrak{p}^e$ is trickier than the rest, because it is *not* of the form $\mathbb{Z}[x]/(m, g)$ where $m \in \mathbb{Z}$ and $g \in \mathbb{Z}[x]$. We have $A \approx (\mathbb{Z}/5^f\mathbb{Z})[x]/(x^2 - 5, 5^{f-1}x)$, and represent elements of A as pairs $(a, b) \in (\mathbb{Z}/5^f) \times (\mathbb{Z}/5^{f-1}\mathbb{Z})$, with arithmetic given by
$$(a, b) + (c, d) = (a + c \mod 5^f,\ b + d \mod 5^{f-1})$$
$$(a, b) \cdot (c, d) = (ac + 5bd \mod 5^f,\ ad + bc \mod 5^{f-1}).$$
We find that $\varphi \in R \mapsto (1/2, 1/2)$.

3. Strategies for finding an elliptic curve attached to a Hilbert modular form

In this section we describe various strategies to find an elliptic curve associated to each of the Hilbert modular forms computed in Section 2. Let f be a rational cuspidal Hilbert newform of weight $(2, 2)$ as in Section 2. According to Conjecture 1.1, there is some elliptic curve E_f over F such that $L(f, s) = L(E_f, s)$. (Note that E_f is only well defined up to isogeny.) Unlike the case for elliptic curves over $\mathbb{Q}$ (see [10]), there seems to be no known *efficient* direct algorithm to find E_f. Nonetheless, there are several approaches coming from various directions, which are each efficient in some cases.

Everywhere below, we continue to assume that Conjecture 1.1 is true and assume that we have computed (as in Section 2) the Hecke eigenvalues $a_{\mathfrak{p}} \in \mathbb{Z}$ of all rational Hilbert newforms of some level $\mathfrak{n}$, for $\mathrm{Norm}(\mathfrak{p}) \leq B$ a good prime, where B is large enough to distinguish newforms. In some cases we will need far more $a_{\mathfrak{p}}$ in order to compute with the L-function attached to a newform. We will also need the $a_{\mathfrak{p}}$ for bad $\mathfrak{p}$ in a few cases, which we obtain using the functional equation for the L-function (as an application of Dokchitser's algorithm [16]).

We define the *norm conductor* of an elliptic curve over F to be the absolute norm of the conductor ideal of the curve.

In Section 3A we give a very simple enumeration method for finding curves, then in Section 3B we refine it by taking into account point counts modulo primes; together, these two methods found a substantial fraction of our curves. Sections 3C and 3D describe methods for searching in certain families of curves, for example, curves with a torsion point of given order or curves with a given irreducible mod ℓ Galois representation. Section 3E is about how to find all twists of a curve with bounded norm conductor. In Section 3F we mention the Cremona-Lingham algorithm, which relies on computing all S-integral points on many auxiliary curves. Finally, Section 3G explains in detail an algorithm of Dembélé that uses explicit computations with special values of L-functions to find curves.

3A. ***Extremely naïve enumeration.*** The most naïve strategy is to systematically enumerate elliptic curves $E\colon y^2 = x^3 + ax + b$, with $a, b \in R$, and for each E, to compute $a_{\mathfrak{p}}(E)$ for $\mathfrak{p}$ not dividing $\mathrm{Disc}(E)$ by counting points on E reduced modulo $\mathfrak{p}$. If all the $a_{\mathfrak{p}}(E)$ match with those of the input newform f up to the bound B, we then compute the conductor $\mathfrak{n}_E$, and if it equals $\mathfrak{n}$, we conclude from the sufficient largeness of B that E is in the isogeny class of E_f.

Under our hypotheses, this approach provides a deterministic and terminating algorithm to find all E_f. However, it can be extremely slow when $\mathfrak{n}$ is small but the simplest curve in the isogeny class of E_f has large coefficients. For example, using this search method it would be infeasible to find the curve (1) computed by Fisher using the visibility of Ш[7].

3B. ***Sieved enumeration.*** A refinement to the approach discussed above uses the $a_{\mathfrak{p}}$ values to impose congruence conditions modulo $\mathfrak{p}$ on E. If f is a newform with Hecke eigenvalues $a_{\mathfrak{p}}$, then $\#\tilde{E}_f(R/\mathfrak{p}) = \boldsymbol{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$. Given $\mathfrak{p}$ not dividing the level $\mathfrak{n}$, we can find all elliptic curves modulo $\mathfrak{p}$ with the specified number of points, especially when $\boldsymbol{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$ has few prime factors. We impose these congruence conditions at multiple primes $\mathfrak{p}_i$, use the Chinese remainder theorem, and lift the resulting elliptic curves modulo $R/\prod \mathfrak{p}_i$ to nonsingular elliptic curves over R.

While this method, like the previous one, will eventually terminate, it too is very ineffective if every E in the class of isogenous elliptic curves corresponding to f has large coefficients. However in practice, by optimally choosing the number of primes $\mathfrak{p}_i$, a reasonably efficient implementation of this method can be obtained.

3C. ***Torsion families.*** We find elliptic curves of small conductor by specializing explicit parametrizations of families of elliptic curves over F having specified torsion subgroups. We use the parametrizations of [25].

Theorem 3.1 (Kamienny and Najman, [22]). *The following is a complete list of torsion structures for elliptic curves over F*:

$$\begin{aligned} &\mathbb{Z}/m\mathbb{Z}, && 1 \le m \le 10,\ m = 12,\\ &\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2m\mathbb{Z}, && 1 \le m \le 4,\\ &\mathbb{Z}/15\mathbb{Z}. \end{aligned}$$

Moreover, there is a unique elliptic curve over F with 15*-torsion.*

We use the following proposition to determine in which family to search.

Proposition 3.2. *Let ℓ be a prime and let E be an elliptic curve over F. Then $\ell \mid \#E'(F)_{\mathrm{tor}}$ for some elliptic curve E' in the isogeny class of E if and only if $\ell \mid \boldsymbol{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$ for all odd primes $\mathfrak{p}$ at which E has good reduction.*

Proof. If $\ell \mid \#E'(F)_{\mathrm{tor}}$, from the injectivity of the reduction map at good primes [23, Appendix], we have that $\ell \mid \#\tilde{E}'(\mathbb{F}_{\mathfrak{p}}) = \boldsymbol{N}(\mathfrak{p}) + 1 - a_{\mathfrak{p}}$. The converse statement is one of the main results of [23]. □

By applying Proposition 3.2 for all $a_{\mathfrak{p}}$ with $\mathfrak{p}$ up to some bound, we can decide whether or not it is *likely* that some elliptic curve in the isogeny class of E contains an F-rational ℓ-torsion point. If this is the case, then we search over those families of elliptic curves with rational ℓ-torsion. With a relatively small search space, we thus find many elliptic curves with large coefficients more quickly than with the algorithm of Section 3A. For example, we first found the elliptic curve E given by

$$y^2 + \varphi y = x^3 + (27\varphi - 43)\, x + (-80\varphi + 128)$$

with norm conductor 145 by searching for elliptic curves with torsion subgroup $\mathbb{Z}/7\mathbb{Z}$.

3D. ***Congruence families.*** Suppose that we are searching for an elliptic curve E and we already know another elliptic curve E' with $E[\ell] \approx E'[\ell]$, where ℓ is some prime and $E[\ell]$ is irreducible. Twists of the modular curve $X(\ell)$ parametrize pairs of elliptic curves with isomorphic ℓ-torsion subgroups, so finding rational points on the correct twist allows us to find curves with the same mod ℓ Galois representation

as E'. Using this idea, we found the curve E given by

$$y^2 + \varphi xy = x^3 + (\varphi - 1)\,x^2 + (-257364\varphi - 159063)\,x + (-75257037\varphi - 46511406) \quad (1)$$

with conductor $-6\varphi + 42$, which has norm 1476. Just given the $a_{\mathfrak{p}}$, we noticed that $E[7] \approx E'[7]$, where E' has norm conductor 369. The curve E' had already been found via naïve search, since it is given by the equation $y^2 + (\varphi + 1)\,y = x^3 + (\varphi - 1)\,x^2 + (-2\varphi)\,x$. For any elliptic curve, the equation for the correct twist of $X(7)$ was found both by Halberstadt and Kraus [21] and by Fisher [18], whose methods also yield formulas for the appropriate twists of $X(9)$ and $X(11)$.

Fisher had already implemented Magma [6] routines to find ℓ-congruent elliptic curves over $\mathbb{Q}$ using these equations and was able to modify his work for $\mathbb{Q}(\sqrt{5})$. Fortunately, our curve E was then easily found.

3E. ***Twisting.*** Let E be an elliptic curve over F. A *twist* E' of E is an elliptic curve over F that is isomorphic to E over some extension of F. A *quadratic twist* is a twist in which the extension has degree 2. We can use twisting to find elliptic curves that may otherwise be difficult to find as follows: Starting with a known elliptic curve E of some (small) conductor, we compute its twists of conductor up to some bound, and add them to our table.

More explicitly, if E is given by $y^2 = x^3 + ax + b$ and $d \in F^*$, then the twist E^d of E by d is given by $dy^2 = x^3 + ax + b$; in particular, we may assume that d is squarefree. The following is well known:

Proposition 3.3. *If* $\mathfrak{n}$ *is the conductor of* E *and* $d \in \mathcal{O}_F$ *is nonzero, squarefree and coprime to* $\mathfrak{n}$*, then the conductor of* E^d *is divisible by* $d^2\mathfrak{n}$.

Proof. There are choices of Weierstrass equations such that $\Delta(E^d) = 2^{12}d^6\Delta(E)$, where Δ is the discriminant. Thus the elliptic curve E^d has bad reduction at each prime that divides d, because twisting introduces a 6th power of the squarefree d into the discriminant, and d is coprime to $\Delta(E)$, so no change of Weierstrass equation can remove this 6th power. Moreover, E^d is isomorphic to E over an extension of the base field, so E^d has potentially good reduction at each prime dividing d. Thus the reduction at each prime dividing d is additive. The conductor is unchanged at the primes dividing $\mathfrak{n}$ because of the formula relating the conductor, discriminant and reduction type (see [31, App. C,§15]), that formation of Néron models commutes with unramified base change, and the fact that at the primes that divide $\mathfrak{n}$ the minimal discriminant of E^d is the same as that of E. □

To find all twists E^d with norm conductor at most B, we twist E by all d of the form $\pm\varphi^\delta d_0 d_1$, where $\delta \in \{0, 1\}$, d_0 is a product of a fixed choice of generators for the prime divisors of $\mathfrak{n}$, d_1 is a squarefree product of a fixed choice of generators of

primes not dividing $\mathfrak{n}$, and $|N(d_1)| \leq \sqrt{B/C}$, where C is the norm of the product of the primes that exactly divide $\mathfrak{n}$. We know from 3.3 that this search is exhaustive.

For example, let E be given by $y^2 + xy + \varphi y = x^3 + (-\varphi - 1)x^2$ of conductor $5\varphi - 3$ having norm 31. Following the above strategy to find twists of norm conductor $\leq B := 1831$, we have $C = 31$ and squarefree d_1 such that $|N(d_1)| \leq \sqrt{B/C} \approx 7.6\ldots$. Thus $d_1 \in \{1, 2, \varphi, 2\varphi\}$ and checking all possibilities for $\varphi^\delta d_0 d_1$, we find the elliptic curve $E^{-\varphi-2}$ having norm conductor 775 and the elliptic curve $E^{5\varphi-3}$ having norm conductor 961. Other twists have larger norm conductors; for example, E^2 has norm conductor $126976 = 2^{12} \cdot 31$.

3F. ***Elliptic Curves with good reduction outside S.*** We use the algorithm of Cremona and Lingham from [11] to find all elliptic curves E having good reduction at primes outside of a finite set $\mathcal{S}$ of primes in F. This algorithm has limitations over a general number field K due to the difficulty of finding a generating set for $E(K)$ and points on E defined over $\mathcal{O}_K$. Using Cremona's Magma implementation of the algorithm, we found several elliptic curves not found by other methods, for example, $y^2 + (\varphi + 1)xy + y = x^3 - x^2 + (-19\varphi - 39)x + (-143\varphi - 4)$, which has norm conductor 1331.

3G. ***Special values of twisted L-series.*** In [15], Lassina Dembélé outlines some methods for finding modular elliptic curves from Hilbert modular forms over real quadratic fields. Formally, these methods are not proven to be any better than a direct search procedure, as they involve making a large number of guesses, and a priori we do not know just how many guesses we will need to make. And unlike other methods described in this paper, this method requires many Hecke eigenvalues, and computing these takes a lot of time. However, this method certainly works extremely well in many cases, and after tuning it by using large tables of elliptic curves that we had already computed, we are able to use it to find more elliptic curves that we would have had no hope of finding otherwise; we will give an example of one of these elliptic curves later.

When the level $\mathfrak{n}$ is not square, Dembélé's method relies on computing or guessing periods of the elliptic curve by using special values of L-functions of twists of the elliptic curve. In particular, the only inputs required are the level of the Hilbert modular form and its L-series. So we suppose that we know the level $\mathfrak{n} = (N)$ of the form, where N is totally positive, and that we have sufficiently many coefficients of its L-series $a_{\mathfrak{p}_1}, a_{\mathfrak{p}_2}, a_{\mathfrak{p}_3}, \ldots$.

Let σ_1 and σ_2 denote the embeddings of F into the real numbers, with $\sigma_1(\varphi) \approx 1.61803\ldots$. For an elliptic curve E over F we get two associated embeddings into the complex numbers, and hence a pair of period lattices. Let Ω_E^+ denote the smallest positive real period corresponding to the embedding σ_1, and similarly define Ω_E^- to be the smallest period which lies on the positive imaginary axis. We

will refer to these as the periods of E, and as the period lattices are interchanged when E is replaced with its conjugate elliptic curve, we let $\Omega_{\overline{E}}^{+}$ and $\Omega_{\overline{E}}^{-}$ denote the least real and imaginary periods of the lattice under the embedding σ_2.

For ease, we write

$$\Omega_E^{++} = \Omega_E^{+}\Omega_{\overline{E}}^{+} \qquad \Omega_E^{+-} = \Omega_E^{+}\Omega_{\overline{E}}^{-}$$
$$\Omega_E^{-+} = \Omega_E^{-}\Omega_{\overline{E}}^{+} \qquad \Omega_E^{--} = \Omega_E^{-}\Omega_{\overline{E}}^{-}.$$

We refer to these numbers as the *mixed periods* of E.

3G.1. *Recovering the elliptic curve from its mixed periods.* If we know these mixed periods to sufficient precision, it is not hard to recover the elliptic curve E. Without the knowledge of the discriminant of the elliptic curve, we do not know the lattice type of the elliptic curve and its conjugate, but there are only a few possibilities for what they might be. This gives us a few possibilities for the j-invariant of E. Observe that $\sigma_1(j(E))$ is either $j(\tau_1(E))$ or $j(\tau_2(E))$ and $\sigma_2(j(E))$ is either $j(\tau_1(\overline{E}))$ or $j(\tau_2(\overline{E}))$, where

$$\tau_1(E) = \frac{\Omega_E^{-+}}{\Omega_E^{++}} = \frac{\Omega_E^{-}}{\Omega_E^{+}} \qquad \tau_2(E) = \frac{1}{2}\left(1 + \frac{\Omega_E^{-+}}{\Omega_E^{++}}\right) = \frac{1}{2}\left(1 + \frac{\Omega_E^{-}}{\Omega_E^{+}}\right)$$
$$\tau_1(\overline{E}) = \frac{\Omega_E^{+-}}{\Omega_E^{++}} = \frac{\Omega_E^{-}}{\Omega_E^{+}} \qquad \tau_2(\overline{E}) = \frac{1}{2}\left(1 + \frac{\Omega_E^{+-}}{\Omega_E^{++}}\right) = \frac{1}{2}\left(1 + \frac{\Omega_{\overline{E}}^{-}}{\Omega_{\overline{E}}^{+}}\right)$$

and $j(\tau)$ is the familiar

$$j(\tau) = e^{-2\pi i\tau} + 744 + 196884e^{2\pi i\tau} + 21493760e^{4\pi i\tau} + \cdots.$$

We try each pair of possible embeddings for $j(E)$ in turn, and recognize possibilities for $j(E)$ as an algebraic number. We then construct elliptic curves E' corresponding to each possibility for $j(E)$. By computing a few $a_{\mathfrak{p}}(E)$, we should be able to determine whether we have chosen the correct j-invariant, in which case E' will be a twist of E. We can then recognize which twist it is in order to recover E.

In practice, of course, as we have limited precision, and as $j(E)$ will not be an algebraic integer, it may not be feasible to directly determine its exact value, especially if its denominator is large.

To get around the problem of limited precision, we suppose that we have some extra information; namely, the discriminant Δ_E of the elliptic curve we are looking for. With Δ_E in hand we can directly determine which τ to choose: If $\sigma_1(\Delta_E) > 0$ then $\sigma_1(j(E)) = j(\tau_1(E))$, and if $\sigma_1(\Delta_E) < 0$ then $\sigma_1(j(E)) = j(\tau_2(E))$, and similarly for σ_2. We then compute $\sigma_1(c_4(E)) = (j(\tau)\sigma_1(\Delta_E))^{1/3}$ and $\sigma_2(c_4(E)) = (j(\tau')\sigma_2(\Delta_E))^{1/3}$.

Using the approximations of the two embeddings of c_4, we can recognize c_4 approximately as an algebraic integer. Specifically, we compute

$$\alpha = \frac{\sigma_1(c_4) + \sigma_2(c_4)}{2} \quad \text{and} \quad \beta = \frac{\sigma_1(c_4) - \sigma_2(c_4)}{2\sqrt{5}}.$$

Then $c_4 = \alpha + \beta\sqrt{5}$, and we can find c_6.

In practice, there are two important difficulties we must overcome: We do not know Δ_E and it may be quite difficult to get high precision approximations to the mixed periods, and thus we may not be able to easily compute c_4. Thus, we actually proceed by choosing a Δ_{guess} from which we compute half-integers α and β and an integer $a + b\varphi \approx \alpha + \beta\sqrt{5}$, arbitrarily rounding either a or b if necessary. We then make some choice of search range M, and for each pair of integers m and n, bounded in absolute value by M, we try each $c_{4,\text{guess}} = (a+m) + (b+n)\varphi$.

Given $c_{4,\text{guess}}$, we attempt to solve

$$c_{6,\text{guess}} = \pm\sqrt{c_{4,\text{guess}}^3 - 1728\Delta_{\text{guess}}},$$

and, if we can, we use these to construct a elliptic curve E_{guess}. If E_{guess} has the correct conductor and the correct Hecke eigenvalues, we declare that we have found the correct elliptic curve; otherwise, we proceed to the next guess.

For a choice of Δ_{guess}, we will generally start with the conductor N_E, and then continue by trying unit multiples and by adding in powers of factors of N_E.

3G.2. *Guessing the mixed periods.* We have thus far ignored the issue of actually finding the mixed periods of the elliptic curve that we are looking for. Finding them presents an extra difficulty as our procedure involves even more guesswork. Dembélé's idea is to use special values of twists of the L-function $L(f, s)$. Specifically, we twist by primitive quadratic Dirichlet characters over $\mathcal{O}_F$, which are homomorphisms $\chi\colon (\mathcal{O}_F/\mathfrak{c})^* \to \pm 1$, pulled back to $\mathcal{O}_F$.

In the case of odd prime conductor, which we will stick to here, there is just a single primitive quadratic character, which is the quadratic residue symbol. A simple way to compute it is by making a table of squares, or by choosing a primitive root of $g \in (\mathcal{O}_F/\mathfrak{c})^*$, assigning $\chi(g) = -1$, and again making a table by extending multiplicatively. Alternatively, one could use a reciprocity formula as described in [7]. For general conductor, one can compute with products of characters having prime conductor.

For a given f and a primitive χ, we can construct the twisted L-function

$$L(f, \chi, s) = \sum_{\mathfrak{m} \subseteq \mathcal{O}_F} \frac{\chi(m) a_{\mathfrak{m}}}{N(\mathfrak{m})^s},$$

where m is a totally positive generator of $\mathfrak{m}$. (Note that χ is not well defined

on ideals, but *is* well defined on totally positive generators of ideals.) $L(f, \chi, s)$ will satisfy a functional equation similar to that of $L(f, s)$, but the conductor is multiplied by $\mathrm{Norm}(\mathfrak{c})^2$ and the sign is multiplied by $\chi(-N)$.

Oda [28] conjectured relations between the periods of f and the associated elliptic curve E and gave some relations between the periods of f and central values of $L(s, \chi, 1)$. Stronger versions of these relations are conjectured, and they are what Dembélé uses to obtain information about the mixed periods of E. Specifically, Dembélé distills the following conjecture from [2], which we further simplify to state specifically for $\mathbb{Q}(\sqrt{5})$.

Conjecture 3.4. *If χ is a primitive quadratic character with conductor $\mathfrak{c}$ relatively prime to the conductor of E, with $\chi(\varphi) = s'$ and $\chi(1-\varphi) = s$, (where $s, s' \in \{+, -\} = \{\pm 1\}$), then*

$$\Omega_E^{s,s'} = c_\chi \tau(\overline{\chi}) L(E, \chi, 1)\sqrt{5},$$

for some integer c_χ, where $\tau(\chi)$ is the Gauss sum

$$\tau(\chi) = \sum_{\alpha \bmod \mathfrak{c}} \chi(\alpha) \exp\left(2\pi i \,\mathrm{Tr}\left(\alpha/m\sqrt{5}\right)\right),$$

with m a totally positive generator of $\mathfrak{c}$.

Remark. The Gauss sum is more innocuous than it seems. For odd conductor $\mathfrak{c}$ it is of size $\sqrt{\mathrm{Norm}(\mathfrak{c})}$, while for an even conductor it is of size $\sqrt{2\,\mathrm{Norm}(\mathfrak{c})}$. Its sign is a 4-th root of unity, and whether it is real or imaginary can be deduced directly from the conjecture, as it matches with the sign of $\Omega_E^{s,s'}$. In particular, $\tau(\chi)$ is real when $\chi(-1) = 1$ and imaginary when $\chi(-1) = -1$, which is a condition on $\mathrm{Norm}(\mathfrak{c}) \bmod 4$, as $\chi(-1) \equiv \mathrm{Norm}(\mathfrak{c}) \pmod 4$. This can all be deduced, for example, from [7].

Also, note that Dembélé writes this conjecture with an additional factor of $4\pi^2$; this factor does not occur with the definition of $L(f, s)$ that we have given.

Remark. Contained in this conjecture is the obstruction to carrying out the method described here when $\mathfrak{n}$ is a square. If the sign of the functional equation of $L(f, s)$ is ϵ_f, then the sign of $L(f, \chi, s)$ will be $\chi(-N)\epsilon_f$. When $\mathfrak{n}$ is a perfect square, this is completely determined by whether or not $\chi(\varphi) = \chi(1-\varphi)$, so we can only obtain information about either Ω^{--} and Ω^{++} or Ω^{-+} and Ω^{+-}, and we need three of these values to find E.

With this conjecture in place, we can describe a method for guessing the mixed periods of E. Now, to proceed, we construct four lists of characters up to some conductor bound M (we are restricting to odd prime modulus here for simplicity,

as primitivity is ensured, but this is not necessary):

$$S^{s,s'} = \{\chi \bmod \mathfrak{p} : \chi(\varphi) = s',\ \chi(1-\varphi) = s,\ (\mathfrak{p}, \mathfrak{n}) = 1, \mathrm{Norm}(\mathfrak{p}) < M,\ \chi(-N) = \epsilon_f\}.$$

Here $s, s' \in \{+,-\} = \{\pm 1\}$ again, and we restrict our choice of characters to force the functional equation of $L(s, \chi, f)$ to have positive sign so that there is a good chance that it does not vanish as the central point. We will consider these lists to be ordered by the norms of the conductors of the characters in increasing order, and index their elements as $\chi_0^{s,s'}, \chi_1^{s,s'}, \chi_2^{s,s'}, \ldots$. For each character we compute the central value of the twisted L-function to get four new lists

$$\mathcal{L}^{s,s'} = \{i^{ss'}\sqrt{5\,\mathrm{Norm}(\mathfrak{p})}L(E, \chi, 1), \chi \in S^{s,s'}\} = \{\mathcal{L}_0^{s,s'}, \mathcal{L}_1^{s,s'}, \ldots\}.$$

These numbers should now all be integer multiples of the mixed periods, so to get an idea of which integer multiples they might be, we compute each of the ratios

$$\frac{\mathcal{L}_0^{s,s'}}{\mathcal{L}_k^{s,s'}} = \frac{c_{\chi_0^{s,s'}}}{c_{\chi_k^{s,s'}}} \in \mathbb{Q}, \quad k = 1, 2, \ldots,$$

attempt to recognize these as rational numbers, and choose as an initial guess

$$\Omega_{E,\mathrm{guess}}^{ss'} = \mathcal{L}_0^{s,s'}\left(\mathrm{lcm}\left\{\mathrm{numerator}\left(\frac{\mathcal{L}_0^{s,s'}}{\mathcal{L}_k^{s,s'}}\right) : k = 1, 2, \ldots\right\}\right)^{-1}.$$

3G.3. *An example.* We give an example of an elliptic curve that we were only able to find by using this method. At level $\mathfrak{n} = (-38\varphi + 26)$ we found a newform f, computed

$$\begin{aligned} a_{(2)}(f) &= -1, \qquad a_{(-2\varphi+1)}(f) = 1, \\ a_{(3)}(f) &= -1, \qquad a_{(-3\varphi+1)}(f) = -1, \quad a_{(-3\varphi+2)}(f) = -6, \\ &\ldots, \\ a_{(200\varphi-101)}(f) &= 168, \end{aligned}$$

and determined, by examining the L-function, that the sign of the functional equation should be -1. (In fact, we do not really need to know the sign of the functional equation, as we would quickly determine that $+1$ is wrong when attempting to find the mixed periods.) Computing the sets of characters described above, and choosing the first 3 of each, we have

$$\begin{aligned} S^{--} &= \{\chi_{(\varphi+6)}, \chi_{(7)}, \chi_{(7\varphi-4)}\}, & S^{-+} &= \{\chi_{(-3\varphi+1)}, \chi_{(5\varphi-2)}, \chi_{(\varphi-9)}\} \\ S^{+-} &= \{\chi_{(-4\varphi+3)}, \chi_{(5\varphi-3)}, \chi_{(-2\varphi+13)}\} & S^{++} &= \{\chi_{(\varphi+9)}, \chi_{(9\varphi-5)}, \chi_{(\varphi+13)}\}. \end{aligned}$$

By using the 5133 eigenvalues above as input to Rubinstein's `lcalc` [29], we compute the lists of approximate values

$$\begin{aligned}
\mathcal{L}^{--} &= \{-33.5784397862407,\ -3.73093775400387,\ -18.6546887691646\ \},\\
\mathcal{L}^{-+} &= \{\ 18.2648617736017i,\ 32.8767511924831i,\ 3.65297235421633i\},\\
\mathcal{L}^{+-} &= \{\ 41.4805656925342i,\ 8.29611313850694i,\ 41.4805677827298i\ \},\\
\mathcal{L}^{++} &= \{\ 32.4909970742969,\ 162.454985515474,\ 162.454973589303\ \}.
\end{aligned}$$

Note that `lcalc` will warn us that we do not have enough coefficients to obtain good accuracy, and we make no claim as far as the accuracy of these values is concerned. Hoping that the ends will justify the means, we proceed forward.

Dividing each list by the first entry, and recognizing the quotients as rational numbers, we get the lists

$$\begin{aligned}
&\{1.000,\ 9.00000000005519,\ \ 1.80000000009351\ \} \approx \{1, 9, 9/5\},\\
&\{1.000,\ 0.555555555555555,\ 5.00000000068986\ \} \approx \{1, 5/9, 5\},\\
&\{1.000,\ 4.99999999999994,\ \ 0.999999949610245\} \approx \{1, 5, 1\},\\
&\{1.000,\ 0.199999999822733,\ 0.200000014505165\} \approx \{1, 1/5, 1/5\},
\end{aligned}$$

which may give an indication of the accuracy of our values. We now proceed with the guesses

$$\begin{aligned}
\Omega^{--}_{E,\mathrm{guess}} &\approx -33.5784397862407/9 \approx -3.73093775402141,\\
\Omega^{-+}_{E,\mathrm{guess}} &\approx \ \ 18.2648617736017i/5 \approx \ \ 3.65297235472034i,\\
\Omega^{+-}_{E,\mathrm{guess}} &\approx \ \ 41.4805656925342i/5 \approx \ \ 8.29611313850683i,\\
\Omega^{++}_{E,\mathrm{guess}} &\approx \ \ 32.4909970742969 \ \ = \ 32.4909970742969.
\end{aligned}$$

These cannot possibly be all correct, as $\Omega^{--}_E\Omega^{++}_E = \Omega^{-+}_E\Omega^{+-}_E$. Still, we can choose any three and get a reasonable guess, and in fact we may choose all possible triples, dividing some of the guesses by small rational numbers, and choosing the fourth guess to be consistent with the first three; we build a list of possible embeddings of $j(E)$, which will contain the possibility $\sigma_1(j(E)) \approx 1.365554233954 \times 10^{12}$, $\sigma_2(j(E)) \approx 221270.95861123$, which is a possibility if

$$\Omega^{-+}_E = \Omega^{-+}_{E,\mathrm{guess}},\quad \Omega^{+-}_E = \Omega^{+-}_{E,\mathrm{guess}},\quad \Omega^{-+}_E = \frac{\Omega^{-+}_{E,\mathrm{guess}}}{2},\quad \Omega^{++}_E = \frac{\Omega^{++}_{E,\mathrm{guess}}}{8}.$$

Cycling through many discriminants, we eventually try

$$\Delta_{\mathrm{guess}} = \varphi \cdot 2^5 \cdot (19\varphi - 13),$$

which leads us to the guess

$$\sigma_1(c_{4,\text{guess}}) = (\sigma_1(j(E))\sigma_1(\Delta_{\text{guess}}))^{1/3} \approx 107850.372979378,$$
$$\sigma_2(c_{4,\text{guess}}) = (\sigma_2(j(E))\sigma_2(\Delta_{\text{guess}}))^{1/3} \approx 476.625892034286.$$

We have enough precision to easily recognize this as

$$c_{4,\text{guess}} = \frac{108327 + 48019\sqrt{5}}{2} = 48019\varphi + 30154,$$

and

$$\sqrt{c_{4,\text{guess}}^3 - 1728\Delta_{\text{guess}}}$$

does in fact have two square roots: $\pm(15835084\varphi + 9796985)$. We try both of them, and the choice with the minus sign gives the elliptic curve

$$y^2 + \varphi xy + \varphi y = x^3 + (\varphi - 1)\,x^2 + (-1001\varphi - 628)\,x + (17899\varphi + 11079)\,,$$

which has the correct conductor. We compute a few values of $a_{\mathfrak{p}}$ for this elliptic curve, and it turns out to be the one that we are looking for.

4. Enumerating the elliptic curves in an isogeny class

Given an elliptic curve E/F, we wish to find representatives up to isomorphism for all elliptic curves E'/F that are isogenous to E via an isogeny defined over F. The analogue of this problem over $\mathbb{Q}$ has an algorithmic solution as explained in [10, §3.8]; it relies on:

(1) Mazur's theorem [27] that if $\psi\colon E \to E'$ is a $\mathbb{Q}$-rational isogeny of prime degree, then $\deg(\psi) \leq 163$.

(2) Formulas of Vélu [34] that provide a way to explicitly enumerate all p-isogenies (if any) with domain E. Vélu's formulas are valid for any number field, but so far there has not been an explicit generalization of Mazur's theorem for any number field other than $\mathbb{Q}$.

Remark. Assume the generalized Riemann hypothesis. Then work of Larson and Vaintrob from [26] implies that there is an effectively computable constant C_F such that if $\varphi\colon E \to E'$ is a prime-degree isogeny defined over F and E' and E are not isomorphic over F, then φ has degree at most C_F.

Since we are interested in specific isogeny classes, we can use the algorithm described in [3] that takes as input a specific non-CM elliptic curve E over a number field K, and outputs a provably finite list of primes p such that E might have a p-isogeny. The algorithm is particularly easy to implement in the case when K is a quadratic field, as explained in [3, §2.3.4]. Using this algorithm combined with

Vélu's formulas, we were able to enumerate *all* isomorphism classes of elliptic curves isogenous to the elliptic curves we found via the methods of Section 3, and thus divide our isogeny classes into isomorphism classes.

5. CM elliptic curves over F

In this section we make some general remarks about CM elliptic curves over F. The main surprise is that there are 31 distinct $\overline{\mathbb{Q}}$-isomorphism classes of CM elliptic curves defined over F, more than for any other quadratic field.

Proposition 5.1. *The field F has more isomorphism classes of CM elliptic curves than any other quadratic field.*

Proof. Let K be a quadratic extension of $\mathbb{Q}$. Let H_D denote the Hilbert class polynomial of the CM order $\mathcal{O}_D$ of discriminant D, so $H_D \in \mathbb{Q}[X]$ is the minimal polynomial of the j-invariant j_D of any elliptic curve $E = E_D$ with CM by $\mathcal{O}_D$. Since K is Galois, we have $j_D \in K$ if and only if H_D is either linear or quadratic with both roots in K. The D for which H_D is linear are the thirteen values $-3, -4, -7, -8, -11, -12, -16, -19, -27, -28, -43, -67, -163$. According to [9], the D for which H_D is quadratic are the following 29 discriminants:

$$-15, -20, -24, -32, -35, -36, -40, -48, -51, -52, -60,$$
$$-64, -72, -75, -88, -91, -99, -100, -112, -115, -123,$$
$$-147, -148, -187, -232, -235, -267, -403, -427.$$

By computing discriminants of these Hilbert class polynomials, we obtain Table 1. The claim follows because the $\mathbb{Q}(\sqrt{5})$ row is largest, containing 9 entries. There are thus $31 = 2 \cdot 9 + 13$ distinct CM j-invariants in $\mathbb{Q}(\sqrt{5})$. □

6. Related future projects

It would be natural to extend the tables to the first known elliptic curve of rank 3 over F, which may be the elliptic curve $y^2 + y = x^3 - 2x + 1$ of norm conductor $163^2 = 26569$. It would also be interesting to make a table in the style of [33], and compute analytic ranks of the large number of elliptic curves that we would find; this would benefit from Sutherland's `smalljac` program, which has very fast code for computing L-series coefficients. Some aspects of the tables could also be generalized to modular abelian varieties A_f attached to Hilbert modular newforms with not necessarily rational Hecke eigenvalues; in particular, we could enumerate the A_f up to some norm conductor, and numerically compute their analytic ranks.

K	D
$\mathbb{Q}(\sqrt{2})$	$-24, -32, -64, -88$
$\mathbb{Q}(\sqrt{3})$	$-36, -48$
$\mathbb{Q}(\sqrt{5})$	$-15, -20, -35, -40, -60, -75, -100, -115, -235$
$\mathbb{Q}(\sqrt{6})$	-72
$\mathbb{Q}(\sqrt{7})$	-112
$\mathbb{Q}(\sqrt{13})$	$-52, -91, -403$
$\mathbb{Q}(\sqrt{17})$	$-51, -187$
$\mathbb{Q}(\sqrt{21})$	-147
$\mathbb{Q}(\sqrt{29})$	-232
$\mathbb{Q}(\sqrt{33})$	-99
$\mathbb{Q}(\sqrt{37})$	-148
$\mathbb{Q}(\sqrt{41})$	-123
$\mathbb{Q}(\sqrt{61})$	-427
$\mathbb{Q}(\sqrt{89})$	-267

Table 1. Quadratic fields K and the values of D for which H_D has roots in K but not in $\mathbb{Q}$.

7. Tables

As explained in Sections 3 and 4, assuming Conjecture 1.1, we found the complete list of elliptic curves with norm conductor up to 1831, which is the first norm conductor of a rank 2 elliptic curve over F. The complete dataset can be downloaded from [5].

In each of the following tables #isom refers to the number of isomorphism classes of elliptic curves, #isog refers to the number of isogeny classes of elliptic curves, $\mathfrak{n}$ refers to the conductor of the given elliptic curve, $\boldsymbol{N}(\mathfrak{n})$ is the norm of the conductor, and Weierstrass equations are given in the form $[\![a_1, a_2, a_3, a_4, a_6]\!]$.

Table 2 gives the number of elliptic curves and isogeny classes we found. Note that in these counts we do not exclude conjugate elliptic curves, that is, if σ denotes

Rank	#Isog	#Isom	Smallest $\boldsymbol{N}(\mathfrak{n})$
0	745	2174	31
1	667	1192	199
2	2	2	1831
Total	1414	3368	—

Table 2. Number of isogeny classes and number of isomorphism classes of elliptic curves over F of norm conductor at most 1831.

	Size of isogeny class							
Bound on $N(\mathfrak{n})$	1	2	3	4	6	8	10	Total
199	2	21	3	20	8	9	1	64
1831	498	530	36	243	66	38	3	1414

Table 3. Number of isogeny classes of a given size for elliptic curves over F with norm conductors no larger than a given bound.

the nontrivial element of $\mathrm{Gal}(F/\mathbb{Q})$, then we count E and E^σ separately if they are not isomorphic.

Table 3 gives counts of the number of isogeny classes of elliptic curves in our data of each size; note that we find some isogeny classes of cardinality 10, which is bigger than what one observes with elliptic curves over $\mathbb{Q}$.

Table 4 gives the number of elliptic curves and isogeny classes up to a given norm conductor bound. Note that the first elliptic curve of rank 1 has norm conductor 199, and there are no elliptic curves of norm conductor 200.

	#Isogeny classes				#Isomorphism classes			
	Rank				Rank			
Bound on $N(\mathfrak{n})$	0	1	2	Total	0	1	2	Total
200	62	2	0	64	257	6	0	263
400	151	32	0	183	580	59	0	639
600	246	94	0	340	827	155	0	982
800	334	172	0	506	1085	285	0	1370
1000	395	237	0	632	1247	399	0	1646
1200	492	321	0	813	1484	551	0	2035
1400	574	411	0	985	1731	723	0	2454
1600	669	531	0	1200	1970	972	0	2942
1800	729	655	0	1384	2128	1178	0	3306
1831	745	667	2	1414	2174	1192	2	3368

Table 4. Number of isogeny classes and number of isomorphism classes of elliptic curves over F with specified rank and with norm conductors no larger than a given bound.

Table 5 gives the number of elliptic curves and isogeny classes with isogenies of each degree; note that we do not see all possible isogeny degrees. For example, the elliptic curve $X_0(19)$ has rank 1 over F, so there are infinitely many elliptic curves over F with degree 19 isogenies (unlike over $\mathbb{Q}$ where $X_0(19)$ has rank 0). We

Type	#Isog	#Isom	Example curve	$N(\mathfrak{n})$
none	498	498	$[\![\varphi+1, 1, 1, 0, 0]\!]$	991
deg 2	652	2298	$[\![\varphi, -\varphi+1, 0, -4, 3\varphi-5]\!]$	99
deg 3	289	950	$[\![\varphi, -\varphi, \varphi, -2\varphi-2, 2\varphi+1]\!]$	1004
deg 5	65	158	$[\![1, 0, 0, -28, 272]\!]$	900
deg 7	19	38	$[\![0, \varphi+1, \varphi+1, \varphi-1, -3\varphi-3]\!]$	1025

Table 5. Number of isogeny classes and number of isomorphism classes of elliptic curves over F of norm conductor at most 1831 having isogenies of a given type. "None" indicates curves having no cyclic isogenies.

also give an example of an elliptic curve (that need not have minimal conductor) with an isogeny of the given degree.

Table 6 gives the number of elliptic curves with each torsion structure, along with an example of an elliptic curve (again, not necessarily with minimal conductor) with that torsion structure.

Group structure	#Isom	Example curve	$N(\mathfrak{n})$
0	796	$[\![0, -1, 1, -8, -7]\!]$	225
$\mathbb{Z}/2\mathbb{Z}$	1453	$[\![\varphi, -1, 0, -\varphi-1, \varphi-3]\!]$	164
$\mathbb{Z}/3\mathbb{Z}$	202	$[\![1, 0, 1, -1, -2]\!]$	100
$\mathbb{Z}/4\mathbb{Z}$	243	$[\![\varphi+1, \varphi-1, \varphi, 0, 0]\!]$	79
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$	312	$[\![0, \varphi+1, 0, \varphi, 0]\!]$	256
$\mathbb{Z}/5\mathbb{Z}$	56	$[\![1, 1, 1, 22, -9]\!]$	100
$\mathbb{Z}/6\mathbb{Z}$	183	$[\![1, \varphi, 1, \varphi-1, 0]\!]$	55
$\mathbb{Z}/7\mathbb{Z}$	13	$[\![0, \varphi-1, \varphi+1, 0, -\varphi]\!]$	41
$\mathbb{Z}/8\mathbb{Z}$	21	$[\![1, \varphi+1, \varphi, \varphi, 0]\!]$	31
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$	51	$[\![\varphi+1, 0, 0, -4, -3\varphi-2]\!]$	99
$\mathbb{Z}/9\mathbb{Z}$	6	$[\![\varphi, -\varphi+1, 1, -1, 0]\!]$	76
$\mathbb{Z}/10\mathbb{Z}$	12	$[\![\varphi+1, \varphi, \varphi, 0, 0]\!]$	36
$\mathbb{Z}/12\mathbb{Z}$	6	$[\![\varphi, \varphi+1, 0, 2\varphi-3, -\varphi+2]\!]$	220
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$	11	$[\![0, 1, 0, -1, 0]\!]$	80
$\mathbb{Z}/15\mathbb{Z}$	1	$[\![1, 1, 1, -3, 1]\!]$	100
$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$	2	$[\![1, 1, 1, -5, 2]\!]$	45

Table 6. Number of isomorphism classes of elliptic curves over F of norm conductor at most 1831 having given torsion subgroups.

We computed the invariants in the Birch and Swinnerton-Dyer conjecture for our elliptic curves, and solved for the conjectural order of Ш; Table 7 gives the number of elliptic curves in our data having each order of Ш, and Table 8 lists elliptic curves of minimal conductor exhibiting each of these orders.

#Ш	1	4	9	16	25	36
#Isom	3191	84	43	16	2	2

Table 7. Number of isomorphism classes of elliptic curves over F of norm conductor at most 1831 having given order of Ш.

#Ш	First elliptic curve over F having Ш of this order	$N(\mathfrak{n})$
1	$[\![1, \varphi+1, \varphi, \varphi, 0]\!]$	31
4	$[\![1, 1, 1, -110, -880]\!]$	45
9	$[\![\varphi+1, -\varphi, 1, -54686\varphi-35336, -7490886\varphi-4653177]\!]$	76
16	$[\![1, \varphi, \varphi+1, -4976733\varphi-3075797, -6393196918\varphi-3951212998]\!]$	45
25	$[\![0, -1, 1, -7820, -263580]\!]$	121
36	$[\![1, -\varphi+1, \varphi, 1326667\varphi-2146665, 880354255\varphi-1424443332]\!]$	1580

Table 8. Elliptic curves over F of smallest norm conductor having Ш of a given order.

Acknowledgments

We would like to thank John Cremona, Noam Elkies, Tom Fisher, Richard Taylor, John Voight, and the anonymous referee for helpful conversations. We would especially like to thank Joanna Gaski for providing (via the method of Section 3A) the explicit table of elliptic curves that kickstarted this project. We used Sage [32] extensively throughout this project. This work was supported by NSF grant DMS-0757627, administered by the American Institute of Mathematics

References

[1] Baur Bektemirov, Barry Mazur, William Stein, and Mark Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 2, 233–254. MR 2009e:11107

[2] Massimo Bertolini, Henri Darmon, and Peter Green, *Periods and points attached to quadratic algebras*, in Darmon and Zhang [13], 2004, pp. 323–367. MR 2005e:11062

[3] Nicolas Billerey, *Critères d'irréductibilité pour les représentations des courbes elliptiques*, Int. J. Number Theory **7** (2011), no. 4, 1001–1032. MR 2012f:11109

[4] B. J. Birch and W. Kuyk (eds.), *Modular functions of one variable*, *IV*: *Proceedings of the International Summer School on Modular Functions of One Variable and Arithmetical Applications*, *RUCA*, *University of Antwerp*, *July* 17–*August* 3, 1972, Lecture Notes in Mathematics, no. 476, Springer, Berlin, 1975. MR 51 #12708

[5] Jon Bober, Alyson Deines, Ariah Klages-Mundt, Ben LeVeque, R. Andrew Ohana, Ashwath Rabindranath, Paul Sharaba, and William Stein, *A Database of Elliptic Curves over* $\mathbf{Q}(\sqrt{5})$, 2012. http://wstein.org/papers/sqrt5

[6] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[7] Hatice Boylan and Nils-Peter Skoruppa, *Explicit formulas for Hecke Gauss sums in quadratic number fields*, Abh. Math. Semin. Univ. Hambg. **80** (2010), no. 2, 213–226. MR 2012c:11163

[8] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over* **Q**: *Wild* 3*-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939. MR 2002d:11058

[9] J. E. Cremona, *Abelian varieties with extra twist, cusp forms, and elliptic curves over imaginary quadratic fields*, J. London Math. Soc. (2) **45** (1992), no. 3, 404–416. MR 93h:11056

[10] ———, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, 1997. MR 99e:11068

[11] J. E. Cremona and M. P. Lingham, *Finding all elliptic curves with good reduction outside a given set of primes*, Experiment. Math. **16** (2007), no. 3, 303–312. MR 2008k:11057

[12] John Cremona, *Elliptic curve data*, 2012. http://homepages.warwick.ac.uk/staff/J.E.Cremona/ftp/data/INDEX.html

[13] Henri Darmon and Shou-Wu Zhang (eds.), *Heegner points and Rankin L-series*: *Papers from the Workshop on Special Values of Rankin L-Series held in Berkeley, CA, December* 2001, Mathematical Sciences Research Institute Publications, no. 49, Cambridge University Press, Cambridge, 2004. MR 2005b:11002

[14] Lassina Dembélé, *Explicit computations of Hilbert modular forms on* $\mathbf{Q}(\sqrt{5})$, Experiment. Math. **14** (2005), no. 4, 457–466. MR 2006h:11050

[15] ———, *An algorithm for modular elliptic curves over real quadratic fields*, Experiment. Math. **17** (2008), no. 4, 427–438. MR 2010a:11119

[16] Tim Dokchitser, *Computing special values of motivic L-functions*, Experiment. Math. **13** (2004), no. 2, 137–149. MR 2005f:11128

[17] Claus Fieker and David R. Kohel (eds.), *Algorithmic number theory*: *Proceedings of the* 5*th International Symposium* (*ANTS-V*) *held at the University of Sydney, July* 7–12, 2002, Lecture Notes in Computer Science, no. 2369, Berlin, Springer, 2002. MR 2004j:11002

[18] T. A. Fisher, *On families of n-congruent elliptic curves*, 2011. https://www.dpmms.cam.ac.uk/~taf1000/papers/highercongr.html

[19] Matthew Greenberg and John Voight, *Computing systems of Hecke eigenvalues associated to Hilbert modular forms*, Math. Comp. **80** (2011), no. 274, 1071–1092. MR 2012c:11103

[20] Benedict H. Gross and Don B. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. http://resolver.sub.uni-goettingen.de/purl?GDZPPN002102773 MR 87j:11057

[21] Emmanuel Halberstadt and Alain Kraus, *Sur la courbe modulaire* $X_E(7)$, Experiment. Math. **12** (2003), no. 1, 27–40. MR 2004m:11090

[22] Sheldon Kamienny and Filip Najman, *Torsion groups of elliptic curves over quadratic fields*, Acta Arith. **152** (2012), no. 3, 291–305. MR 2885789

[23] Nicholas M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. **62** (1981), no. 3, 481–502. MR 82d:14025

[24] Victor Alecsandrovich Kolyvagin, *On the Mordell-Weil group and the Shafarevich-Tate group of modular elliptic curves*, in Satake [30], 1991, pp. 429–436. http://www.mathunion.org/ICM/ICM1990.1/Main/icm1990.1.0429.0436.ocr.pdf MR 93c:11046

[25] Daniel Sion Kubert, *Universal bounds on the torsion of elliptic curves*, Proc. London Math. Soc. (3) **33** (1976), no. 2, 193–237. MR 55 #7910

[26] Eric Larson and Dmitry Vaintrob, *Determinants of Subquotients of Galois Representations Associated to Abelian Varieties*, 2011. arXiv 1110.0255 [math.NT]

[27] B. Mazur, *Rational isogenies of prime degree* (*with an appendix by D. Goldfeld*), Invent. Math. **44** (1978), no. 2, 129–162. MR 80h:14022

[28] Takayuki Oda, *Periods of Hilbert modular surfaces*, Progress in Mathematics, no. 19, Birkhäuser, Boston, 1982. MR 83k:10057

[29] Michael O. Rubinstein, `lcalc`, 2012. http://code.google.com/p/l-calc/

[30] Ichirō Satake (ed.), *Proceedings of the International Congress of Mathematicians* (*Kyoto*, 1990), vol. 1, Tokyo, Mathematical Society of Japan, 1991. MR 92m:00054

[31] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics, no. 106, Springer, Dordrecht, 2009. MR 2010i:11005

[32] W. A. Stein et al., *Sage Mathematics Software* (*version* 4.8), The Sage Development Team, 2012. http://www.sagemath.org

[33] William A. Stein and Mark Watkins, *A database of elliptic curves—first report*, in Fieker and Kohel [17], 2002, pp. 267–275. MR 2005h:11113

[34] Jacques Vélu, *Isogénies entre courbes elliptiques*, C. R. Acad. Sci. Paris Sér. A-B **273** (1971), A238–A241. http://gallica.bnf.fr/ark:/12148/bpt6k56191248/f52.image MR 45 #3414

[35] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551. MR 96d:11071

[36] Shouwu Zhang, *Heights of Heegner points on Shimura curves*, Ann. of Math. (2) **153** (2001), no. 1, 27–147. MR 2002g:11081

JONATHAN BOBER: `jwbober@math.washington.edu`
Department of Mathematics, University of Washington, Seattle, WA 98195-4350, United States
Current address: *Howard House, University of Bristol, Queens Avenue, Bristol, BS8 1SN, United Kingdom*
http://sage.math.washington.edu/home/bober/www/

ALYSON DEINES: `adeines@math.washington.edu`
Department of Mathematics, University of Washington, Amherst, MA 01002, United States
http://www.math.washington.edu/∼adeines/Alyson_Deines/Welcome.html

ARIAH KLAGES-MUNDT: `aklagesmundt12@amherst.edu`
Department of Mathematics, Amherst College, 1555 Keefe Campus Center, Amherst, MA 01002, United States
https://www.amherst.edu/users/K/aklagesmundt12

BENJAMIN LEVEQUE: `ben.leveque@gmail.com`
Mathematics Department, Brown University, Providence, RI 02906, United States

R. ANDREW OHANA: `ohanar@math.washington.edu`
Department of Mathematics, University of Washington, Seattle, WA 98195, United States

ASHWATH RABINDRANATH: `ashwathr@umich.edu`
Department of Mathematics, University of Michigan, 2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, United States

PAUL SHARABA: `paul.sharaba@gmail.com`
Department of Mathematics, Cleveland State University, Cleveland, OH 44115, United States

WILLIAM STEIN: `wstein@uw.edu`

Department of Mathematics, University of Washington, 423 Padelford Hall,
Seattle, WA 98195-4361, United States
http://www.williamstein.org

msp

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.167

msp

Finding simultaneous Diophantine approximations with prescribed quality

Wieb Bosma and Ionica Smeets

We give an algorithm that finds a sequence of approximations with Dirichlet coefficients bounded by a constant only depending on the dimension. The algorithm uses LLL lattice basis reduction. We present a version of the algorithm that runs in polynomial time of the input.

1. Introduction

The regular continued fraction algorithm is a classical algorithm to approximate reals by rational numbers. The denominators of continued fraction convergents furnish, for every $a \in \mathbb{R}$, infinitely many integers q such that

$$\|q\,a\| < q^{-1},$$

where $\|x\|$ denotes the distance between x and the nearest integer. The exponent -1 of q is minimal; if it is replaced by any number $e < -1$, then there exist real numbers a such that only finitely many integers q satisfy $\|q\,a\| < q^e$.

Hurwitz [9] proved that the continued fraction algorithm finds, for every $a \in \mathbb{R} \setminus \mathbb{Q}$, an infinite sequence of increasing integers q_n with

$$\|q_n\,a\| < \frac{1}{\sqrt{5}}\,q_n^{-1}.$$

If the constant $1/\sqrt{5}$ is replaced by any smaller one, then this statement is false. Legendre [15] showed that the continued fraction algorithm finds all good approximations, in the sense that if for some positive integer q

$$\|q\,a\| < \tfrac{1}{2}\,q^{-1},$$

then q is one of the q_n found by the algorithm.

MSC2010: primary 11J13; secondary 11Y16, 11J70.
Keywords: simultaneous Diophantine approximation, LLL lattice reduction.

As to the generalization of approximations in higher dimensions Dirichlet [16] proved the following theorem; see Chapter II of [19].

Theorem 1.1. *Let an $n \times m$ matrix A with entries $a_{ij} \in \mathbb{R} \setminus \mathbb{Q}$ be given and suppose that $1, a_{i1}, \ldots, a_{im}$ are linearly independent over $\mathbb{Q}$ for some i with $1 \le i \le n$. There exist infinitely many coprime m-tuples of integers $(q_1, \ldots, q_m)$ such that, with $q = \max_j |q_j| \ge 1$, we have*

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| < q^{-m/n}. \tag{1}$$

If the exponent $-m/n$ is replaced by any smaller number, there exists a matrix A for which the inequality holds for only finitely many coprime tuples $(q_1, q_2, \ldots, q_m)$.

Definition 1.2. Let an $n \times m$ matrix A with entries $a_{ij} \in \mathbb{R} \setminus \mathbb{Q}$ be given. The *Dirichlet coefficient* of an m-tuple $(q_1, \ldots, q_m)$ is $q^{m/n} \max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\|$.

The proof of the theorem does not give an efficient way of finding a series of approximations with a Dirichlet coefficient less than 1. For the case $m = 1$ the first multidimensional continued fraction algorithm was given by Jacobi [10]. Many more followed, see for instance Perron [18], Brun [5; 6], Lagarias [14] and Just [11]. Brentjes [4] gives a detailed history and description of such algorithms. Schweiger's book [20] gives a broad overview. For $n = 1$ there is, amongst others, the algorithm by Ferguson and Forcade [8]. However, there is no efficient algorithm guaranteed to find a series of approximations with Dirichlet coefficient smaller than 1.

In 1982 the LLL algorithm for lattice basis reduction was published in [17]. The authors noted that their algorithm could be used for finding Diophantine approximations of given rationals with Dirichlet coefficient only depending on the dimension; see Corollary 2.4. Just [11] developed an algorithm based on lattice reduction that detects $\mathbb{Z}$-linear dependence in the a_i, in the case $m = 1$. If no such dependence is found her algorithm returns integers q with

$$\max_i \|q a_i\| \le c \left(\sum_{i=1}^{n} a_i^2 \right)^{1/2} q^{-1/(2n(n-1))},$$

where c is a constant depending on n. The exponent $-1/(2n(n-1))$ is larger than the Dirichlet exponent $-1/n$. Lagarias [13] used the LLL algorithm in a series of lattices to find good approximations for the case $m = 1$. Let $a_1, \ldots, a_n \in \mathbb{Q}$ and let N be a positive integer; suppose there exists $Q \in \mathbb{N}$ with $1 \le Q \le N$ such that $\max_j \|Q a_j\| < \varepsilon$. Then Lagarias's algorithm on input $a_1, \ldots, a_n$ and N finds in polynomial time a q with $1 \le q \le 2^{n/2} N$ such that $\max_j \|q a_j\| \le \sqrt{5n} 2^{(n-1)/2} \varepsilon$. One difference with our work is that Lagarias focuses on the quality $\|q a_j\|$, while we focus on the Dirichlet coefficient $q^{1/n} \|q a_j\|$. We also consider the case $m > 1$.

The main result of the present paper is an algorithm that by iterating the LLL algorithm gives a series of approximations of given rationals with optimal Dirichlet exponent. Where the LLL algorithm gives one approximation, our dynamic algorithm gives a series of successive approximations. To be more precise: For a given $n \times m$-matrix A with entries $a_{ij} \in \mathbb{Q}$ and a given upper bound $q_{\max}$ the algorithm returns a sequence of m-tuples $(q_1, \ldots, q_m)$ such that for every Q with $2^{(m+n+3)(m+n)/(4m)} \leq Q \leq q_{\max}$ one of these m-tuples satisfies

$$\max_j |q_j| \leq Q$$

and

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \leq 2^{(m+n+3)(m+n)/(4n)} Q^{-m/n}.$$

The exponent $-m/n$ of Q can not be improved, and therefore we say that these approximations have *optimal Dirichlet exponent*.

Our algorithm is a multidimensional continued fraction algorithm in the sense that we work in a lattice basis and that we only interchange basis vectors and add integer multiples of basis vectors to another. Our algorithm differs from other multidimensional continued fraction algorithms in that the lattice is not fixed across iterations. In Lemma 3.6 we show that if there exists an extremely good approximation, our algorithm finds a very good one. We derive in Theorem 3.8 how the output of our algorithm gives a lower bound on the quality of possible approximations with coefficients up to a certain limit. In Section 4 we show that a slightly modified version of our algorithm runs in polynomial time. In Section 5 we present some numerical data.

An earlier version of this paper appeared as Chapter V of Smeets's thesis [21].

2. Lattice reduction and the LLL algorithm

In this section we give the definitions and results that we need for our algorithm.

Let r be a positive integer. A subset L of the r-dimensional real Euclidean vector space $\mathbb{R}^r$ is called a *lattice* if there exists a basis $b_1, \ldots, b_r$ of $\mathbb{R}^r$ such that

$$L = \sum_{i=1}^{r} \mathbb{Z} b_i = \left\{ \sum_{i=1}^{r} z_i b_i \;\middle|\; z_i \in \mathbb{Z} \text{ for } i = 1, \ldots, r \right\}.$$

We say that $b_1, \ldots, b_r$ is a *basis* for L. The *determinant* of the lattice L is defined by $|\det(b_1, \ldots, b_r)|$ and we denote it as $\det L$.

For any linearly independent $b_1, \ldots, b_r \in \mathbb{R}^r$ the Gram-Schmidt process yields an orthogonal basis $b_1^*, \ldots, b_r^*$ for $\mathbb{R}^r$, defined inductively by

$$b_i^* = b_i - \sum_{j=1}^{i-1} \mu_{ij} b_j^* \quad \text{for } 1 \leq i \leq r$$

and

$$\mu_{ij} = \frac{(b_i, b_j^*)}{(b_j^*, b_j^*)},$$

where $(\ ,\ )$ denotes the ordinary inner product on $\mathbb{R}^r$.

We call a basis $b_1, \ldots, b_r$ for a lattice L *reduced* if

$$|\mu_{ij}| \leq \tfrac{1}{2} \qquad \text{for } 1 \leq j < i \leq r$$

and

$$\left|b_i^* + \mu_{ii-1} b_{i-1}^*\right|^2 \leq \tfrac{3}{4} |b_{i-1}^*|^2 \qquad \text{for } 1 \leq i \leq r,$$

where $|x|$ denotes the Euclidean length of x.

Proposition 2.1 [17, Proposition 1.6]. *Let $b_1, \ldots, b_r$ be a reduced basis for a lattice L in $\mathbb{R}^r$. Then*

(1) $|b_1| \leq 2^{(r-1)/4} (\det L)^{1/r}$,

(2) $|b_1|^2 \leq 2^{r-1} |x|^2$ *for every nonzero* $x \in L$,

(3) $\prod_{i=1}^{r} |b_i| \leq 2^{r(r-1)/4} \det L$.

Proposition 2.2 [17, Proposition 1.26]. *Let $L \subset \mathbb{Z}^r$ be a lattice with a basis $b_1, b_2, \ldots, b_r$, and let $F \in \mathbb{R}$, $F \geq 2$, be such that $|b_i|^2 \leq F$ for $1 \leq i \leq r$. Then the number of arithmetic operations needed by the LLL algorithm is $O(r^4 \log F)$ and the integers on which these operations are performed each have binary length $O(r \log F)$.*

In the following lemma the approach suggested in the original LLL-paper for finding (simultaneous) Diophantine approximations is generalized to the case $m > 1$.

Lemma 2.3. *Let an $n \times m$-matrix A with entries $a_{ij} \in \mathbb{R}$ and an $\varepsilon \in (0, 1)$ be given. Let L be the lattice formed by the columns of the $(m+n) \times (m+n)$-matrix*

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 & a_{11} & \cdots & a_{1m} \\ 0 & 1 & & 0 & a_{21} & \cdots & a_{2m} \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 1 & a_{n1} & \cdots & a_{nm} \\ 0 & \cdots & 0 & 0 & c & & 0 \\ \vdots & & \vdots & \vdots & & \ddots & \\ 0 & \cdots & 0 & 0 & 0 & & c \end{bmatrix}, \tag{2}$$

with $c = (2^{-(m+n-1)/4} \varepsilon)^{(m+n)/m}$.

The LLL algorithm applied to L will yield an m-tuple $(q_1, \ldots, q_m)$ of integers with

$$\max_j |q_j| \le 2^{(m+n-1)(m+n)/(4m)} \varepsilon^{-n/m} \tag{3}$$

and

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \le \varepsilon.$$

Proof. The LLL algorithm finds a reduced basis $b_1, \ldots, b_{m+n}$ for the lattice L. For each vector b in this basis there exist $p_i \in \mathbb{Z}$, for $1 \le i \le n$, and $q_j \in \mathbb{Z}$, for $1 \le j \le m$, such that

$$b = \begin{bmatrix} q_1 a_{11} + \cdots + q_m a_{1m} - p_1 \\ \vdots \\ q_1 a_{n1} + \cdots + q_m a_{nm} - p_n \\ cq_1 \\ \vdots \\ cq_m \end{bmatrix}.$$

Proposition 2.1(i) gives an upper bound for the length of the first basis vector,

$$|b_1| \le 2^{(m+n-1)/4} c^{m/(m+n)}.$$

From this vector b_1 we find integers $q_1, \ldots, q_m$, such that

$$\max_j |q_j| \le 2^{(m+n-1)/4} c^{-n/(m+n)} \tag{4}$$

and

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \le 2^{(m+n-1)/4} c^{m/(m+n)}. \tag{5}$$

Substituting $c = (2^{-(m+n-1)/4}\varepsilon)^{(m+n)/m}$ gives the results. □

From (4) and (5) we obtain the following corollary.

Corollary 2.4. *For any $n \times m$-matrix A with entries $a_{ij} \in \mathbb{R}$ the LLL algorithm can be used to obtain an m-tuple $(q_1, \ldots, q_m)$ that satisfies, with $q = \max_j |q_j|$,*

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \le 2^{(m+n-1)(m+n)/(4n)} q^{-m/n}.$$

3. The iterated LLL algorithm

We iterate the LLL algorithm over a series of lattices to find a sequence of approximations. We start with a lattice determined by a basis of the form (2). After the LLL algorithm finds a reduced basis for this lattice, we decrease the constant c by dividing the last m rows of the matrix by a constant d greater than 1. By doing

so, ε is divided by $d^{m/(m+n)}$. We repeat this process until the upper bound (3) for $\max |q_j|$ guaranteed by the LLL algorithm exceeds a given upper bound $q_{\max}$.

To ease notation we put $d = 2$ and $\varepsilon = 1/2$.

Algorithm 3.1 (Iterated LLL algorithm (ILLL)).

Input: An $n \times m$-matrix A with entries a_{ij} in $\mathbb{R}$, and an upper bound $q_{\max} > 1$.

Output: For each integer $k \geq 1$ no larger than the k' defined in (8), a vector $q(k) \in \mathbb{Z}^m$ with

$$\max_j |q_j(k)| \leq 2^{(m+n-1)(m+n)/(4m)}\, 2^{kn/m} \tag{6}$$

and

$$\max_i \|q_1(k)\, a_{i1} + \cdots + q_m(k)\, a_{im}\| \leq 1/2^k. \tag{7}$$

1. Construct the basis matrix B as given in (2) from A.
2. Apply the LLL algorithm to B.
3. Deduce $q_1, \ldots, q_m$ from the first vector in the reduced basis returned by the LLL algorithm.
4. Divide the last m rows of B by $2^{(m+n)/m}$
5. Stop if the upper bound for q guaranteed by the algorithm (6) exceeds $q_{\max}$; else go to Step 2.

Remark 3.2. The number $2^{(m+n)/m}$ in Step 4 may be replaced by $d^{(m+n)/m}$ for any real number $d > 1$. When we additionally set $\varepsilon = 1/d$ this yields

$$\max_j |q_j(k)| \leq 2^{(m+n-1)(m+n)/(4m)} d^{kn/m}$$

and

$$\max_i \|q_1(k)a_{i1} + \cdots + q_m(k)a_{im}\| < d^{-k}.$$

In this paper, with the exception of the numerical examples in Section 5, we always take $d = 2$ and $\varepsilon = 1/2$.

Define

$$k' := \left\lceil -\frac{(m+n-1)(m+n)}{4n} + \frac{m \log_2 q_{\max}}{n} \right\rceil. \tag{8}$$

Lemma 3.3. *Let an $n \times m$-matrix A with entries a_{ij} in $\mathbb{R}$ and an upper bound $q_{\max} > 1$ be given. With this input, the number of times the ILLL algorithm applies the LLL algorithm equals k' from* (8).

Proof. One derives the number of iterations by solving k from the stopping criterion (6)

$$q_{\max} \leq 2^{(m+n-1)(m+n)/(4m)} 2^{kn/m},$$

that is:

$$\frac{m}{n}\log_2 q_{\max} \le \frac{(m+n-1)(m+n)}{4n} + k.$$

We stop iterating as soon as the integer k reaches the ceiling k' as in (8). □

For each $k \ge 1$ we define

$$c(k) = (2^{-k-(m+n-5)/4}\varepsilon)^{(m+n)/m}.$$

Note that $c(1)$ is the constant c from Lemma 2.3. In the k-th iteration we are working in the lattice defined by the basis in (2) with c replaced by $c(k)$.

Lemma 3.4. *The output $q(k) = (q_1(k), q_2(k), \ldots, q_m(k))$ of the ILLL algorithm satisfies* (6) *and* (7), *for $1 \le k \le k'$.*

Proof. Since we take $\varepsilon = 1/2$, in the k-th iteration we use

$$c(k) = (2^{-k-(m+n-1)/4})^{(m+n)/m}.$$

Substituting $c(k)$ for c in (4) and (5) yields (6) and (7), respectively. □

The following theorem gives the main result of the present paper, as mentioned in the introduction. The algorithm returns a sequence of approximations with all coefficients smaller than Q, optimal Dirichlet exponent and Dirichlet coefficient only depending on the dimensions m and n.

Theorem 3.5. *Let an $n \times m$-matrix A with entries a_{ij} in $\mathbb{R}$, and $q_{\max} > 1$ be given. The ILLL algorithm finds a sequence of m-tuples $(q_1, \ldots, q_m)$ of integers such that for every Q with $2^{(m+n+3)(m+n)/(4m)} \le Q \le q_{\max}$ one of these m-tuples satisfies*

$$\max_j |q_j| \le Q \quad \text{and}$$

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \le 2^{(m+n+3)(m+n)/(4n)} Q^{-m/n}.$$

Proof. Take $k \in \mathbb{N}$ such that

$$2^{(k-1)n/m} \le Q \cdot 2^{4m/((m+n+3)(m+n))} < 2^{kn/m}. \tag{9}$$

From Lemma 3.4 we know that $q(k) = (q_1(k), q_2(k), \ldots, q_m(k))$ satisfies the inequality

$$\max_j |q_j(k)| \le 2^{(m+n+3)(m+n)/(4m)}\, 2^{(k-1)n/m} \le Q.$$

From the right side of inequality (9) it follows that

$$\frac{1}{2^k} < 2^{(m+n+3)(m+n)/(4n)} Q^{-m/n}.$$

From Lemma 3.4 and this inequality we derive that

$$\max_i \|q_1(k)\, a_{i1} + \cdots + q_m(k)\, a_{im}\| \le \frac{1}{2^k} < 2^{(m+n+3)(m+n)/(4n)} Q^{-m/n}. \quad \square$$

Proposition 2.1(2) guarantees that if there exists an extremely short vector in the lattice, then the LLL algorithm finds a rather short lattice vector. We extend this result to the realm of successive approximations. In the next lemma we show that for every very good approximation (satisfying (11)), the ILLL algorithm finds a rather good one (satisfying (14)) not too far away from it (as specified by (13)).

Lemma 3.6. *Let an $n \times m$-matrix A with entries a_{ij} in $\mathbb{R}$, a real number $0 < \delta < 1$, and an integer $s > 1$ be given. If there exists an m-tuple $(s_1, \ldots, s_m)$ of integers with*

$$s = \max_j |s_j| > 2^{(m+n-1)n/(4m)} \left(\frac{n\delta^2}{m}\right)^{n/(2(m+n))} \tag{10}$$

and

$$\max_i \|s_1 a_{i1} + \cdots + s_m a_{im}\| \le \delta s^{-m/n}, \tag{11}$$

then applying the ILLL algorithm with

$$q_{\max} > 2^{(m^2+m(n-1)+4n)/(4m)} \left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s \tag{12}$$

yields an m-tuple $(q_1, \ldots, q_m)$ of integers with

$$\max_j |q_j| \le 2^{(m^2+m(n-1)+4n)/(4m)} \left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s \tag{13}$$

and

$$\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \le 2^{(m+n)/2} \sqrt{n}\delta s^{-m/n}. \tag{14}$$

Proof. Let $1 \le k \le k'$ be an integer. Proposition 2.1(2) gives that for each m-tuple $q(k)$ found by the algorithm, we have

$$\sum_{i=1}^{n} \|q_1(k)a_{i1} + \cdots + q_m(k)a_{im}\|^2 + c(k)^2 \sum_{j=1}^{m} q_j(k)^2 \le 2^{m+n-1}\left(\sum_{i=1}^{n} \|s_1 a_{11} + \cdots + s_m a_{im}\|^2 + c(k)^2 \sum_{j=1}^{m} s_j^2\right).$$

From this and (10) and (11) it follows that

$$\max_i \|q_1(k)a_{i1} + \cdots + q_m(k)a_{im}\|^2 \le 2^{m+n-1}\left(n\delta^2 s^{-2m/n} + c(k)^2 m s^2\right). \tag{15}$$

Take the smallest positive integer K such that

$$c(K) \le \sqrt{\frac{n}{m}}\,\delta s^{-(m+n)/n}. \tag{16}$$

We find for the K-th iteration from (15) and (16)

$$\max_i \|q_1(K)a_{i1} + \cdots + q_m(K)a_{im}\| \le 2^{(m+n)/2}\sqrt{n}\delta s^{-m/n},$$

which gives (14).

We show that under assumption (12) the ILLL algorithm performs at least K iterations. We may assume $K > 1$, since the ILLL algorithm always performs at least 1 iteration. From Lemma 3.3 we find that if $q_{\max}$ satisfies

$$q_{\max} > 2^{Kn/m}2^{(m+n-1)(m+n)/(4m)},$$

then the ILLL algorithm performs at least K iterations. Our choice of K implies

$$c(K-1) = \frac{c(1)}{2^{(m+n)(K-2)/m}} = \frac{2^{-(m+n+3)(m+n)/(4m)}}{2^{(m+n)(K-2)/m}} > \sqrt{\frac{n}{m}}\delta s^{-(m+n)/n},$$

and we obtain

$$2^{Kn/m} < 2^{-(m+n-5)n/(4m)}\left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s.$$

From this we find that

$$q_{\max} > 2^{(m^2+m(n-1)+4n)/(4m)}\left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s$$

is a sufficient condition to guarantee that the algorithm performs at least K iterations.

Furthermore, either $2^{-(m+n)/m}\sqrt{n/m}\,\delta s^{-(m+n)/n} < c(K)$ or $K = 1$. In the former case we find from (4) that

$$\max_j |q_j(K)| \le 2^{(m+n-1)/4}c(K)^{-n/(m+n)} < 2^{(m+n-1)/4}2^{n/m}\left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s.$$

In the latter case we obtain from (4) that

$$\max_j |q_j(1)| \le 2^{(m+n-1)/4}c(1)^{-n/(m+n)} = 2^{(m+n-1)/4}2^{(m+n+3)n/(4m)}$$

and, by (10),

$$\begin{aligned} 2^{(m+n-1)/4}2^{(m+n+3)n/(4m)} &= 2^{(m+n-1)/4}2^{n/m}2^{(m+n-1)n/(4m)} \\ &< 2^{(m+n-1)/4}2^{n/m}\left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s. \end{aligned}$$

We conclude that for all $K \geq 1$,

$$\max_j |q_j(K)| \leq 2^{(m^2+m(n-1)+4n)/(4m)} \left(\frac{m}{n\delta^2}\right)^{n/(2(m+n))} s. \qquad \square$$

From (13) and (14) we obtain the following corollary.

Corollary 3.7. *With the assumptions of Lemma* 3.6, *the ILLL algorithm can be used to obtain an m-tuple* $(q_1, \ldots, q_m)$ *of integers that satisfies*

$$q^{m/n} \max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \leq 2^{(m^2+m(3n-1)+4n+2n^2)/(4n)} m^{m/(2(m+n))} (n\delta^2)^{n/(2(m+n))},$$

where again $q = \max_j |q_j|$.

Theorem 3.8. *Let an* $n \times m$*-matrix A with entries* a_{ij} *in* $\mathbb{R}$ *and* $q_{\max} > 1$ *be given. Assume that* γ *is such that for every m-tuple* $(q_1, \ldots, q_m)$ *returned by the ILLL algorithm, we have*

$$q^{m/n} \max_i \|q_1 a_{i1} + \ldots q_m a_{im}\| > \gamma, \quad \text{where } q = \max_j |q_j|. \tag{17}$$

Set

$$\delta = 2^{-(m+n)(m^2+m(3n-1)+4n+2n^2)/(4n^2)} m^{-m/(2n)} n^{-1/2} \gamma^{(m+n)/n}. \tag{18}$$

Let $(s_1, \ldots, s_m)$ *be an m-tuple of integers, and set* $s = \max_j |s_j|$. *If*

$$s > 2^{(m+n-1)n/(4m)} \left(\frac{n\delta^2}{m}\right)^{n/(2(m+n))} \tag{19}$$

and

$$s < 2^{-(m^2+m(n-1)+4n)/(4m)} \left(\frac{n\delta^2}{m}\right)^{n/(2(m+n))} q_{\max} \tag{20}$$

then

$$s^{m/n} \max_i \|s_1 a_{i1} + \cdots + s_m a_{im}\| > \delta. \tag{21}$$

Proof. Assume that every vector returned by our algorithm satisfies (17) and that there exists an m-tuple $(s_1, \ldots, s_m)$ satisfying (19) and (20) but not satisfying Equation (21). From Equation (20) it follows that $q_{\max}$ satisfies (12). We apply Lemma 3.6 and find that the algorithm finds an m-tuple $(q_1, \ldots, q_m)$ that satisfies (17). Substituting δ as given in (18) gives

$$q^{m/n} \max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| \leq \gamma,$$

in contradiction with our assumption. $\square$

4. A polynomial time version of the ILLL algorithm

We have used real numbers in our theoretical results, but in a practical implementation of the algorithm we only use rational numbers. Without loss of generality we may assume that these numbers are in the interval $[0, 1]$. In this section we describe the changes to the algorithm and we show that this modified version of the algorithm runs in polynomial time.

As input for the rational algorithm we take

- the dimensions m and n,
- a rational number $\varepsilon \in (0, 1)$,
- an integer M that is large compared to $\dfrac{(m+n)^2}{m} - \dfrac{m+n}{m} \log \varepsilon$,
- an $n \times m$-matrix A with entries $0 < a_{ij} \leq 1$, where each $a_{ij} = p_{ij}/2^M$ for some integer p_{ij},
- an integer $q_{\max} < 2^M$.

Remark 4.1. In this rational algorithm all irrational numbers are approximated by rational numbers with denominator 2^M. Thus M denotes the precision that is used.

When we construct the matrix B in Step 1 of the ILLL algorithm we approximate c as given in (2) by a rational number

$$\hat{c} = 2^{-M} \lceil 2^M c \rceil = 2^{-M} \left\lceil 2^M \left(2^{-(m+n-1)/4} \varepsilon\right)^{(m+n)/m} \right\rceil. \tag{22}$$

Hence $c < \hat{c} \leq c + 2^{-M}$.

In iteration k we use a rational $\hat{c}\,(k)$ that for $k \geq 2$ is given by

$$\hat{c}\,(k) = 2^{-M} \left\lceil 2^M \hat{c}\,(k-1) 2^{-(m+n)/m} \right\rceil \quad \text{and} \quad \hat{c}\,(1) = \hat{c} \text{ as in (22)},$$

and we change Step 4 of the ILLL algorithm to "Multiply the last m rows of B by $\hat{c}\,(k-1)/\,\hat{c}\,(k)$." The other steps of the rational iterated algorithm are as described in Section 3.

The running time of the rational algorithm.

Theorem 4.2. *Let the input be given as described above. Then the number of arithmetic operations needed by the ILLL algorithm and the binary length of the integers on which these operations are performed are both bounded by a polynomial in m, n, and M.*

Proof. The number of times we apply the LLL algorithm is not changed by rationalizing c, so we find the number of iterations k' from Lemma 3.3

$$k' = \left\lceil -\frac{(m+n-1)(m+n)}{4n} + \frac{m \log_2 q_{\max}}{n} \right\rceil < \left\lceil \frac{mM}{n} \right\rceil.$$

It is obvious that Steps 1, 3, 4 and 5 of the algorithm are polynomial in the size of the input and we focus on the LLL-step. We determine an upper bound for the length of a basis vector used at the beginning of an iteration in the ILLL algorithm.

In the first application of the LLL algorithm the length of the initial basis vectors as given in (2) is bounded by

$$|b_i|^2 \le \max_j \{1, a_{1j}^2 + \cdots + a_{nj}^2 + m\hat{c}^{\,2}\} \le m+n \quad \text{for } 1 \le i \le m+n,$$

where we use that $0 < a_{ij} < 1$ and $\hat{c} \le 1$.

The input of each following application of the LLL algorithm is derived from the reduced basis found in the previous iteration by making some of the entries strictly smaller. Part (2) of Proposition 2.1 yields that for every vector b_i in a reduced basis we have

$$|b_i|^2 \le 2^{(m+n)(m+n-1)/2} (\det L)^2 \prod_{\substack{j=1 \\ j \ne i}}^{m+n} |b_i|^{-2}.$$

The determinant of our starting lattice is given by $\hat{c}^{\,m}$ and the determinants of all subsequent lattices are strictly smaller. Every vector b_i in the lattice is at least as long as the shortest nonzero vector in the lattice. Thus for each i we have $|b_i|^2 \ge \frac{1}{2^M}$. Combining this yields

$$|b_i|^2 \le 2^{(m+n+2M)(m+n-1)/2}\, \hat{c}^{\,2m} \le 2^{(m+n+2M)(m+n-1)/2}$$

for every vector used as input for the LLL-step after the first iteration.

Thus we have

$$|b_i|^2 < \max\{m+n\,,\, 2^{(m+n+2M)(m+n-1)/2}\} = 2^{(m+n+2M)(m+n-1)/2} \qquad (23)$$

for any basis vector that is used as input for an LLL-step in the ILLL algorithm.

Proposition 2.2 shows that for a given basis $b_1, \ldots, b_{m+n}$ for $\mathbb{Z}^{m+n}$ with $F \in \mathbb{R}$, $F \ge 2$ such that $|b_i|^2 \le F$ for $1 \le i \le m+n$ the number of arithmetic operations needed to find a reduced basis from this input is $O((m+n)^4 \log F)$. For matrices with entries in $\mathbb{Q}$ we need to clear denominators before applying this proposition. Thus for a basis with basis vectors $|b_i|^2 \le F$ and rational entries that can all be written as fractions with denominator 2^M the number of arithmetic operations is $O((m+n)^4 \log(2^{2M} F))$.

Combining this with (23) and the number of iterations yields the theorem. □

Approximation results from the rational algorithm. Assume that the input matrix A (with entries $a_{ij} = 2^{-M} p_{ij} \in \mathbb{Q}$) is an approximation of an $n \times m$-matrix $\mathcal{A}$

(with entries $\alpha_{ij} \in \mathbb{R}$), found by putting $a_{ij} = 2^{-M}\lceil 2^M \alpha_{ij}\rceil$. In this subsection we derive the approximation results guaranteed by the rational iterated algorithm for the $\alpha_{ij} \in \mathbb{R}$.

According to (4) and (5) the LLL algorithm, applied with $\hat{c}$ instead of c, is guaranteed to find an m-tuple $(q_1, \ldots, q_m)$ such that

$$q = \max_j |q_j| \le 2^{(m+n-1)(m+n)/(4m)} \varepsilon^{-n/m}$$

and

$$\begin{aligned}\max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| &\le 2^{(m+n-1)/4}\left(\left(2^{-(m+n-1)/4}\varepsilon\right)^{(m+n)/m} + 2^{-M}\right)^{m/(m+n)} \\ &\le \varepsilon + 2^{(m+n-1)/4 - Mm/(m+n)},\end{aligned}$$

the last inequality following from the fact that $(x+y)^\alpha \le x^\alpha + y^\alpha$ if $\alpha < 1$ and $x, y > 0$.

For the α_{ij} we find that

$$\begin{aligned}\max_i \|q_1 \alpha_{i1} + \cdots + q_m \alpha_{im}\| &\le \max_i \|q_1 a_{i1} + \cdots + q_m a_{im}\| + mq\, 2^{-M} \\ &\le \varepsilon + 2^{(m+n-1)/4 - Mm/(m+n)} + m\, \varepsilon^{-n/m}\, 2^{(m+n-1)(m+n)/(4m) - M}.\end{aligned}$$

In the introduction to Section 4 we have chosen M large enough to guarantee that the error introduced by rationalizing the entries is negligible.

We show that the difference between $\hat{c}\,(k)$ and $c(k)$ is bounded by $2/2^M$.

Lemma 4.3. *For each integer $k \ge 0$,*

$$c(k) \le \hat{c}\,(k) < c(k) + 2^{-M} \sum_{i=0}^{k} 2^{-i(m+n)/m} < c(k) + \frac{2}{2^M}.$$

Proof. We use induction. For $k = 0$ we have $\hat{c}\,(0) = 2^{-M}\left\lceil c(0) 2^M \right\rceil$ and trivially

$$c(0) \le \hat{c}\,(0) < c(0) + \frac{1}{2^M}.$$

Assume that

$$c(k-1) \le \hat{c}\,(k-1) < c(k-1) + 2^{-M} \sum_{i=0}^{k-1} 2^{-i(m+n)/m}$$

and consider $\hat{c}\,(k)$. From the definition of $\hat{c}\,(k)$ and the induction assumption it

follows that

$$\begin{aligned}\hat{c}\,(k) &= 2^{-M}\left\lceil \hat{c}\,(k-1)\,2^{-(m+n)/m}2^{M}\right\rceil \\ &\geq 2^{-(m+n)/m}\,\hat{c}\,(k-1) \geq 2^{-(m+n)/m}\,c(k-1) = c(k)\end{aligned}$$

and

$$\begin{aligned}\hat{c}\,(k) &= 2^{-M}\left\lceil \hat{c}\,(k-1)\,2^{-(m+n)/m}2^{M}\right\rceil \\ &< 2^{-(m+n)/m}\,\hat{c}\,(k-1) + 2^{-M} \\ &< 2^{-(m+n)/m}\left(c(k-1) + 2^{-M}\sum_{i=0}^{k-1} 2^{-i(m+n)/m}\right) + 2^{-M} \\ &= c(k) + 2^{-M}\sum_{i=0}^{k} 2^{-i(m+n)/m}.\end{aligned}$$

Finally note that $\sum_{i=0}^{k} 2^{-i(m+n)/m} < 2$ for all k. □

One can derive analogues of Theorem 3.5, Lemma 3.6 and Theorem 3.8 for the polynomial version of the ILLL algorithm by carefully adjusting for the introduced error. We do not give the details, since in practice this error is negligible.

5. Experimental data

In this section we present some experimental data from the rational ILLL algorithm. In our experiments we choose the dimensions m and n and iteration speed d, so $\varepsilon = \frac{1}{d}$. We fill the $m \times n$ matrix A with random numbers in the interval $[0, 1]$ and repeat the entire ILLL algorithm for a large number of these random matrices to find our results. First we look at the distribution of the approximation quality. Then we look at the growth of the denominators q found by the algorithm.

The distribution of the approximation qualities. For one-dimensional continued fractions the approximation coefficients Θ_k are defined as

$$\Theta_k = q_k^2\left|a - \frac{p_k}{q_k}\right|,$$

where p_k/q_k is the kth convergent of a.

For the multidimensional case we define Θ_k in a similar way:

$$\Theta_k = q(k)^{m/n}\max_i \|q_1(k)\,a_{i1} + \cdots + q_m(k)\,a_{im}\|.$$

The one-dimensional case $m = n = 1$. We compare the distribution of the Θ_k found by the ILLL algorithm for $m = n = 1$ and various values of d with the distribution of the Θ_k as produced by the continued fraction algorithm with the

best approximation properties. For this optimal continued fraction algorithm it was shown in [2] that for almost all a, the limit

$$\lim_{N\to\infty} \frac{1}{N} \#\{k : 1 \le k \le N \text{ and } \Theta_k \le z\}$$

is equal to $F(z)$, where

$$F(z) = \begin{cases} \dfrac{z}{\log G} & \text{if } 0 \le z \le 1/\sqrt{5}, \\ \dfrac{1}{\log G}\left(\sqrt{1-4z^2} + \log\left(G\dfrac{1-\sqrt{1-4z^2}}{2z}\right)\right) & \text{if } 1/\sqrt{5} \le z \le 1/2, \\ 1 & \text{if } 1/2 \le z \le 1, \end{cases}$$

with $G = (\sqrt{5}+1)/2$.

The optimal continued fraction algorithm finds rational approximations of which the denominators grow with maximal rate, and it finds all approximations with $\Theta_k < 1/2$; for all this, see [1; 2; 3].

The following figures display distribution functions for Θ_k; that is, we show the fraction of the Θ_k found up to the value given on the horizontal axis.

We plot the distribution of the Θ_k found by the ILLL algorithm for $m = n = 1$ and $d = 2$ in Figure 1. The ILLL algorithm might find the same approximation more than once. We see in Figure 1 that for $d = 2$ the distribution function differs depending on whether we leave in the duplicates or sort them out. With the duplicate approximations removed the distribution of Θ_k strongly resembles $F(z)$ of the optimal continued fraction. The duplicates that the ILLL algorithm finds are usually good approximations: If they are much better than necessary they will also be an admissible solution in the next few iterations.

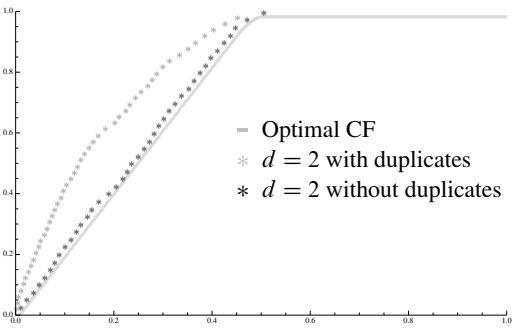


Figure 1. Distribution function for Θ_k from ILLL with $m = n = 1$ and $d = 2$, with and without the duplicate approximations, compared to that of Θ_k for optimal continued fractions.

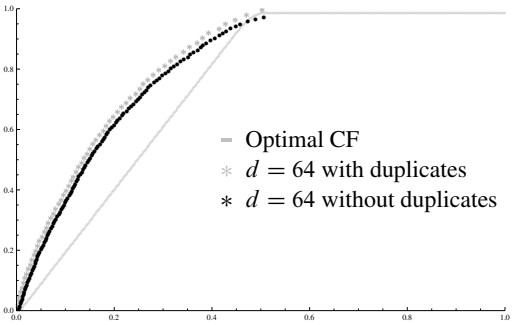


Figure 2. Distribution function for Θ_k from ILLL with $m = n = 1$ and $d = 64$, with and without the duplicate approximations, compared to that of Θ_k for optimal continued fractions.

For larger d we do not find so many duplicates, because the quality has to improve much more in every iteration; also see Figure 2 for an example with $d = 64$.

From now on we remove duplicates from our results.

The multidimensional case. In this section we show some results for the distribution of the Θ_k's found by the ILLL algorithm. For fixed m and n there also appears to be a limit distribution for Θ_k as d grows. See Figure 3 (right) for an example with $m = 3$ and $n = 2$, and compare this with the left half of the same figure. In this section we fix $d = 512$.

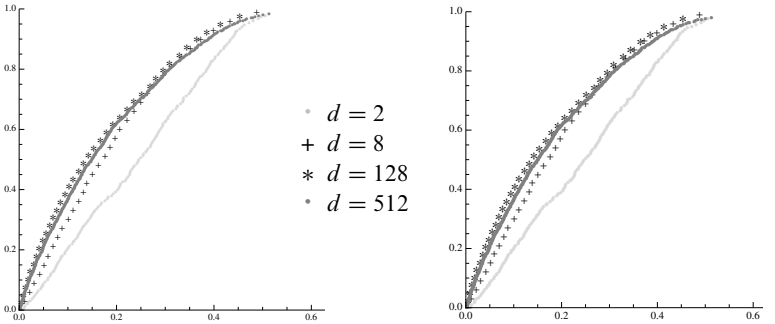


Figure 3. Distribution function for Θ_k from ILLL (with duplicates removed) for $d = 2, 8, 128$ and 512. Left: $m = n = 1$. Right: $m = 3$ and $n = 2$.

In Figure 4 we show some distributions for cases where either m or n is 1.

In Figure 5 we show some distributions for cases where $m = n$.

Remark 5.1. Very rarely the ILLL algorithm returns an approximation with $\Theta_k > 1$.

The denominators q. For regular continued fractions, denominators grow exponentially fast; to be more precise, for almost all x we have (see Section 3.5 of [7])

$$\lim_{k\to\infty} q_k^{1/k} = e^{\pi^2/(12\log 2)},$$

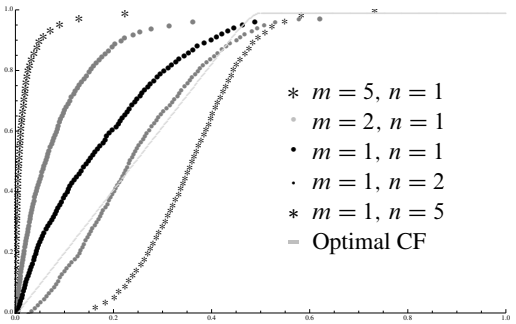


Figure 4. Distribution for Θ_k from ILLL when either $m = 1$ or $n = 1$.

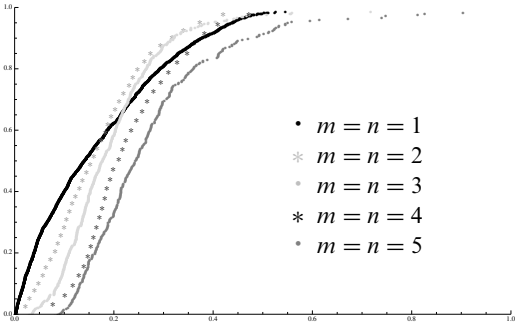


Figure 5. Distribution of Θ_k from ILLL when $m = n$.

For optimal continued fractions, the constant $\pi^2/(12 \log 2)$ in this expression is replaced by $\pi^2/(12 \log G)$, where $G = (\sqrt{5} + 1)/2$. For multidimensional continued fraction algorithms little is known about the distribution of the denominators q_j. Lagarias defined in [12] the notion of a best simultaneous Diophantine approximation and showed that for the ordered denominators $1 = q_1 < q_2 < \cdots$ of best approximations for $a_1, \ldots, a_n$ we have

$$\lim_{k \to \infty} \inf q_k^{1/k} \geq 1 + \frac{1}{2^{n+1}}.$$

We look at the growth of the denominators $q = \max_j |q_j|$ that are found by the ILLL algorithm. Dirichlet's Theorem 1.1 suggests that if q grows exponentially with a rate of m/n, then infinitely many approximations with Dirichlet coefficient smaller than 1 can be found. In the iterated LLL algorithm it is guaranteed by (6) that $q(k)$ is smaller than a constant times $d^{kn/m}$. Our experiments indicate that $q(k)$ is about $d^{kn/m}$, or equivalently that $e^{(m \log q_k)/(kn)}$ is about d; see Figure 6, which gives a histogram of solutions that satisfy $e^{(m \log q_k)/(kn)} = x$.

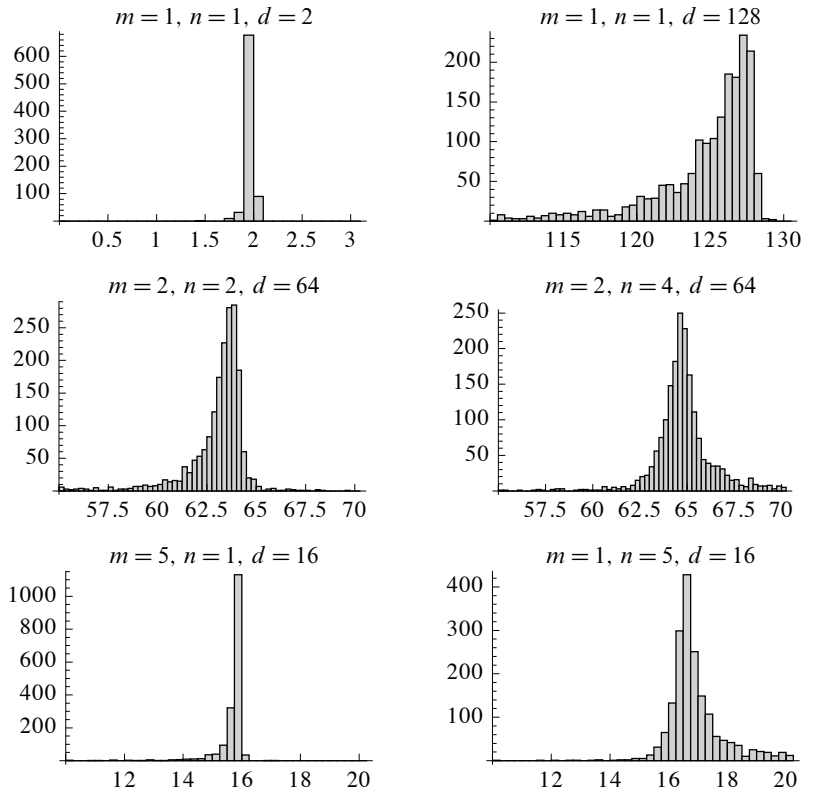


Figure 6. Histograms of $e^{(m \log q(k))/(kn)}$ for various values of m, n and d. In these experiments we used $q_{\max} = 10^{40}$ and repeated the ILLL algorithm $\lfloor 2000/k' \rfloor$ times, with k' from Lemma 3.3.

References

[1] Wieb Bosma, *Optimal continued fractions*, Nederl. Akad. Wetensch. Indag. Math. **49** (1987), no. 4, 353–379. http://dx.doi.org/10.1016/1385-7258(87)90001-1 MR 89b:40001

[2] Wieb Bosma and Cor Kraaikamp, *Metrical theory for optimal continued fractions*, J. Number Theory **34** (1990), no. 3, 251–270. MR 91d:11095

[3] _____, *Optimal approximation by continued fractions*, J. Austral. Math. Soc. Ser. A **50** (1991), no. 3, 481–504. MR 92f:11093

[4] A. J. Brentjes, *Multidimensional continued fraction algorithms*, Mathematical Centre Tracts, no. 145, Mathematisch Centrum, Amsterdam, 1981. http://tinyurl.com/brentjes145 MR 83b:10038

[5] Viggo Brun, *En generalisation av kjedebrøken, I*, Skr. Vidensk. Selsk. Kristiania I **1919** (1919), no. 6, 1–29, with an abstract in French. http://archive.org/stream/skrifterutgitavv1917chri#page/n764/ JFM 47.0168.01

[6] _____, *En generalisation av kjedebrøken, II*, Skr. Vidensk. Selsk. Kristiania I **1920** (1920), no. 6, 1–24, with an abstract in French. http://archive.org/stream/skrifterutgitavv201chri#page/n460/ JFM 47.0168.01

[7] Karma Dajani and Cor Kraaikamp, *Ergodic theory of numbers*, Carus Mathematical Monographs, no. 29, Mathematical Association of America, Washington, DC, 2002. MR 2003f:37014

[8] H. R. P. Ferguson and R. W. Forcade, *Generalization of the Euclidean algorithm for real numbers to all dimensions higher than two*, Bull. Amer. Math. Soc. (N.S.) **1** (1979), no. 6, 912–914. MR 80i:10039

[9] A. Hurwitz, *Ueber die angenäherte Darstellung der Zahlen durch rationale Brüche*, Math. Ann. **44** (1894), no. 2-3, 417–436. MR 1510845

[10] C. G. J. Jacobi, *Allgemeine Theorie der kettenbruchähnlichen Algorithmen, in welchen jede Zahl aus* drei *vorhergehenden gebildet wird*, J. Reine Angew. Math. **69** (1868), 29–64.

[11] Bettina Just, *Generalizing the continued fraction algorithm to arbitrary dimensions*, SIAM J. Comput. **21** (1992), no. 5, 909–926. MR 93k:11065

[12] J. C. Lagarias, *Best simultaneous Diophantine approximations. I. Growth rates of best approximation denominators*, Trans. Amer. Math. Soc. **272** (1982), no. 2, 545–554. MR 84d:10039a

[13] ———, *The computational complexity of simultaneous Diophantine approximation problems*, SIAM J. Comput. **14** (1985), no. 1, 196–209. MR 86m:11048

[14] ———, *Geodesic multidimensional continued fractions*, Proc. London Math. Soc. (3) **69** (1994), no. 3, 464–488. MR 95j:11066

[15] A. M. Legendre, *Essai sur la théorie des nombres*, Chez Duprat, Paris, 1798. http://gallica.bnf.fr/ark:/12148/btv1b8626880r

[16] G. Lejeune Dirichlet, *Verallgemeinerung eines Satzes aus der Lehre von den Kettenbrüchen nebst einige Anwendungen auf die Theorie der Zahlen*, Verh. Kön. Preuss. Akad. Wiss. (1842), 93–95, reprinted in *Mathematische Werke*, I, edited by L. Kronecker (G. Reimer, Berlin, 1889, and Chelsea, Bronx, NY, 1969, two volumes in one), pp. 635–638. http://books.google.com/books?id=cihJAAAAcAAJ&pg=PA93

[17] A. K. Lenstra, H. W. Lenstra, Jr., and L. Lovász, *Factoring polynomials with rational coefficients*, Math. Ann. **261** (1982), no. 4, 515–534. MR 84a:12002

[18] Oskar Perron, *Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus*, Math. Ann. **64** (1907), no. 1, 1–76. MR 1511422

[19] Wolfgang M. Schmidt, *Diophantine approximation*, Lecture Notes in Mathematics, no. 785, Springer, Berlin, 1980. MR 81j:10038

[20] Fritz Schweiger, *Multidimensional continued fractions*, Oxford University Press, Oxford, 2000. MR 2005i:11090

[21] Ionica Smeets, *On continued fraction algorithms*, Ph.D. thesis, Leiden University, 2010. http://www.math.leidenuniv.nl/scripties/SmeetsThesis.pdf

WIEB BOSMA: bosma@math.ru.nl
Mathematics Department, Radboud University Nijmegen, PO Box 9010, 6500 Nijmegen, The Netherlands

IONICA SMEETS: ionica.smeets@gmail.com
Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.187

msp

Success and challenges in determining the rational points on curves

Nils Bruin

We give an overview of current computational methods for determining the rational points on algebraic curves. We discuss how two methods, based on embedding a curve in an abelian variety, provide a practical method for deciding whether the curve has rational points and, if some additional technical condition is met, for the determination of these points.

While we cannot prove the methods are always successful, we do have a heuristic that makes us expect so. This means that the main problem becomes the determination of rational points on abelian varieties, in particular the determination of the free rank of the finitely generated group they form. We discuss some methods that provide bounds on this rank.

Finally, we report on some recent progress on applying these methods to non-hyperelliptic curves of genus 3.

1. Introduction

This article is an extended abstract from an invited lecture delivered on July 13, 2012, as part of the Tenth Algorithmic Number Theory Symposium (ANTS X), at the University of California, San Diego. It discusses current computational methods for determining the rational points on algebraic curves. Two methods, *Mordell-Weil sieving* (see Section 4) and *Chabauty's method* (see Section 5) together provide a procedure that often decides whether a curve has any rational points and, if so, determines them. While we cannot prove that these methods will always succeed, we do have some heuristics that indicate that this is quite likely.

Both methods rely on embedding a curve in an abelian variety J and on having a rather detailed description of the rational points on J. There is presently no proven

MSC2010: primary 11G30; secondary 11G10, 14G25, 14H45.
Keywords: Selmer group, descent, Mordell-Weil sieving, rational points, curves, Chabauty, coverings.

algorithm for determining the rational points on an abelian variety, but here too we have methods that frequently work in practice. In fact, if Tate-Shafarevich groups are finite, as they are conjectured to be, then it would theoretically be possible to compute the rational points on an abelian variety.

The main point of this article is that the computational bottleneck for determining rational points on curves presently lies in the determination of rational points on abelian varieties. Our main tool is the computation of Selmer sets via finite descent.

After reviewing the Mordell-Weil sieve and Chabauty's method in Sections 4 and 5, we give a brief description in Section 7 of recent joint work [14] with Bjorn Poonen and Michael Stoll to provide a description of descent computations which, to our knowledge, encompasses all previous methods for doing such computations for curves.

We note in Section 8 that descent methods also help in deciding whether a curve C can be embedded in its Jacobian, a requirement for the curve to have rational points and for the application of the Mordell-Weil sieve and Chabauty's method. A good description of Selmer groups also helps in constructing *covering collections*, which can be used to transform problems where Chabauty's method does not apply into problems where it may.

The most difficult ingredient in descent computations usually is the determination of unit groups and ideal class groups of number fields. Especially for number fields of larger degrees, this can be extremely challenging. In Section 7 we describe some ways one can reduce the maximal degree to be considered: from 63 to 28 in the case of smooth plane quartic curves. This has allowed us to perform the required calculations for some genus-3 curves. To our knowledge, these are the first examples of curves with simple Jacobians and trivial automorphism groups to which the methods have been successfully applied. Previous applications made essential use of decompositions of the Jacobian or of the automorphisms to get descriptions more favorable to computation.

Since in general curves have trivial automorphism groups, we believe these examples present evidence that these methods are indeed quite generally applicable, although the computational challenges can be daunting.

We cannot hope to give an exhaustive account of the subject here. Instead, we intend to provide the reader with a bit of insight into how the different methods interact and what the fundamental ideas and problems are. We have also included ample literature references for further reading.

2. Statement of the problems

Consider the equation

$$x^4 + y^4 + x^2 y + 2xy - y^2 + 1 = 0. \tag{1}$$

Can you determine the solutions $x, y \in \mathbb{Q}$ to this equation? Can you determine whether this equation has any rational solutions at all? These are questions about *rational points on curves*, and such questions are about as old as mathematics itself. (See Proposition 9.3 for results on this particular equation.)

We concern ourselves with curves C defined over $\mathbb{Q}$, and we want to study the set of rational points $C(\mathbb{Q})$. Every curve has a projective closure, which has at most finitely many additional rational points. Furthermore, every curve admits a morphism from a nonsingular curve that is an isomorphism outside the finitely many singularities, which are easily determined and tested for rationality. We can therefore restrict our attention to nonsingular, absolutely irreducible, projective curves.

The reader does not lose much, and may gain a more concrete conception, by thinking of C as a smooth plane curve such as the projective closure of the curve defined by Equation (1). Although much of what we discuss holds with suitable modifications over arbitrary number fields, we will limit ourselves to $\mathbb{Q}$ for the sake of concreteness and ease of notation.

A common theme in arithmetic geometry is that *geometry* determines *arithmetic*: The geometric classification of curves C has deep ramifications for the structure of $C(\mathbb{Q})$. There are:

- *Curves of genus* 0. These are always isomorphic to plane conics. Either such a curve C has no rational points at all, or C admits a parametrization $\phi\colon \mathbb{P}^1 \to C$, providing an explicit bijection between $\mathbb{P}^1(\mathbb{Q})$ and $C(\mathbb{Q})$.
- *Curves of genus* 1. If C has any rational points, then C is isomorphic to an elliptic curve. In that case, Mordell's Theorem [46] implies that $C(\mathbb{Q})$ can be described as a finitely generated abelian group.
- *Curves of general type* (genus at least 2). Faltings's Theorem [28] states that $C(\mathbb{Q})$ is a finite set.

We concentrate on two explicit questions.

Decision Problem. Given a curve C over $\mathbb{Q}$, decide if $C(\mathbb{Q}) = \varnothing$.

Determination Problem. Given a curve C over $\mathbb{Q}$, give an explicit description of $C(\mathbb{Q})$.

We assume that the curve is given to us in a sufficiently explicit way, for instance by explicit equations like Equation (1). For genus-0 curves, both questions have a reasonably satisfactory solution [44, pp. 512–513] (and see [58] for a modern algorithmic perspective). For genus-1 curves, a satisfactory answer to the determination problem is usually considered to be an explicit listing of a finite set of generators of $C(\mathbb{Q})$ equipped with its group structure. We are primarily interested in curves

of general type. For those curves the set $C(\mathbb{Q})$ is finite, so an explicit listing of the set would provide a satisfactory solution to the determination problem.

As we discuss in Section 4, the most important step is to realize C as a subvariety of an abelian variety J. If we take J to be the *Jacobian* of C then a rational point on C gives rise to such an embedding. If we can prove no such embedding exists, then we can conclude that $C(\mathbb{Q})$ is empty.

Challenge A. Given a curve C over $\mathbb{Q}$ of positive genus, determine an embedding of C into its Jacobian or prove no such embedding exists.

The main advantage of considering C as a subvariety of an abelian variety J, rather than of a rational space such as $\mathbb{P}^2$, is that the set of rational points of J is much sparser: The Mordell-Weil Theorem [63] states that $J(\mathbb{Q})$ is a finitely generated group. We can use knowledge about $J(\mathbb{Q})$ to obtain information about $C(\mathbb{Q})$. This leads to our second challenge.

Challenge B. Given a curve C of positive genus, determine $J(\mathbb{Q})$, where J is the Jacobian of C.

Note that if C is of genus 1, then an embedding as in Challenge A establishes an isomorphism between C and J, so Challenge B provides a solution to the determination problem. In the remainder of this text we take C to be a curve of general type.

A major component in determining $J(\mathbb{Q})$ is determining the rank of its free part. A conjectural link suggested by Birch and Swinnerton-Dyer [4] for elliptic curves connects this rank to the vanishing of an L-function at a special point. For elliptic curves over $\mathbb{Q}$ with an L-function that vanishes to order at most 1, this is now proved [38; 43], but for more general abelian varieties even the existence of the function at the point is not generally established.

The only general unconditional approach uses *descent* to provide a hopefully sharp upper bound on the rank. The ideas are most easily explained in the language of Galois cohomology (see Section 6).

Once a bound on the rank is determined, one can try to prove that the bound is sharp by exhibiting sufficiently many independent points on J. Finding them is only a computational problem. Since these points can be drawn from an obviously enumerable set of candidates, generators will eventually be found. Finding generators *efficiently* is a serious computational problem, but we will ignore it here.

The traditional way of showing that a set generates all of $J(\mathbb{Q})$ is by computing *canonical heights*. However, a good algorithm for computing canonical heights efficiently is only available for curves of genus up to 2; see [31; 35; 60; 61]. For our purposes, one only needs a subgroup of $J(\mathbb{Q})$, of finite index prime to some predetermined number B. Proving that a set generates such a group is usually much easier to establish; see Remark 4.6.

Since (sharply) bounding the rank of $J(\mathbb{Q})$ is a crucial step for the methods in Sections 4 and 5, we describe in Section 7 a way to actually compute or approximate the rather abstract objects introduced in Section 6. While one can concentrate on the geometry of J (see [3; 37]), this becomes unwieldy for more complicated J. Another approach emphasizes that J represents the group $\mathrm{Pic}^0(C)$ of degree-0 divisor classes on C and tries to express as much of the data as possible in terms of objects directly related to the curve [18; 21; 23; 48; 49; 53; 54; 56]. We closely follow the exposition in [14].

In Section 8 we describe how the constructions in Section 7 can also be used to attack some related problems, and in Section 9 we give some examples, taken from [14], of successful applications of these methods to smooth plane quartic curves. To our knowledge, these are the first examples fully carried out for curves with trivial automorphism groups. Previous applications all made essential use of nontrivial automorphisms to simplify computations. The fact that these procedures are also shown to be practical when no such automorphisms are available is a hopeful sign that they are applicable in generality.

3. Local considerations

Let C be a curve over $\mathbb{Q}$ and let $K \supseteq \mathbb{Q}$ be a field extension. Then $C(\mathbb{Q}) \subseteq C(K)$. Hence, if C has a $\mathbb{Q}$-rational point then $C(\mathbb{R}) \neq \varnothing$ and $C(\mathbb{Q}_p) \neq \varnothing$ for all primes p.

We introduce some notation to express this observation more concisely. We call $\mathbb{R}$ the completion of $\mathbb{Q}$ at the *infinite* prime and write $\mathbb{R} = \mathbb{Q}_\infty$. We write

$$\Omega_{\mathbb{Q}} = \{p \in \mathbb{Z}_{>1} : p \text{ is prime}\} \cup \{\infty\}.$$

The consideration of all completions of $\mathbb{Q}$ at once leads to the ring $\mathbb{A}$ of *adèles*. We will only use it here as a concise piece of opaque notation and define for a projective curve C the set

$$C(\mathbb{A}) := \prod_{v \in \Omega_{\mathbb{Q}}} C(\mathbb{Q}_v).$$

The observation above now translates to

$$C(\mathbb{Q}) \neq \varnothing \quad \text{implies} \quad C(\mathbb{A}) \neq \varnothing. \tag{2}$$

Fact 3.1. *One can decide algorithmically whether $C(\mathbb{A}) = \varnothing$.*

Determining whether $C(\mathbb{R}) = \varnothing$ is a straightforward application of calculus and the intermediate value theorem. Determining whether $C(\mathbb{Q}_p) = \varnothing$ is also computable thanks to Hensel's lifting criterion (see [10] for a collection of algorithms). Furthermore, for all but a finite and explicitly computable set of primes p we can immediately conclude that $C(\mathbb{Q}_p)$ is nonempty.

The implication (2) is mainly useful for its contrapositive: if we can show that $C(\mathbb{A})$ is empty (that is, that $C(\mathbb{Q}_v)$ is empty for some v) then we can conclude that $C(\mathbb{Q})$ is empty. The converse of implication (2), known as the *local-global* principle, is known to hold for genus-0 curves. Hence, if C is a genus-0 curve and $C(\mathbb{A}) \neq \varnothing$ then C has a rational point.

However, for curves of positive genus the local-global principle is known to fail. For instance, for curves of genus 2 over $\mathbb{Q}$, one can prove that the subset of curves C with $C(\mathbb{A}) \neq \varnothing$ has asymptotic density about 0.85, measured with respect to an appropriate height [50]. However, one would expect the set of curves with a rational point to have asymptotic density 0 — see for instance [52, Conjecture 2.2(i)] for a formal statement of this folklore conjecture in the case of plane curves — so many curves with points everywhere locally should have no rational points at all.

4. The Mordell-Weil sieve

Let C be a smooth projective curve of genus $g \geq 2$. In this section we discuss a method that allows us to obtain significant information on $C(\mathbb{Q})$ by considering an embedding of C into an abelian variety J (usually its Jacobian) for which we can determine $J(\mathbb{Q})$. We write $\iota: C \to J$ for the embedding.

The rational points on an abelian variety are sufficiently sparse that the topological closure $\overline{J(\mathbb{Q})} \subset J(\mathbb{A})$ is significantly smaller than $J(\mathbb{A})$. We observe that

$$C(\mathbb{Q}) \subset C(\mathbb{A}) \cap \overline{J(\mathbb{Q})}.$$

The latter set is amenable to computation, or at least to approximation. As it turns out, the small step of taking into consideration a little bit of extra global data, in the form of $\overline{J(\mathbb{Q})}$, provides considerable extra information.

In [57], Scharaschkin presents the method and shows, subject to the standard conjecture that $Ш(J/\mathbb{Q})$ is finite, that the obstruction to the existence of rational points on C that this method exhibits can be interpreted in terms of the *Brauer-Manin obstruction* [59]. See [12; 33; 49] for applications and [15] for a larger scale experiment. Details are provided in [17], including an optimal strategy for avoiding a combinatorial explosion to which this method is prone. See also [20] for an application of to determining integral points on curves.

Let p be a prime of good reduction of the embedding $\iota: C \to J$, meaning that there are smooth proper models $\mathscr{C}$ and $\mathscr{J}$ over $\mathbb{Z}_p$ of C and J, respectively, and a morphism $\iota': \mathscr{C} \to \mathscr{J}$ that restricts to ι on the generic fiber. (The conditions on the type of reduction can be significantly relaxed.) We write $C(\mathbb{F}_p) = \mathscr{C}(\mathbb{F}_p)$ and $J(\mathbb{F}_p) = \mathscr{J}(\mathbb{F}_p)$. We use that $J(\mathbb{Q}_p) = \mathscr{J}(\mathbb{Z}_p)$ and write $\rho_p: J(\mathbb{Q}) \to J(\mathbb{F}_p)$ for the induced reduction map. Via the same principle we obtain a reduction map

$C(\mathbb{Q}) \to C(\mathbb{F}_p)$. Furthermore, we write $\iota_p : C(\mathbb{F}_p) \to J(\mathbb{F}_p)$ for the map that ι' induces on the rational points of the reductions.

Let us fix a finite set S of primes of good reduction of J and a positive integer B. We consider the commutative diagram

$$\begin{array}{ccc} C(\mathbb{Q}) & \xrightarrow{\quad\iota\quad} & \dfrac{J(\mathbb{Q})}{BJ(\mathbb{Q})} \\ \downarrow & & \downarrow \rho_S \\ \displaystyle\prod_{p\in S} C(\mathbb{F}_p) & \xrightarrow[\iota_S]{} & \displaystyle\prod_{p\in S} \frac{J(\mathbb{F}_p)}{B \operatorname{im} \rho_p}, \end{array}$$

where ρ_S and ι_S are the obvious maps induced by $\{\rho_p : p \in S\}$ and $\{\iota_p : p \in S\}$ respectively.

Each of the four sets in this diagram is finite, so determining

$$V_{S,B} = \operatorname{im} \rho_S \cap \operatorname{im} \iota_S$$

is a matter of combinatorics. For sufficiently large B and S, the map $\rho_S \circ \iota$ will be an injection, so in that case the size of $V_{S,B}$ provides an upper bound on the size of $C(\mathbb{Q})$. In any case, if $V_{S,B}$ is empty, then C has no rational points.

If the domains of ρ_S and ι_S are sufficiently small relative to their codomain, one would expect the intersection of their images to be rather small. One can formulate a reasonable heuristic argument that supports this.

Heuristic 4.1 (Poonen [47]). Subject to plausible assumptions that $\operatorname{im} \iota_S$ and $\operatorname{im} \rho_S$ behave in a way that can be suitably modeled by a random process, one expects that for suitably chosen B and S, the set $V_{S,B}$ consists only of images of $C(\mathbb{Q})$.

While ρ_S and ι_S are maps between finite sets, both B, S have to be quite large in practice for Heuristic 4.1 to apply. So, while $V_{S,B}$ is likely a very small set, it tends to be an intersection of two rather large sets. For practical computations, one has to take some care in constructing the set via appropriate steps. See [17] for some strategies for doing so.

We are left with finding an appropriate embedding $\iota : C \to J$ into an abelian variety. A canonical choice for J is the *Jacobian* of C. It is a g-dimensional abelian variety representing the degree-0 divisor classes on C; that is, $J(\overline{\mathbb{Q}}) = \operatorname{Pic}^0(C/\overline{\mathbb{Q}})$. This equality is Galois-equivariant, so $J(\mathbb{Q})$ consists of the Galois-invariant divisor classes $\operatorname{Pic}^0(C/\overline{\mathbb{Q}})^{\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}$. The latter can be strictly larger than $\operatorname{Pic}^0(C/\mathbb{Q})$, the set of linear equivalence classes that contain divisors that are defined over $\mathbb{Q}$. However, for the problem at hand, this is not an issue (see [7], for instance, for some related theory).

Lemma 4.2 (Standard result). *Let C be a curve over a field k, where k is either a finite field or a number field such that $C(k_v)$ is nonempty for all places v of k. Then every Galois-invariant divisor class on C contains a divisor defined over k.*

For our applications, if $C(\mathbb{Q}_v) = \varnothing$ for any place $v \in \Omega_{\mathbb{Q}}$, the results in Section 3 already imply that $C(\mathbb{Q}) = \varnothing$, so we only need to work with $J(\mathbb{Q})$ when we can represent its points by divisors over $\mathbb{Q}$. This allows us to avoid constructing a projective model for J as a variety.

A point on a curve C gives rise to a degree-1 divisor class. Since on a curve of positive genus no two such divisors are linearly equivalent, we obtain an injection $C(\mathbb{Q}) \to \mathrm{Pic}^1(C/\mathbb{Q})$. Similarly to how J is a variety that represents Pic^0, there is also a variety $\underline{\mathrm{Pic}}^1(C)$, that represents Pic^1. Indeed, there is a natural morphism $C \to \underline{\mathrm{Pic}}^1(C)$. There is a natural action of J on $\underline{\mathrm{Pic}}^1(C)$, corresponding to addition of divisor classes, that equips $\underline{\mathrm{Pic}}^1(C)$ with the structure of a $\mathbb{Q}$-torsor under J. A rational point on $\underline{\mathrm{Pic}}^1(C)$ induces an isomorphism between J and $\underline{\mathrm{Pic}}^1(C)$. If there is no such point, then C has no degree 1 divisors and hence certainly no rational points. Therefore, a reformulation of Challenge A is:

Challenge A′. Given a curve C over $\mathbb{Q}$ of positive genus, determine a divisor class $\mathfrak{d} \in \mathrm{Pic}^1(C/\mathbb{Q})$ or prove no such divisor class exists.

If $\mathfrak{d}$ exists then the map $\iota\colon C \to J$ it induces corresponds to

$$\begin{aligned} C(\mathbb{Q}) &\longrightarrow \mathrm{Pic}^0(C/\mathbb{Q}) \\ P &\longmapsto [P] - \mathfrak{d}. \end{aligned}$$

For an appropriate reduction $\mathfrak{d}_p$ modulo p, we get the corresponding map

$$C(\mathbb{F}_p) \to \mathrm{Pic}^0(C/\mathbb{F}_p)$$

given by $P \mapsto [P] - \mathfrak{d}_p$. This suggests the procedure below for solving the decision problem. First note that a choice of smooth projective model for C also provides us with an explicitly enumerable set containing $C(\mathbb{Q})$ — namely, $\mathbb{P}^n(\mathbb{Q})$ — so if C has a rational point we can find it in finite time by enumeration (but see Remark 4.5 for drastic improvements).

Remark 4.3. We use the term *algorithm* in the strict sense: a Turing machine or an equivalent computing device that is guaranteed to produce a correct answer in finite time when given correct input. We use the word *procedure* for a less formal concept than an algorithm. We allow a procedure to include steps that are not guaranteed to succeed, and we do not require that a procedure will stop for all valid input. We do require the guarantee that *if* a procedure finishes then its output is correct.

Procedure 4.4 (Decision procedure).

Input: A curve C over $\mathbb{Q}$ (or more generally, a number field).

Output: A rational point on C or a proof that there is none.

First parallel thread:

0. Enumerate candidates for $C(\mathbb{Q})$. If a point is found, we have shown that $C(\mathbb{Q})$ is not empty.

Second parallel thread:

1. Test if $C(\mathbb{A}) = \varnothing$. If that is the case then $C(\mathbb{Q})$ is empty too. See Fact 3.1.
2. (Challenge A′) Find $\mathfrak{d} \in \mathrm{Pic}^1(C/\mathbb{Q})$ or prove it doesn't exist. We either obtain an embedding $\iota\colon C \to J$ or we prove that $C(\mathbb{Q})$ is empty.
3. (Challenge B) Find a finite set of generators for $J(\mathbb{Q})$.
4. Choose appropriate S and B.
5. Compute $V_{S,B}$. This involves computing $J(\mathbb{F}_p)$, using for instance [39; 42].
6. If $V_{S,B} = \varnothing$ then $C(\mathbb{Q})$ is empty. Otherwise, increase S and B and go to step 5.

Remark 4.5. Once we have determined generators for $J(\mathbb{Q})$, we can enumerate candidates for $C(\mathbb{Q})$ much more efficiently by enumerating $J(\mathbb{Q})$. Furthermore, the set $V_{S,B}$ provides us with a list of cosets modulo $BJ(\mathbb{Q})$ that may contain elements of $C(\mathbb{Q})$, further reducing the number of candidates to consider. This makes it feasible to search up to height bounds that are doubly exponential in time. See [20] for an application to finding *integral* points on curves.

We do not have a proof that this procedure will always terminate, but Heuristic 4.1 suggests it should. Indeed, in [15] we describe an experiment where we test how well the decision procedure works in practice. We consider genus-2 curves admitting models of the form

$$y^2 = f_6x^6 + f_5x^5 + \cdots + f_0 \quad \text{with} \quad f_0, \ldots, f_6 \in \{-3, -2, \ldots, 3\}.$$

For nearly all the roughly 200,000 isomorphism classes represented, we were able to solve the decision problem. For 42 curves we were unable to unconditionally complete step 2. For those we obtained a presumably accurate bound on the rank of $J(\mathbb{Q})$ by assuming the Birch and Swinnerton-Dyer conjecture. The Mordell-Weil sieving itself never posed an insurmountable problem.

The main practical problem with the procedure above is that if either of steps 2 or 3 fails, we have no way of continuing. We can weaken the requirement for step 3 slightly.

Remark 4.6. We only need a set of elements in $J(\mathbb{Q})$ that generate $J(\mathbb{Q})/BJ(\mathbb{Q})$, so a subgroup of finite index prime to B in $J(\mathbb{Q})$ would already be enough. If

one knows the rank of $J(\mathbb{Q})$ then one can usually quickly deduce that a given set generates such a group by considering its image under

$$J(\mathbb{Q}) \longrightarrow \prod_{p\in S} J(\mathbb{F}_p).$$

for some suitable set of primes S. For instance, let q be a prime dividing B. If we know that $J(\mathbb{Q})/qJ(\mathbb{Q}) \simeq (\mathbb{Z}/q\mathbb{Z})^t$ and the codomain has a direct factor of the form $\prod_{i=1}^{t}(\mathbb{Z}/q^{e_i}\mathbb{Z})$, with $e_1, \ldots, e_t \geq 1$, onto which the group generated by our given set surjects, then the set generates a subgroup of finite index prime to q.

5. Isolating rational points: Chabauty's method

While Mordell-Weil sieving can provide a proof that $C(\mathbb{Q})$ is empty, it will not prove that $C(\mathbb{Q})$ is finite, let alone determine $C(\mathbb{Q})$, if there is a rational point on C. Yet for large enough B and S the map $C(\mathbb{Q}) \to V_{S,B}$ is injective, and Heuristic 4.1 predicts that for suitable values of B and S it is surjective as well. Thus, given a rational point $P \in C(\mathbb{Q})$, we mainly need a way to prove the equality

$$\iota C(\mathbb{Q}) \cap (\iota(P) + BJ(\mathbb{Q})) = \{\iota(P)\}. \tag{3}$$

Inspired by Skolem's ideas for subvarieties of multiplicative groups, Chabauty [24] observed that one can construct a nonzero p-adic analytic function

$$\Theta_p \colon J(\mathbb{Q}_p) \longrightarrow \mathbb{Q}_p$$

that vanishes on $J(\mathbb{Q})$, provided that the rank r of $J(\mathbb{Q})$ is strictly smaller than the dimension g of J. (Actually, he observed that one can construct such functions locally, and gets the desired result by doing so on a finite open covering of the rational points.) The fact that analytic functions have isolated zeros allows one to conclude that C has only finitely many rational points and, with a bit of extra work, to establish statements like equality (3). See [26] for one of the first modern treatments of the method and [23], [32], and [34] for a flexible way of applying it.

In order to avoid some technical complications, we take a prime p at which C has good reduction. We write $J^{(1)}(\mathbb{Q}_p)$ for the kernel of the reduction homomorphism $J(\mathbb{Q}_p) \to J(\mathbb{F}_p)$ and we write $\Lambda_p = J(\mathbb{Q}) \cap J^{(1)}(\mathbb{Q}_p)$ for the part of the Mordell-Weil group that lies in the kernel of reduction.

The function Θ_p in question arises from the p-adic integration of a regular differential ω. We consider regular differentials obtained by lifting a regular differential $\overline{\omega}$ on C over $\mathbb{F}_p$, so our differentials have *good reduction* at p as well. We sketch the details here.

Let $P \in C(\mathbb{Q}_p)$. We choose a uniformizer $\bar{t} \in \mathbb{F}_p(C)$ at the reduction $\overline{P} \in C(\mathbb{F}_p)$ of P and lift it to a uniformizer $t \in \mathbb{Q}_p(C)$ at P. Let ω be a regular differential on C with good reduction as described above. We have $\omega = h\,dt$ for some

function $h \in \mathbb{Q}_p(C)$ regular at P. Localization at P provides a homomorphism $\mathbb{Q}_p(C) \to \mathbb{Q}_p((t))$. Regularity and good reduction imply that when we identify h with its image, we have $h(t) \in \mathbb{Z}_p[[t]]$. We can compute a formal power series

$$\int_{t=0}^{z} h(t)dt \in \mathbb{Q}_p[[z]],$$

and it is straightforward to check that its radius of convergence is at least 1. Let $\overline{\mathbb{Q}}_p$ be an algebraic closure of $\mathbb{Q}_p$, and extend the p-adic absolute value in the natural way to $\overline{\mathbb{Q}}_p$. For any point $Q \in C(\overline{\mathbb{Q}}_p)$ that reduces to $\overline{P} \in C(\mathbb{F}_p)$, we have $|t(Q)|_p < 1$. Hence we can define the integral of ω from P to Q by the formula

$$\int_P^Q \omega = \int_0^{t(Q)} h(t)dt,$$

which is easily checked to not depend on the choice of t. Note that every divisor class in $J^{(1)}(\mathbb{Q}_p)$ admits a representative of the form

$$[Q_1 + \cdots + Q_g - gP],$$

where each $Q_i \in C(\overline{\mathbb{Q}}_p)$ reduces to $\overline{P} \in C(\overline{\mathbb{F}}_p)$. We define the integral of ω over this divisor class by

$$\int_{[Q_1+\cdots+Q_g-gP]} \omega = \sum_{i=1}^{g} \int_P^{Q_i} \omega.$$

One can check that the regularity of ω implies that this provides a well-defined group homomorphism $J^1(\mathbb{Q}_p) \to \mathbb{Q}_p$.

Let $\overline{\omega}_1, \ldots, \overline{\omega}_g$ be a basis of the space of regular differentials of the reduction of C at p and let $\omega_1, \ldots, \omega_g$ be a lift of that basis. We have a $\mathbb{Z}_p$-bilinear pairing

$$J^{(1)}(\mathbb{Q}_p) \times (\mathbb{Z}_p)^g \to \mathbb{Q}_p$$

taking $\big(D, (\lambda_1, \ldots, \lambda_g)\big)$ to

$$\int_D \lambda_1\omega_1 + \cdots + \lambda_g\omega_g.$$

We see that if the $\mathbb{Z}$-rank r of $J(\mathbb{Q})$ is strictly less than g, then the $\mathbb{Z}_p$-submodule generated by $\Lambda_p \subset J^{(1)}(\mathbb{Q}_p)$ has $\mathbb{Z}_p$-rank at most $r < g$, so there is a nonzero differential ω_p such that

$$\int_D \omega_p = 0 \quad \text{for all } D \in \Lambda_p = J(\mathbb{Q}) \cap J^{(1)}(\mathbb{Q}_p).$$

In particular, for a rational point $P \in C(\mathbb{Q})$, we can define

$$\Theta_{p,P}(Q) = \int_P^Q \omega_p \quad \text{for } Q \in C(\mathbb{Q}_p) \text{ that reduce to } \overline{P} \in C(\mathbb{F}_p).$$

It follows that $\Theta_{p,P}(Q) = 0$ for every $Q \in C(\mathbb{Q})$ with the same reduction as P modulo p. The following is straightforward to prove by applying Hensel's lemma to the appropriate power series expansion.

Proposition 5.1 [26, proof of Theorem 4]. *If $P \in C(\mathbb{Q})$ and the reduction $\overline{\omega}_p$ is nonzero at $\rho_p(P) \in C(\mathbb{F}_p)$, then we have*

$$\iota C(\mathbb{Q}) \cap (\iota(P) + \Lambda_p) = \{\iota(P)\}.$$

We obtain the following procedure (see Remark 4.3 for the technical meaning of this word).

Procedure 5.2 (Determination procedure).

Input: A curve C of genus $g > 1$ with $J(\mathbb{Q})$ of free rank $r < g$.

Output: The elements of $C(\mathbb{Q})$.

1. Choose S and B, and search for points $\{P_1, \ldots, P_k\} \subset C(\mathbb{Q})$ such that
$$\{P_1, \ldots, P_k\} + BJ(\mathbb{Q}) = V_{S,B} + BJ(\mathbb{Q}).$$
2. For each point P_i, find a prime p such that $BJ(\mathbb{Q}) \subset \Lambda_p$ and $\overline{\omega}_p(\overline{P}) \neq 0 \in \mathbb{F}_p$. If this succeeds, you have proved that
$$C(\mathbb{Q}) = \{P_1, \ldots, P_k\}.$$
3. If step 2 fails, go to step 1 and choose larger S and B.

Remark 5.3. The linearity of the integration pairing in the first component implies that for any $D \in J^{(1)}(\mathbb{Q}_p)$ and $m \in \mathbb{Z}$ we have that

$$\int_{mD} \omega = m \int_D \omega.$$

Since $J^{(1)}(\mathbb{Q}_p) \subset J(\mathbb{Q}_p)$ is of finite index, say index m, we have for any $D \in J(\mathbb{Q}_p)$ that $mD \in J^{(1)}(\mathbb{Q}_p)$, so we can use this identity to extend the integration pairing to all of $J(\mathbb{Q}_p)$. This provides a rigid analytic continuation of $\Theta_{p,P}$ to all of $C(\mathbb{Q}_p)$ that vanishes at $C(\mathbb{Q})$ — see also [2].

We cannot prove that step 1 of the determination procedure will succeed, but Heuristic 4.1 suggests it should. We cannot prove that step 2 will succeed eventually either, but given that $\overline{\omega}_p(\rho_p(P)) = 0$ requires the vanishing of a power series coefficient in $\mathbb{F}_p$, we expect that this happens only one in p cases on average. Indeed, in practice finding an appropriate p in step 2 never seems to be a problem.

Combining Mordell-Weil sieving with Chabauty's method yields the significant benefit that larger residue characteristics pose no problem. Results typical for Chabauty's method by itself bound $\#C(\mathbb{Q})$ in terms of $\#C(\mathbb{F}_p)$, and these bounds are rarely sharp (see [26] and [62]).

A more significant restriction is that the procedure is not guaranteed to apply at all if $r \geq g$. One remedy is to use *covers*. One determines a finite set of covers $\phi_i \colon D_i \to C$ with $i = 1, \ldots, m$ and where the D_i are curves of genus larger than g, such that

$$C(\mathbb{Q}) = \bigcup_{i=1}^{m} \phi_i(D_i(\mathbb{Q})),$$

in the hope that the determination procedure does apply to each of $D_1, \ldots, D_r$. In Section 8C we see how the ideas from Section 6, in particular Proposition 6.3, can be used to construct such covering sets.

6. Theory of finite descent

Let us first consider Challenge B, finding (a finite index subgroup of) the group $J(\mathbb{Q})$. The first observation is that $J^{(1)}(\mathbb{Q}_p)$ is torsion-free for $p > 2$ (see [41]), so the reduction map $J(\mathbb{Q}) \to J(\mathbb{F}_p)$ is injective on the torsion subgroup $J(\mathbb{Q})_{\text{tors}}$. As a consequence, by computing $J(\mathbb{F}_p)$ for a small number of primes p, which we have to do for Mordell-Weil sieving anyway, we easily obtain a bound on the size of $J(\mathbb{Q})_{\text{tors}}$. This bound is often sharp, so simply exhibiting enough torsion points usually suffices for determining $J(\mathbb{Q})_{\text{tors}}$.

More generally, the kernel of the multiplication-by-n morphism $J \to J$, denoted by $J[n]$, is 0-dimensional. Determining an approximation of it points over, say, $\mathbb{C}$, is straightforward. One can then recognize which of these torsion points are defined over $\mathbb{Q}$. Once $J(\mathbb{Q})_{\text{tors}}$ is obtained, we are left with determining the free part. The structure theorem for finitely generated abelian groups gives us that

$$J(\mathbb{Q}) \simeq J(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r \quad \text{and} \quad \frac{J(\mathbb{Q})}{nJ(\mathbb{Q})} \simeq \frac{J(\mathbb{Q})_{\text{tors}}}{nJ(\mathbb{Q})_{\text{tors}}} \times (\mathbb{Z}/n\mathbb{Z})^r.$$

That means that if we can compute the size of $J(\mathbb{Q})/nJ(\mathbb{Q})$, we can compute r.

Since the multiplication-by-n morphism $J \xrightarrow{n} J$ is surjective over algebraically closed fields, we have a short exact sequence of Galois modules

$$0 \longrightarrow J[n](\overline{\mathbb{Q}}) \longrightarrow J(\overline{\mathbb{Q}}) \xrightarrow{n} J(\overline{\mathbb{Q}}) \longrightarrow 0. \tag{4}$$

The abstract language of Galois cohomology allows us to derive a description of the set $J(\mathbb{Q})/nJ(\mathbb{Q})$ that facilitates a clean proof of the weak Mordell-Weil theorem. It also provides a road map for computing bounds on r. In this section we make a detour into this abstract world. In the next section we investigate how to compute some of the objects introduced here.

For a Galois module $M(\overline{\mathbb{Q}})$ we write $H^i(\mathbb{Q}, M) = H^i(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}), M(\overline{\mathbb{Q}}))$. Taking cohomology of the short exact sequence (4), we obtain the exact sequence

$$0 \longrightarrow \frac{J(\mathbb{Q})}{nJ(\mathbb{Q})} \xrightarrow{\gamma} H^1(\mathbb{Q}, J[n]) \longrightarrow H^1(\mathbb{Q}, J). \tag{5}$$

Thus, if we can bound the size of the image of the connecting homomorphism γ then a corresponding bound on r follows.

Indeed, we can consider the same sequence over localizations $\mathbb{Q}_v$ of $\mathbb{Q}$, and by identifying each $\mathrm{Gal}(\overline{\mathbb{Q}}_v/\mathbb{Q}_v)$ with a decomposition subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ we obtain the following commutative diagram:

$$\begin{array}{ccccc} 0 & \longrightarrow & \dfrac{J(\mathbb{Q})}{nJ(\mathbb{Q})} & \xrightarrow{\gamma} & H^1(\mathbb{Q}, J[n]) \\ & & \downarrow & & \downarrow{\scriptstyle \mathrm{res}_v} \\ 0 & \longrightarrow & \dfrac{J(\mathbb{Q}_v)}{nJ(\mathbb{Q}_v)} & \xrightarrow{\gamma_v} & H^1(\mathbb{Q}_v, J[n]). \end{array}$$

Since rational points are also $\mathbb{Q}_v$-rational, it follows that $\mathrm{im}\,\gamma$ lies in the *n-Selmer group* of J, defined by

$$\mathrm{Sel}^n(J/\mathbb{Q}) = \left\{\delta \in H^1(\mathbb{Q}, J[n]) : \mathrm{res}_v(\delta) \in \mathrm{im}\,\gamma_v \text{ for all } v \in \Omega_{\mathbb{Q}}\right\}.$$

Part of the proof that $J(\mathbb{Q})$ is finitely generated is establishing that $\mathrm{Sel}^n(J/\mathbb{Q})$ is finite, which is known as the *weak Mordell-Weil theorem*. This fact follows from another interpretation of the set $H^1(\mathbb{Q}, J[n])$, which also has computational significance. Some technical language is required to properly formulate this interpretation.

Let k be a field with separable closure $\bar{k}$, let M be a finite group with a $\mathrm{Gal}(\bar{k}/k)$-action and let X and Y be k-varieties. By limiting ourselves here to a *finite* group M, we guarantee that M can be represented by an affine group scheme; this helps in proving Proposition 6.1 below and simplifies the definition of an X-torsor under M. Dropping the assumption that M be finite invalidates the statement in general (see [5, §6.7]), but the statement does hold under various alternative conditions.

An *X-torsor under a finite M* is an unramified morphism $\phi: Y \to X$ of degree $\#M$ between k-varieties, together with an isomorphism $M \to \mathrm{Aut}_{\bar{k}}(Y/X)$ of groups with $\mathrm{Gal}(\bar{k}/k)$-action; see [45, § III.4].

Let $\phi: Y \to X$ and $\phi': Y' \to X$ be X-torsors under a finite M. An *isomorphism* of X-torsors is an isomorphism of k-varieties $\sigma: Y \to Y'$ such that $\phi = \phi' \circ \sigma$

and such that the induced isomorphism $\mathrm{Aut}_{\bar{k}}(Y/X) \to \mathrm{Aut}_{\bar{k}}(Y'/X)$ is compatible with the isomorphisms $M \to \mathrm{Aut}_{\bar{k}}(Y/X)$ and $M \to \mathrm{Aut}_{\bar{k}}(Y'/X)$.

Let $X_{\bar{k}}$ be the base change of X to $\bar{k}$. Via base change, we can obtain from any X-torsor under M an $X_{\bar{k}}$-torsor under $M_{\bar{k}}$. We say that two torsors are *twists* of one another if they becomes isomorphic to one another upon base change to $\bar{k}$.

If M is not abelian there is still an object denoted $H^1(k, M)$, but it is no longer a group — it is merely a set with a distinguished element, called the *trivial class*. From Theorem III.4.3(a) (p. 121) and Proposition III.4.6 (p. 123) of [45] we obtain the following result.

Proposition 6.1 (Twisting principle). *Let $\phi: Y \to X$ be an X-torsor under a finite M. There is a bijection between $H^1(k, M)$ and the set of isomorphism classes of twists of $\phi: Y \to X$, and a natural map $\gamma: X(k) \to H^1(k, M)$, such that*

(1) *the bijection sends the trivial class of $H^1(k, M)$ to the class of ϕ, and*

(2) *for every $x \in X(k)$, if $\gamma(x)$ corresponds to a twist $\phi_x: Y_x \to X$, then x has a k-rational preimage on Y_x.*

In fact, if a twist $\phi': Y' \to X$ has a point $y \in Y'(k)$ then Y' is isomorphic to Y_x, where $x = \phi'(y)$. It follows that the image of γ consists exactly of those twists for which $Y'(k)$ is nonempty. We can approximate the image by considering those that have adelic points.

Definition 6.2. Let $\phi: Y \to X$ be an M-cover over $\mathbb{Q}$. We define the *Selmer set* to be

$$\begin{aligned}\mathrm{Sel}(\mathbb{Q}, Y \xrightarrow{\phi} X) &= \left\{[\phi': Y' \to X] \in H^1(\mathbb{Q}, M) : Y'(\mathbb{A}) \neq \varnothing\right\} \\ &= \left\{\delta \in H^1(\mathbb{Q}, M) : \mathrm{res}_v(\delta) \in \mathrm{im}\, \gamma_v \text{ for all } v \in \Omega_{\mathbb{Q}}\right\}.\end{aligned}$$

Note that the multiplication-by-n morphism in the exact sequence (4) yields a J-torsor $J \to J$ under the group $M = J[n](\overline{\mathbb{Q}})$. Indeed, this map γ and the connecting homomorphism in Equation (5) agree, as do the concepts of Selmer set and group.

Of particular importance for us is the case where k is a number field. For ease of notation, we restrict to the case $k = \mathbb{Q}$. Let $\mathbb{Q}_v^{\mathrm{unr}}$ be the maximal unramified extension of $\mathbb{Q}_v$ in $\overline{\mathbb{Q}}_v$. We say a class is *unramified* if it becomes trivial under the restriction $H^1(\mathbb{Q}_v, M) \to H^1(\mathbb{Q}_v^{\mathrm{unr}}, M)$. A class in $H^1(\mathbb{Q}, M)$ is *unramified at v* if res_v maps it to an unramified class. For a finite set $S \subset \Omega_{\mathbb{Q}}$ we write $H^1(\mathbb{Q}, M; S)$ for the subgroup of classes unramified at all places outside of S. We find that $H^1(\mathbb{Q}, M; S)$ is finite; this is analogous to Hermite's result that there are only finitely many number fields of bounded degree unramified outside a finite set of primes.

Proposition 6.3 (Chevalley-Weil [25]). *Let X and Y be smooth projective varieties over $\mathbb{Q}$, let M be a finite* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$*-group, and let $\phi\colon Y \to X$ be an X-torsor under M. Let $S \subset \Omega_{\mathbb{Q}}$ contain the archimedean places, the places of bad reduction of ϕ, and the places of residue characteristic dividing $|M|$. Then*

$$\gamma(X(\mathbb{Q})) \subset \mathrm{Sel}(\mathbb{Q}, Y \xrightarrow{\phi} X) \subset H^1(\mathbb{Q}, M; S).$$

In particular, $\gamma(X(\mathbb{Q}))$ is finite.

The version in [25] states that $\phi^{-1}(X(\mathbb{Q}))$ lies in $Y(L)$ for some fixed number field L, the compositum of degree-$|M|$ extensions unramified outside S of the splitting field of M. This formulation is not very conducive to computation. A more promising approach is to try to find reasonable computational descriptions of $H^1(k, M)$ and γ for $k = \mathbb{Q}$ and $k = \mathbb{Q}_v$. General theory gives us that the map γ_v for $k = \mathbb{Q}_v$ is continuous and therefore locally constant. If we can determine the neighborhood on which γ_v is constant, we can determine $\operatorname{im}\gamma_v$ and thus compute $\mathrm{Sel}(\mathbb{Q}, Y \xrightarrow{\phi} X)$.

7. Computing Selmer groups

In this section we describe a method for computing (or at least approximating) Selmer groups that goes back to Cassels (see [21] for a survey), and that has been developed and used by many others [23; 48; 53; 54; 56]. The presentation here closely follows that in [14].

We continue our philosophy that points on J are most conveniently represented by divisors on C. We would like to describe $J[n]$ as a Galois module. We do so by presenting a finite Galois-stable set of generators $\Delta = \{\theta_1, \ldots, \theta_d\}$. Since this is a finite $\mathrm{Gal}(\overline{k}/k)$-set, it can be viewed as the $\overline{k}$-points of an affine 0-dimensional variety over k, which we also denote by Δ. Its coordinate ring is some finite k-algebra L. Note that L is a field only when $\mathrm{Gal}(\overline{k}/k)$ acts transitively on Δ. In general, L is a direct sum of fields, corresponding to the Galois orbits of Δ.

A certificate that $\theta \in \Delta$ is n-torsion as a divisor class on C can be given as a function f_θ whose divisor is linearly equivalent to $n\theta$. If we take these functions Galois-covariantly, we can combine them into a function $f \in k(C) \otimes_k L$.

We construct an n-torsion Galois module directly from Δ by taking the *twisted power*

$$\left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^{\Delta} := \bigoplus_{i=1}^{d} \left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)\theta_i,$$

which as a group is simply $(\mathbb{Z}/n\mathbb{Z})^d$, but has its Galois action twisted so that the coordinates are permuted according to the action on Δ. The fact that Δ *generates*

$J[n]$ is expressed in the surjectivity of the third arrow in the short exact sequence

$$0 \longrightarrow R \longrightarrow \left(\frac{\mathbb{Z}}{n\mathbb{Z}}\right)^{\Delta} \longrightarrow J[n] \longrightarrow 0,$$

where the map to $J[n]$ consists of evaluating the formal linear combinations and R is defined to be the kernel of that map. If we are able to choose a Galois-stable *basis* for $J[n]$ then R is trivial and we obtain an isomorphism to $J[n]$. In general, we have to choose Δ larger than that. In fact, the Galois group may act transitively on the nonzero elements of $J[n]$, in which case $\Delta = J[n] \setminus \{0\}$ is the only choice.

If M is a finite Galois module, we let $M^\vee$ denote the Cartier dual $\mathrm{Hom}(M, \bar{k}^\times)$ of M. We note that

$$((\mathbb{Z}/n\mathbb{Z})^{\Delta})^\vee = ((\mathbb{Z}/n\mathbb{Z})^\vee)^{\Delta} = \mu_n^{\Delta},$$

and, thanks to the Weil pairing, that $J[n]^\vee = J[n]$. We obtain

$$0 \longrightarrow J[n] \longrightarrow \mu_n^{\Delta} \longrightarrow R^\vee \longrightarrow 0.$$

Taking Galois cohomology yields a map from $H^1(k, J[n])$ to $H^1(k, \mu_n^{\Delta})$. From Kummer theory we know that $H^1(k, \mu_n) = k^\times / k^{\times n}$, and with a little extra work we find that $H^1(k, \mu_n^{\Delta}) = L^\times / L^{\times n}$. Hence we obtain the following commutative diagram with exact rows:

$$\begin{array}{ccccccccc}
 & & & & & & \dfrac{J(k)}{nJ(k)} & \xrightarrow{\tilde{\gamma}} & \dfrac{L^\times}{L^{\times n}} \\
 & & & & & & \gamma \downarrow & & \| \\
0 & \longrightarrow & J[n](k) & \longrightarrow & (\mu_n^{\Delta})(k) & \longrightarrow & R^\vee(k) & \longrightarrow & H^1(k, J[n]) \longrightarrow H^1(k, \mu_n^{\Delta})
\end{array}$$

Note that we represent elements of $J(k)/nJ(k)$ by divisors on C. Our function f provides a partial map

$$\begin{aligned}
\mathrm{Div}(C/k) &\dashrightarrow L^\times \\
\sum_{P \in C(\bar{k})} n_P P &\longmapsto \prod_{P \in C(\bar{k})} f(P)^{n_P}
\end{aligned}$$

defined for divisors supported away from poles and zeros of f. The main work, for which we refer the reader to [14], is to prove that this map induces the map $\tilde{\gamma}$ above.

For $k = \mathbb{Q}$ and $S \subset \Omega_{\mathbb{Q}}$ a finite set containing the infinite place and the primes dividing n, we also need to describe the subgroup $H^1(\mathbb{Q}, \mu_n^{\Delta}; S)$. To that end, we denote by $\mathbb{O}_{L,S}$ the ring of the elements of L that are integral over $\mathbb{Z}_S$. This

ring decomposes into a direct product of Dedekind domains, namely the rings of S-integers of the number fields constituting L. If $\mathcal{O}_{L,S}$ is a principal ideal ring, which can be ensured by enlarging S if necessary, then

$$H^1(\mathbb{Q}, \mu_n^\Delta; S) = \frac{\mathcal{O}_{L,S}^\times}{\mathcal{O}_{L,S}^{\times n}}.$$

Computing an explicit representation amounts to determining class groups and unit groups in number fields.

Our explicit description of the map $\tilde{\gamma}$ also makes it possible to determine neighborhoods on which the local version $\tilde{\gamma}_v$ is constant. The arguments used are similar to those that show that elements $u, v \in \mathbb{Q}_2^\times$ represent the same class in $\mathbb{Q}_2^\times/\mathbb{Q}_2^{\times 2}$ when $2^\epsilon(u-v) \in 1 + 8\mathbb{Z}_2$ for some $\epsilon \in \mathbb{Z}$.

For appropriate sets $S, T \subset \Omega_{\mathbb{Q}}$ we define

$$\begin{aligned}\mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J) &:= \left\{\delta \in H^1(\mathbb{Q}, \mu_n^\Delta) : \mathrm{res}_v(\delta) \in \mathrm{im}\,\tilde{\gamma}_v\right\} \\ &\subseteq \left\{\delta \in \mathcal{O}_{L,S}^\times/\mathcal{O}_{L,S}^{\times n} : \mathrm{res}_v(\delta) \in \mathrm{im}\,\tilde{\gamma}_v \text{ for all } v \in T\right\},\end{aligned}$$

where, for a large enough finite set $T \subset \Omega_{\mathbb{Q}}$, the inclusion stabilizes to an equality.

We have a map $\mathrm{Sel}^n(J/\mathbb{Q}) \to \mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J)$ but this need be neither surjective nor injective. We do know that the kernel is contained in the group K defined by the exact sequence

$$0 \longrightarrow J[n](k) \longrightarrow \mu_n^\Delta(k) \longrightarrow R^\vee(k) \longrightarrow K \longrightarrow 0,$$

and in practice K is frequently trivial. In any case, we can use $\mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J)$ to obtain an upper bound on the rank of $J(\mathbb{Q})$. It may be larger than the one that can be derived from the actual Selmer group, but it has the advantage that it is more easily computed. There are also auxiliary computations one can do to obtain more detailed information on the difference; see [14, Appendix A].

Requiring a set Δ as above is often too demanding. Indeed, in general one does not expect a more favorable choice than $\Delta = J[n] \setminus \{0\}$ to be available. In that case, L is usually a number field of degree $n^{2g} - 1$, where g is the genus of C. So even in the case $g = 3$ and $n = 2$ one expects to have to compute with a number field of degree 63.

At the expense of getting even further removed from a description of $H^1(\mathbb{Q}, J[n])$, one can use a smaller set Δ. We restrict to the case $n = 2$. We take Δ to be a set so that the *differences* of elements of Δ generate $J[2]$. We consider the submodule E of even weight vectors,

$$0 \longrightarrow E \longrightarrow (\mathbb{Z}/2\mathbb{Z})^\Delta \xrightarrow{\text{sum}} \mathbb{Z}/2\mathbb{Z} \longrightarrow 0,$$

and we obtain a short exact sequence

$$0 \longrightarrow R \longrightarrow E \longrightarrow J[n] \longrightarrow 0.$$

Taking cohomology of the dual sequence gives

$$H^1(k, \mu_2) \longrightarrow H^1(k, \mu_2^\Delta) \longrightarrow H^1(k, E^\vee),$$

which leads to

$$\frac{L^\times}{L^{\times 2} k^\times} \subset H^1(k, E^\vee).$$

Provided that $\mathrm{Pic}^0(C/k) = J(k)$, which holds for us by Lemma 4.2, we can show that the part of $H^1(k, E^\vee)$ relevant to us lies in the subgroup we can describe, and we obtain a map

$$\tilde{\gamma}\colon \frac{J(k)}{nJ(k)} \longrightarrow \frac{L^\times}{L^{\times 2} k^\times}$$

which we can use in essentially the same way as above. One can choose Δ to be the set of classes of *odd theta characteristics*, which has size $2^{g-1}(2^g - 1)$, less than half of what we needed before. For $g = 3$ this results in an algebra L of degree 28.

8. Application of descent to other problems

8A. ***Descent on the curve.*** If we embed the curve C in its Jacobian then we can restrict the maps γ and $\tilde{\gamma}$ to C. In that case we can construct a C-torsor under $J[n]$ by pulling C back along the multiplication-by-n map $J \to J$. The result is an unramified cover $\phi\colon D \to C$ of degree n^{2g}.

We can compute approximations of $\mathrm{Sel}(\mathbb{Q}, D \xrightarrow{\phi} C)$ using the same approach as in Section 7. If that approximation turns out to be empty, then C has no rational points. This can happen even if $C(\mathbb{A})$ is nonempty. When it works, this method is easier to apply than Mordell-Weil sieving, because the data we need is required for determining $J(\mathbb{Q})$ anyway, and we do not have to actually find generators for $J(\mathbb{Q})$.

Given that the map $\tilde{\gamma}$ is computed by evaluating a function on representative divisors, we can evaluate $\tilde{\gamma}$ directly on C, without choosing an embedding in J, and even if such an embedding does not exist. See [16] for a more thorough analysis of this method for hyperelliptic curves.

8B. ***Finding an embedding in J.*** A curve C has a degree-n point for some n. For instance, on a curve of genus $g \geq 2$, the canonical divisor class always contains a rational effective divisor, so one can take $n \leq 2g - 2$. It follows that $\underline{\mathrm{Pic}}^n(C) \simeq J$ and hence that multiplication-by-n yields a cover $\underline{\mathrm{Pic}}^1 \to J$. Note that over $\overline{\mathbb{Q}}$ we have $J \simeq \underline{\mathrm{Pic}}^1(C)$ in a way that is compatible with the multiplication-by-n map, so this cover expresses $\underline{\mathrm{Pic}}^1$ as a J-torsor under $J[n]$. By Proposition 6.1, this

torsor corresponds to some class in $H^1(\mathbb{Q}, J[n])$. In fact, if $C(\mathbb{A}) \neq \varnothing$, we have $[\underline{\mathrm{Pic}}^1(C)] \in \mathrm{Sel}^n(J/\mathbb{Q})$. If we have succeeded in determining $J(\mathbb{Q})$, we can check if $[\underline{\mathrm{Pic}}^1(C)]$ lies in the image of $J(\mathbb{Q})$. If it does, then we have an explicit rational point that we can lift to $\underline{\mathrm{Pic}}^1(C)$. If it does not then we have proved that $\underline{\mathrm{Pic}}^1(C)$ does not have a rational point, and therefore neither does C.

We can also adapt the ideas from Section 8A to do further descent computations on $\underline{\mathrm{Pic}}^1$, although doing a descent directly on C yields stronger information for our purposes — see [27].

8C. ***Covering collections.*** Proposition 6.3 also provides useful information when Chabauty's method (Section 5) does not apply because $J(\mathbb{Q})$ is of too high rank. As we saw in Section 8A, we can use the embedding $C \to J$ to obtain unramified Galois covers $D \xrightarrow{\phi} C$. As Proposition 6.3 shows, one has

$$C(\mathbb{Q}) = \bigcup_{[D' \xrightarrow{\phi'} C] \in \mathrm{Sel}(\mathbb{Q}, D \xrightarrow{\phi} C)} \phi'(D'(\mathbb{Q})).$$

Note that D (and hence any of the D') is of higher genus than C, so Chabauty's method might apply to D' even if it does not to C; see also [64]. *A priori* it may seem computationally unattractive to compute with a curve of much higher genus. However, by construction, the curve D is far from general; for example, it has many automorphisms. That usually means that its Jacobian can be decomposed into factors of lower dimension. For instance, if C is a hyperelliptic curve and D is a C-torsor under $J[2]$, the Jacobian of D has many elliptic isogeny factors, although not necessarily over $\mathbb{Q}$. This means that many of the computations that would normally take place on the Jacobian of D can now be done on elliptic curves. This greatly simplifies computations and has led to a variant of Chabauty's method commonly referred to as *elliptic curve Chabauty*. See [36] for a special case and [8], [9] for the general case, as well as an application that amounts to a Chabauty computation on a 12-dimensional abelian variety. See also [16] on how to use descent computation to determine which twists to consider and [13] for an iterated application of these ideas. See [11] for an application to a curve of genus 3 admitting a double cover; this example involves Mordell-Weil sieving and a Chabauty computation on a genus-5 curve embedded in an abelian surface presented as the Jacobian of an otherwise unrelated curve of genus 2.

9. Smooth plane quartics

As an example, let us see how the ideas in the previous sections apply to smooth plane quartics — that is, nonhyperelliptic genus-3 curves. In a way, this is the simplest collection of truly *general* curves, in the sense that genus-2 curves are always hyperelliptic and hence necessarily have a nontrivial automorphism. The

examples come from [14], to which the reader is referred for further details and references.

Let $C \subset \mathbb{P}^2$ be a smooth plane quartic curve over $\mathbb{Q}$. We apply the procedure described in Section 7 for $n = 2$. The set Δ has a particularly explicit description. A smooth plane quartic has 28 *bitangents*. If l and m are degree-1 forms on C that describe bitangents, then l/m obviously induces a function on the curve whose divisor is twice another divisor. That divisor therefore represents a 2-torsion class. It is a matter of combinatorics to compute that every nonzero 2-torsion point can be described this way (in fact, in 6 different ways). Let $\Delta \subset (\mathbb{P}^2)^*$ be the 0-dimensional, degree-28 locus in the dual space corresponding to these 28 bitangents, and let L be the affine coordinate ring of Δ, so that L is a finite algebra over $\mathbb{Q}$ of degree 28.

The Galois group of (a splitting field of) L is a subgroup of $\mathrm{Sp}_6(\mathbb{F}_2)$, which is also the generic Galois group of $J[2]$. For this full group, the module $(\mathbb{Z}/2\mathbb{Z})^\Delta$ has unique submodules E and R of dimensions 27 and 21 respectively, giving us a unique sequence of $\mathrm{Sp}_6(\mathbb{F}_2)$-modules

$$0 \longrightarrow R \longrightarrow E \longrightarrow J[2] \longrightarrow 0.$$

If we identify the conjugacy class in $\mathrm{Sp}_6(\mathbb{F}_2)$ of the group through which $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on L, then we can determine the action on the sequence via restriction. This means we can determine the sequence

$$0 \longrightarrow J[2](\mathbb{Q}) \longrightarrow E^\vee(\mathbb{Q}) \longrightarrow R^\vee(\mathbb{Q})$$

by identifying the Galois group of L as a subgroup of $\mathrm{Sp}_6(\mathbb{F}_2)$. Determining Galois groups is one of the classic problems in computational algebraic number theory.

For each $\theta \in \Delta$ we obtain a linear form $l_\theta \in L[x, y, z]$, where x, y, z are the coordinates on $\mathbb{P}^2$. Evaluating $\tilde{\gamma}$ at a point on C amounts to evaluating l_θ at that point.

In order to compute $\mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J)$, we need to compute the ideal class group and unit group of L, for which we need an integral basis as well. The computation of class groups, unit groups, and integral bases are three further classical problems in computational algebraic number theory.

We give some examples.

Proposition 9.1. *If C is the curve*

$$x^3y - x^2y^2 - x^2z^2 - xy^2z + xz^3 + y^3z = 0$$

in $\mathbb{P}^2_{\mathbb{Q}}$, then $J(\mathbb{Q}) = \langle[(0:1:0) - (0:0:1)]\rangle \simeq \mathbb{Z}/51\mathbb{Z}$ and

$$C(\mathbb{Q}) = \{(1:1:1), (0:1:0), (0:0:1), (1:0:0), (1:1:0), (1:0:1)\}.$$

For this example the Galois group of L is a member of the unique index-36 conjugacy class of $\mathrm{Sp}_6(\mathbb{F}_2)$. For that group we find that $R^\vee(\mathbb{Q}) = \mathbb{Z}/2\mathbb{Z}$ and that $E^\vee(\mathbb{Q}) = 0$. *A priori* this leaves room for a nontrivial kernel in

$$\mathrm{Sel}^2(\mathbb{Q}, J) \to \mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J).$$

However, we find that $R^\vee(\mathbb{Q}_2) = R^\vee(\mathbb{Q})$ and $E^\vee(\mathbb{Q}_2) = E^\vee(\mathbb{Q})$ and that the image of $R^\vee(\mathbb{Q}_2)$ does not lie in the image of γ_2. This means that the map is an injection anyway and, since $\mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J) = 0$, that $J(\mathbb{Q})$ is finite and of odd order. Further investigation shows that there is 51-torsion. Finding the rational points of C from the finite set $J(\mathbb{Q})$ is trivial.

Proposition 9.2. *Let C be the curve*

$$x^2y^2 - xy^3 - x^3z - 2x^2z^2 + y^2z^2 - xz^3 + yz^3 = 0$$

in $\mathbb{P}^2_{\mathbb{Q}}$. If the generalized Riemann hypothesis holds, then $J(\mathbb{Q}) \simeq \mathbb{Z}$ and

$$C(\mathbb{Q}) = \{(1:1:0), (-1:0:1), (0:-1:1), (0:1:0),\\ (1:1:-1), (0:0:1), (1:0:0), (1:4:-3)\}.$$

For this curve, the Galois group of L is all of $\mathrm{Sp}_6(\mathbb{F}_2)$. Then $R^\vee(\mathbb{Q}) = 0$, so

$$\mathrm{Sel}^2(\mathbb{Q}, J) \subseteq \mathrm{Sel}^{\tilde{\gamma}}(\mathbb{Q}, J).$$

Further computation shows that the latter has size 2, so $J(\mathbb{Q})$ has rank at most 1. Furthermore, we have $J(\mathbb{F}_3) \simeq \mathbb{Z}/85\mathbb{Z}$ and $J(\mathbb{F}_7) \simeq \mathbb{Z}/336\mathbb{Z}$. These group orders are coprime, so $J(\mathbb{Q})$ is torsion free. It is straightforward to exhibit a nontrivial point in $J(\mathbb{Q})$, so it follows the rank is 1. A straightforward application of Chabauty's method yields the rest of the statement.

We invoke the generalized Riemann hypothesis to verify the class group information. The Minkowski bound of L (which is a field in this case) is 1,008,340,641, so a dedicated enthusiast could probably confirm the class group information unconditionally.

Proposition 9.3. *Let C be the curve in $\mathbb{P}^2_{\mathbb{Q}}$ defined by*

$$x^4 + y^4 + x^2yz + 2xyz^2 - y^2z^2 + z^4 = 0.$$

Then $C(\mathbb{R}) \neq \varnothing$ and $C(\mathbb{Q}_p) \neq \varnothing$ for all p, but if the generalized Riemann hypothesis holds, then $C(\mathbb{Q}) = \varnothing$.

For this curve we verify that its $\tilde{\gamma}$-Selmer set is empty. The Minkowski bound for L exceeds 10^{22} so unconditional verification is out of the question.

Acknowledgments

I would like to thank the ANTS X Program Committee, and in particular Everett Howe and Kiran Kedlaya, for their hard work in organizing an excellent symposium. I would also like to thank an anonymous referee for comments on an earlier draft of this article. It helped clarify the exposition greatly.

This research was supported by NSERC.

References

[1] Scott D. Ahlgren, George E. Andrews, and Ken Ono (eds.), *Topics in number theory*: *Proceedings of the conference held at the Pennsylvania State University*, *University Park*, *PA*, *July* 31–*August* 3, 1997, Mathematics and its Applications, no. 467, Dordrecht, Kluwer Academic Publishers, 1999. MR 99m:11004

[2] Jennifer S. Balakrishnan, *Iterated Coleman integration for hyperelliptic curves*, in Howe and Kedlaya [40], 2013, pp. 41–61.

[3] B. J. Birch and H. P. F. Swinnerton-Dyer, *Notes on elliptic curves*, *I*, J. Reine Angew. Math. **212** (1963), 7–25. MR 26 #3669

[4] ———, *Notes on elliptic curves*, *II*, J. Reine Angew. Math. **218** (1965), 79–108. MR 31 #3419

[5] Siegfried Bosch, Werner Lütkebohmert, and Michel Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), no. 21, Springer, Berlin, 1990. MR 91i:14034

[6] Wieb Bosma and John Cannon (eds.), *Discovering mathematics with Magma*: *Reducing the abstract to the concrete*, Algorithms and Computation in Mathematics, no. 19, Springer, Berlin, 2006. MR 2007h:00016

[7] N. Bruin and E. V. Flynn, *Rational divisors in rational divisor classes*, in Buell [19], 2004, pp. 132–139. MR 2005m:11117

[8] N. R. Bruin, *Chabauty methods and covering techniques applied to generalized Fermat equations*, CWI Tract, no. 133, Stichting Mathematisch Centrum Centrum voor Wiskunde en Informatica, Amsterdam, 2002. http://persistent-identifier.org/?identifier=urn:nbn:nl:ui:18-13154 MR 2003i:11042

[9] Nils Bruin, *Chabauty methods using elliptic curves*, J. Reine Angew. Math. **562** (2003), 27–49. MR 2004j:11051

[10] ———, *Some ternary Diophantine equations of signature* $(n, n, 2)$, in Bosma and Cannon [6], 2006, pp. 63–91. MR 2007m:11047

[11] ———, *The arithmetic of Prym varieties in genus* 3, Compos. Math. **144** (2008), no. 2, 317–338. MR 2009f:11074

[12] Nils Bruin and Noam D. Elkies, *Trinomials* $ax^7 + bx + c$ *and* $ax^8 + bx + c$ *with Galois groups of order* 168 *and* $8 \cdot 168$, in Fieker and Kohel [30], 2002, pp. 172–188. MR 2005d:11094

[13] Nils Bruin and E. Victor Flynn, *Towers of* 2*-covers of hyperelliptic curves*, Trans. Amer. Math. Soc. **357** (2005), no. 11, 4329–4347. MR 2006k:11118

[14] Nils Bruin, Bjorn Poonen, and Michael Stoll, *Generalized explicit descent and its application to curves of genus* 3, 2012. arXiv 1205.4456 [math.NT]

[15] Nils Bruin and Michael Stoll, *Deciding existence of rational points on curves*: *an experiment*, Experiment. Math. **17** (2008), no. 2, 181–189. MR 2009d:11100

[16] ———, *Two-cover descent on hyperelliptic curves*, Math. Comp. **78** (2009), no. 268, 2347–2370. MR 2010e:11059

[17] ———, *The Mordell-Weil sieve: Proving non-existence of rational points on curves*, LMS J. Comput. Math. **13** (2010), 272–306. MR 2011j:11118

[18] Armand Brumer and Kenneth Kramer, *The rank of elliptic curves*, Duke Math. J. **44** (1977), no. 4, 715–743. MR 56 #15658

[19] Duncan Buell (ed.), *Algorithmic number theory: Proceedings of the 6th International Symposium (ANTS-VI) held at the University of Vermont, Burlington, VT, June* 13–18, 2004, Lecture Notes in Computer Science, no. 3076, Berlin, Springer, 2004. MR 2005m:11002

[20] Yann Bugeaud, Maurice Mignotte, Samir Siksek, Michael Stoll, and Szabolcs Tengely, *Integral points on hyperelliptic curves*, Algebra Number Theory **2** (2008), no. 8, 859–885. MR 2010b:11066

[21] J. W. S. Cassels, *Diophantine equations with special reference to elliptic curves*, J. London Math. Soc. **41** (1966), 193–291. MR 33 #7299

[22] ———, *Corrigenda: "Survey article: Diophantine equations with special reference to elliptic curves"*, J. London Math. Soc. **42** (1967), 183. MR 34 #2523

[23] J. W. S. Cassels and E. V. Flynn, *Prolegomena to a middlebrow arithmetic of curves of genus* 2, London Mathematical Society Lecture Note Series, no. 230, Cambridge University Press, 1996. MR 97i:11071

[24] Claude Chabauty, *Sur les points rationnels des courbes algébriques de genre supérieur à l'unité*, C. R. Acad. Sci. Paris **212** (1941), 882–885. MR 3,14d

[25] Claude Chevalley and André Weil, *Un théorème d'arithmétique sur les courbes algébriques*, C. R. Acad. Sci. Paris **195** (1932), 570–572.

[26] Robert F. Coleman, *Effective Chabauty*, Duke Math. J. **52** (1985), no. 3, 765–770. MR 87f:11043

[27] Brendan Creutz, *Explicit descent in the Picard group of a cyclic cover of the projective line*, in Howe and Kedlaya [40], 2013, pp. 295–315.

[28] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Invent. Math. **73** (1983), no. 3, 349–366. MR 85g:11026a

[29] ———, *Erratum: "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern"*, Invent. Math. **75** (1984), no. 3, 381. MR 85g:11026b

[30] Claus Fieker and David R. Kohel (eds.), *Algorithmic number theory: Proceedings of the 5th International Symposium (ANTS-V) held at the University of Sydney, July* 7–12, 2002, Lecture Notes in Computer Science, no. 2369, Berlin, Springer, 2002. MR 2004j:11002

[31] E. V. Flynn, *An explicit theory of heights*, Trans. Amer. Math. Soc. **347** (1995), no. 8, 3003–3015. MR 95j:11052

[32] ———, *A flexible method for applying Chabauty's theorem*, Compositio Math. **105** (1997), no. 1, 79–94. MR 97m:11083

[33] ———, *The Hasse principle and the Brauer-Manin obstruction for curves*, Manuscripta Math. **115** (2004), no. 4, 437–466. MR 2005j:11047

[34] E. V. Flynn, Bjorn Poonen, and Edward F. Schaefer, *Cycles of quadratic polynomials and rational points on a genus-*2 *curve*, Duke Math. J. **90** (1997), no. 3, 435–463. MR 98j:11048

[35] E. V. Flynn and N. P. Smart, *Canonical heights on the Jacobians of curves of genus* 2 *and the infinite descent*, Acta Arith. **79** (1997), no. 4, 333–352. MR 98f:11066

[36] E. Victor Flynn and Joseph L. Wetherell, *Finding rational points on bielliptic genus* 2 *curves*, Manuscripta Math. **100** (1999), no. 4, 519–533. MR 2001g:11098

[37] Daniel M. Gordon and David Grant, *Computing the Mordell-Weil rank of Jacobians of curves of genus two*, Trans. Amer. Math. Soc. **337** (1993), no. 2, 807–824. MR 93h:11057

[38] Benedict H. Gross and Don B. Zagier, *Heegner points and derivatives of L-series*, Invent. Math. **84** (1986), no. 2, 225–320. MR 87j:11057

[39] F. Hess, *Computing Riemann-Roch spaces in algebraic function fields and related topics*, J. Symbolic Comput. **33** (2002), no. 4, 425–445. MR 2003j:14032

[40] Everett W. Howe and Kiran S. Kedlaya (eds.), *Algorithmic number theory*: *Proceedings of the* 10*th Biennial International Symposium* (*ANTS-X*) *held in San Diego*, *July* 9–13, 2012, The Open Book Series, no. 1, Berkeley, Mathematical Sciences Publishers, 2013, THIS VOLUME.

[41] Nicholas M. Katz, *Galois properties of torsion points on abelian varieties*, Invent. Math. **62** (1981), no. 3, 481–502. MR 82d:14025

[42] Kamal Khuri-Makdisi, *Asymptotically fast group operations on Jacobians of general curves*, Math. Comp. **76** (2007), no. 260, 2213–2239. MR 2009a:14072

[43] V. A. Kolyvagin, *Finiteness of $E(\mathbf{Q})$ and $SH(E, \mathbf{Q})$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671. MR 89m:11056

[44] Adrien-Marie le Gendre, *Recherches d'analyse indéterminée*, Histoire de l'Académie royale des sciences **1785** (1788), 465–559. http://gallica.bnf.fr/ark:/12148/bpt6k35847/f649

[45] James S. Milne, *Étale cohomology*, Princeton Mathematical Series, no. 33, Princeton University Press, 1980. MR 81j:14002

[46] L. J. Mordell, *On the rational solutions of the indeterminate equations of the third and fourth degrees*, Cambr. Phil. Soc. Proc. **21** (1922), 179–192. JFM 48.1156.03

[47] Bjorn Poonen, *Heuristics for the Brauer-Manin obstruction for curves*, Experiment. Math. **15** (2006), no. 4, 415–420. MR 2008d:11062

[48] Bjorn Poonen and Edward F. Schaefer, *Explicit descent for Jacobians of cyclic covers of the projective line*, J. Reine Angew. Math. **488** (1997), 141–188. MR 98k:11087

[49] Bjorn Poonen, Edward F. Schaefer, and Michael Stoll, *Twists of $X(7)$ and primitive solutions to $x^2 + y^3 = z^7$*, Duke Math. J. **137** (2007), no. 1, 103–158. MR 2008i:11085

[50] Bjorn Poonen and Michael Stoll, *A local-global principle for densities*, in Ahlgren et al. [1], 1999, pp. 241–244. MR 2000e:11082

[51] Bjorn Poonen and Yuri Tschinkel (eds.), *Arithmetic of higher-dimensional algebraic varieties*: *Proceedings of the Workshop on Rational and Integral Points of Higher-Dimensional Varieties held in Palo Alto*, *CA*, *December* 11–20, 2002, Progress in Mathematics, no. 226, Boston, Birkhäuser, 2004. MR 2004h:11001

[52] Bjorn Poonen and José Felipe Voloch, *Random Diophantine equations*, in Poonen and Tschinkel [51], 2004, pp. 175–184. MR 2005g:11055

[53] Edward F. Schaefer, 2*-descent on the Jacobians of hyperelliptic curves*, J. Number Theory **51** (1995), no. 2, 219–232. MR 96c:11066

[54] ———, *Computing a Selmer group of a Jacobian using functions on the curve*, Math. Ann. **310** (1998), no. 3, 447–471. MR 99h:11063

[55] ———, *Erratum*: *"Computing a Selmer group of a Jacobian using functions on the curve"* [*Math. Ann.* 310 (1998), *no.* 3, 447–471], Math. Ann. **339** (2007), no. 1, 1. MR 2008f:11063

[56] Edward F. Schaefer and Michael Stoll, *How to do a p-descent on an elliptic curve*, Trans. Amer. Math. Soc. **356** (2004), no. 3, 1209–1231. MR 2004g:11045

[57] Victor Scharaschkin, *Local-global problems and the Brauer-Manin obstruction*, Ph.D. thesis, University of Michigan, Ann Arbor, MI, 1999, p. 59. http://search.proquest.com/docview/304517948 MR 2700328

[58] Denis Simon, *Solving quadratic equations using reduced unimodular quadratic forms*, Math. Comp. **74** (2005), no. 251, 1531–1543. MR 2005k:11246

[59] Alexei Skorobogatov, *Torsors and rational points*, Cambridge Tracts in Mathematics, no. 144, Cambridge University Press, 2001. MR 2002d:14032

[60] Michael Stoll, *On the height constant for curves of genus two*, Acta Arith. **90** (1999), no. 2, 183–201. MR 2000h:11069

[61] _____, *On the height constant for curves of genus two, II*, Acta Arith. **104** (2002), no. 2, 165–182. MR 2003f:11093

[62] _____, *Independence of rational points on twists of a given curve*, Compos. Math. **142** (2006), no. 5, 1201–1214. MR 2007m:14025

[63] André Weil, *L'arithmétique sur les courbes algébriques*, Acta Math. **52** (1929), no. 1, 281–315. MR 1555278

[64] Joseph Loebach Wetherell, *Bounding the number of rational points on certain curves of high rank*, Ph.D. thesis, University of California, Berkeley, Ann Arbor, MI, 1997, p. 61. http://search.proquest.com/docview/304343505 MR 2696280

NILS BRUIN: nbruin@sfu.ca
Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.213

msp

Solving quadratic equations in dimension 5 or more without factoring

Pierre Castel

Let Q be a 5×5 symmetric matrix with integral entries and with $\det Q \neq 0$, but neither positive nor negative definite. We describe a probabilistic algorithm which solves the equation ${}^t xQx = 0$ over $\mathbb{Z}$ without factoring $\det Q$. The method can easily be generalized to forms of higher dimensions by reduction to a suitable subspace.

1. Introduction

Solving quadratic equations in dimension 1 is trivial: Since the equation is $ax^2 = 0$, the only solution is $x = 0$. In two dimensions, the homogeneous equation is $ax^2 + bxy + cy^2 = 0$, and the solution is obtained by computing a square root. In dimension 3, the equation is

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz = 0,$$

where the coefficients are integers. Since the polynomial becomes more complicated as the dimension increases, we use matrix notation instead. We define Q as the associated quadratic form. If we denote by $X = (x, y, z)$ the row vector containing the variables, the equation becomes

$${}^t X \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} X = 0.$$

MSC2010: primary 11E20; secondary 11D09.
Keywords: general quadratic forms, factorization, general quadratic equations, isotropic spaces, algorithmic number theory, Cebotarev density theorem.

If the equation has a solution, several algorithms exist for finding solutions, for instance see Simon [11] or Cremona [3]. In dimension 3 it is known that finding a (nontrivial) isotropic vector is equivalent to factoring the determinant of the form.

The situation is almost the same in dimension 4 when the determinant is a square: Solutions may not exist, and if a solution exists, finding one is equivalent to factoring the determinant.

The situation is quite different in dimensions greater than or equal to 5. The Hasse-Minkowski theorem [9] asserts that in such dimensions a nontrivial solution always exists. It is easy to see that one just needs the result in dimension 5, since larger dimensions can handled by restricting the form to a subspace of dimension 5 where the form has a suitable signature. This is why we will focus on quadratic forms in dimension 5. As in dimensions 3 and 4, there exist algorithms such as the ones given in [10], but since they are generalizations of algorithms in smaller dimensions, they still need the factorization of the determinant, which rapidly becomes prohibitive. Thus, if we know the factorization of the determinant we can easily find a solution, so the question is whether it is possible to find a solution (in polynomial time) without factorizing the determinant. The goal of this paper is to show that this is indeed possible; in other words, we will give an algorithm which finds a (nontrivial) isotropic vector for a 5-dimensional quadratic form which does not require the factorization of the determinant.

As already mentioned, this algorithm can also be used for forms of higher dimensions by restricting the form to a dimension 5 subspace where the restricted form has a suitable signature. The solution is found over the integers, but since the equation is homogeneous, this is equivalent to finding a rational solution.

The first part of this paper gives the definitions needed to understand the algorithm, the second part explains how the algorithm works, and the last part gives some ideas of the complexity of the method. The full analysis of its complexity is not done here, since it requires a number of tools from analytic number theory and the Cebotarev density theorem [6]. I refer the interested reader to [1].

Basic definitions and notation

To begin, we give definitions and basic properties which we need.

We denote the set of integral quadratic forms as follows.

Definition 1.1. Let n be a nonzero positive integer. We denote by $\mathrm{Sym}(n,\mathbb{Z})$ the set of $n \times n$ symmetric matrices with nonzero determinant and integral entries.

We recall the definition of the Smith normal form of a matrix; for more details, see [2].

Definition 1.2 (Smith normal form). Let A be an $n \times n$ matrix with coefficients in $\mathbb{Z}$ and nonzero determinant. There exists a unique matrix in Smith normal form B such that $B = VAU$ with U and V elements of $\mathrm{GL}_n(\mathbb{Z})$. If we set $d_i = b_{i,i}$, the d_i are called the *elementary divisors* of the matrix A, and we have

$$A = U^{-1} \begin{bmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & d_n \end{bmatrix} V^{-1}$$

with $d_{i+1} \mid d_i$ for $1 \le i < n$.

Definition 1.3. For a matrix $M \in \mathcal{M}_n(\mathbb{Z})$ with nonzero determinant, we denote by $d_1(M), \dots, d_n(M)$ its elementary divisors (given by its Smith normal form). If there is no possible confusion, they will be denoted $d_1, \dots, d_n$.

We can now add a restriction to the set of quadratic forms.

Definition 1.4. Let n be a nonzero positive integer. We denote by $\mathrm{Sym}^*(n, \mathbb{Z})$ the set of $n \times n$ symmetric matrices with nonzero determinant and integral entries, such that their coefficient d_2 as defined above is equal to 1.

2. The algorithm

2A. ***The main idea.*** The key idea of the method is to increase by 1 the dimension of the form by adding a row and a column, then to use an efficient algorithm to find solutions to our new form, and finally to deduce a solution to the original form by considering intersections of hyperbolic spaces of suitable dimensions.

Since Simon's algorithm [10] is very efficient when the factorization of the determinant is known, we are going to build a new 6-dimensional quadratic form Q_6 starting from Q, whose determinant will be equal to $2p$ where p is an odd prime number. We will call this the *completion step*. To do this, we choose an integral vector $X = (x_1, \dots, x_5)$ of dimension 5 and an integer z and we complete Q in the following way:

$$Q_6 = \left[\begin{array}{ccc:c} & & & x_1 \\ & Q & & \vdots \\ & & & x_5 \\ \hdashline x_1 & \cdots & x_5 & z \end{array}\right]. \tag{1}$$

Lemma 2.1. *Let Q be a symmetric matrix with integral entries and with* $\det Q \neq 0$. *If we complete Q to the form Q_6 as described in* (1) *above, then we have*

$$\det Q_6 = z \det Q - {}^tX \operatorname{Co}(Q) X, \tag{2}$$

where $\mathrm{Co}(Q)$ *is the matrix of cofactors of the matrix* Q.

Proof. Simply use the formula involving the cofactors of Q_6 for computing its determinant, and expand it along the row and then the column containing the x_i. □

Some special cases may occur: There exist cases where all the values taken by $\det Q_6$ have a common factor. To avoid these cases we will have to do some minimizations of the form Q before completing it. In order to be able to do a complexity analysis of the algorithm we will need the determinant of Q_6 to be odd, so we will also have to perform a reduction of the even part of the determinant.

2B. ***Minimizations.*** The values taken by the determinant of the form Q_6 will follow from the next result.

Theorem 2.2. *Let* $Q \in \mathrm{Sym}(5, \mathbb{Z})$ *and* $\Delta = \det Q$. *Then for all* $X \in \mathbb{Z}^5$ *and for all* $z \in \mathbb{Z}$ *we have that* $d_2(Q)$ *divides* $\det Q_6$, *where* Q_6 *is defined by* (1).

Proof. Consider the Smith normal form of Q: There exist three matrices D, U, and V with integer entries such that D is diagonal with the elementary divisors on the diagonal, U and V have determinant ± 1, and $D = UQV$. Because of the relation (2), let us consider the values of $-{}^tX\,\mathrm{Co}(Q)X \pmod{\Delta}$. We have

$$\begin{aligned}\mathrm{Co}(Q) &= \mathrm{Co}(V^{-1})\,\mathrm{Co}(D)\,\mathrm{Co}(U^{-1})\\ &= (\det V)(\det U)\,{}^tV\,\mathrm{Co}(D)\,{}^tU\\ &= \pm{}^t\big(U\,{}^t\mathrm{Co}(D)V\big)\\ &= \pm{}^t(U\,\mathrm{Co}(D)V).\end{aligned}$$

Since D is the diagonal matrix of elementary divisors, it follows that $\mathrm{Co}(D)$ is also diagonal and that every coefficient is divisible by $d_2(Q)$. We thus have

$$\begin{aligned}{}^tX\,\mathrm{Co}(Q)X &= \pm{}^tX\,{}^t(U\,\mathrm{Co}(D)V)X\\ &\equiv 0 \pmod{d_2(Q)}.\end{aligned}$$

Combining this congruence with the formula (2) proves the result. □

Remark. If $d_1(Q) \neq \det Q$ it will not be possible to have $\det Q_6$ equal to a prime or twice an odd prime number, so we will first need to minimize Q so as to obtain an equivalent form Q' such that $d_2(Q') = 1$.

Remark 2.3. If we perform a change of basis using the matrix V of the previous result with $d_i(Q) \neq 1$ and $d_{i+1}(Q) = 1$, the first i columns and rows will be divisible by $d_i(Q)$.

We are now going to explain what to do in order to avoid the case $d_2(Q) \neq 1$.

Case $d_5 \neq 1$.

Proposition 2.4. *Let* $Q \in \mathrm{Sym}(5, \mathbb{Z})$ *such that* $d_5(Q) \neq 1$. *There exist two* 5×5 *matrices with integral entries* G *and* Q_f *such that*

$$d_5 Q_f = {}^t G Q G,$$
$$\det Q_f = \frac{1}{d_5^5} \det Q.$$

The proof is given by the following algorithm.

Algorithm 2.5 (Minimization 5).

Input: $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_5(Q) \neq 1$ and $m \neq 1 \in \mathbb{Z}$ dividing $d_5(Q)$.

Output: Q_f: a form equivalent to Q such that $\det Q_f = (1/m^5) \det Q$;
G : the corresponding change of basis such that $d_5 Q_f = {}^t G Q G$.

1. Set $G := \mathrm{Id}_5$.
2. Set $Q_f := (1/m) Q$.
3. Return Q_f, G.

When the coefficient d_5 of the Smith normal form of Q is different from 1, the whole matrix Q is divisible by d_5, so the minimization simply consists in dividing the matrix by d_5 and the corresponding change of basis G is equal to Id_5.

Case $d_4 \neq 1$ *and* $d_5 = 1$.

Proposition 2.6. *Let* $Q \in \mathrm{Sym}(5, \mathbb{Z})$ *such that* $d_4(Q) \neq 1$ *and* $d_5(Q) = 1$. *There exist two* 5×5 *matrices with integral entries* G *and* Q_f *such that*

$$d_4 Q_f = {}^t G Q G,$$
$$\det Q_f = \frac{1}{d_4^3} \det Q.$$

The proof is given by the following algorithm.

Algorithm 2.7 (Minimization 4).

Input: $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_4(Q) \neq 1$ and $d_5(Q) = 1$, $m \neq 1 \in \mathbb{Z}$ dividing $d_4(Q)$.

Output: Q_f: a form equivalent to Q such that $\det Q_f = (1/m^3) \det Q$;
G : the corresponding change of basis such that $m Q_f = {}^t G Q G$.

1. Let V be the V matrix given by the SNF of Q.
2. Let H be the diagonal matrix such that for $1 \leq i \leq 4$, $H_{i,i} = 1$ and $H_{5,5} = m$.
3. Set $G := V \times H$; $Q' := (1/m)\, {}^t G Q G$.

4. Apply the LLL algorithm for indefinite forms to Q' (see [11] for more details). Let Q_f be the returned form and G' the corresponding change of basis.
5. Set $G := G \times G'$.
6. Return Q_f, G.

As stated in Remark 2.3, after the change of basis in step 1, the first four columns and rows are divisible by d_4. Thus we apply this change of basis, multiply the last row and column by d_4, and divide the whole matrix by d_4.

Remark. The notion of equivalence between quadratic forms used here simply means that both corresponding quadratic equations have the same solutions up to a change of basis.

Case $d_3 \neq 1$ and $d_4 = 1$.

Proposition 2.8. *Let $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_3(Q) \neq 1$ and $d_4(Q) = 1$. There exist two 5×5 matrices with integer entries G and Q_f such that*

$$d_3 Q_f = {}^t G Q G,$$
$$\det Q_f = \frac{1}{d_3} \det Q.$$

The proof is given by the following algorithm:

Algorithm 2.9 (Minimization 3).

Input: $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_3(Q) \neq 1$ and $d_4(Q) = 1$, $m \neq 1 \in \mathbb{Z}$ dividing $d_3(Q)$.

Output: Q_f: a form equivalent to Q such that $\det Q_f = (1/m) \det Q$;
G : the corresponding change of basis such that $mQ_f = {}^t G Q G$.

1. Let V be the V matrix given by the SNF of Q.
2. Let H be the diagonal matrix such that for $1 \leq i \leq 3$, $H_{i,i} = 1$ and $H_{4,4} = H_{5,5} = m$.
3. Set $G := V \times H$; $Q' := (1/m)\, {}^t G Q G$.
4. Apply the LLL algorithm to Q'. Let Q_f be the returned form and G' the corresponding change of basis.
5. Set $G := G \times G'$.
6. Return Q_f, G.

The minimizing method for this case is essentially the same as for the previous one.

Case $d_2 \neq 1$ and $d_3 = 1$. This case is much more complicated than the previous ones. If we try to do it in the same way, we will multiply the determinant by some factor which is of course not what we want. The idea is first to perform a change of basis thanks to the matrix V given by the SNF of Q, and then to work on the 3×3 block that remains which may not be divisible by $d_2(Q)$. What we need to do in order to be able to apply the same method is to be in the case where the upper-left coefficient of this block is already divisible by $d_2(Q)$. We are thus going to do a special change of basis in order to succeed. The method is given by the following result.

Proposition 2.10. *Let $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_2(Q) \neq 1$ and $d_3 = 1$. Let m be an integer such that $m \neq 1$ and $m \mid d_2(Q)$. There exist two 5×5 matrices with integral entries G and Q_f, with G unimodular, and such that*

$$mQ_f = {}^tGQG,$$
$$\det Q_f = \frac{1}{m} \det Q.$$

Proof. We first compute the SNF of Q, so that $D = UQV$ where D, U, V have integral entries and U and V are unimodular. We apply the change of basis given by the matrix V. The quadratic form $Q' = {}^tVQV$ is equivalent to the form Q and its first two rows and columns are divisible by m. Denote by Q_3 the restriction of Q' to the space spanned by the last three columns of the matrix V. This corresponds to the submatrix $(Q_3)_{i,j} = (Q')_{i,j}$ with $3 \leq i \leq 5$, $3 \leq j \leq 5$. We now want to have $Q_{3\,1,1} \equiv 0 \pmod{m}$. We apply a Gram-Schmidt orthogonalization process to the matrix Q_3 modulo m. If we find a noninvertible element modulo m, this means that we have found a factor of m. In that case we start the process again by replacing m by its divisor. During the process, if we find a vector whose norm is 0 modulo m, we just have to skip this step since this vector is exactly the one we need. Otherwise the process ends and gives us a change of basis such that in this new basis, the form $Q_3 \pmod{m}$ has the shape

$$\begin{bmatrix} a & & 0 \\ & b & \\ 0 & & c \end{bmatrix} \pmod{m}.$$

We must now solve the following quadratic equation:

$$ax^2 + by^2 + cz^2 \equiv 0 \pmod{m}. \tag{3}$$

Since we do not want to factor m, we have to use a method which does not use its factorization. Such a method is described in [8]: If the coefficient a is not invertible modulo m we have found a factor of m, so we can continue the process with both factors, obtain the solution for each of them and combine them using the Chinese

remainder theorem and Hensel lifting if needed. We are thus reduced to the case where a is invertible modulo m. Solving (3) is equivalent to solving the equation

$$x^2 + ba^{-1}y^2 \equiv -ca^{-1}z^2 \pmod{m}. \tag{4}$$

If we take the arbitrary choice $z = 1$, we have exactly the type of equation that is solved in [8]. We thus use this method to obtain a solution S of (3). We complete the single vector family $\{S\}$ to a unimodular matrix G, and we extend the matrix G to a matrix G' of dimension 5 by taking the identity matrix Id_5 and replacing the 3×3 lower-right block by G. We now apply G' to Q' and obtain Q'' which has the form

$${}^tG'Q'G' = Q'' = \left[\begin{array}{c|ccc} mM_{2,2} & & mM_{2,3} & \\ \hline & m* & * & * \\ mM_{3,2} & * & * & * \\ & * & * & * \end{array}\right],$$

where the $*$ are integers. It is now possible to use the same methods explained in the previous cases: We multiply the last rows and columns by m and divide the whole matrix by m. □

Remark. The case where we find a factor of m practically never happens. The reason is simply that the forms used to test the algorithm always have a determinant which is very hard to factor. So finding a factor in such a way is quite hopeless.

The corresponding algorithm is the following.

Algorithm 2.11 (Minimization 2).

Input: $Q \in \mathrm{Sym}(5, \mathbb{Z})$ such that $d_2(Q) \neq 1$ and $d_3(Q) = 1$, $m \neq 1 \in \mathbb{Z}$ dividing $d_2(Q)$.

Output: Q_f: a form equivalent form to Q;
G : the corresponding change of basis such that $m'Q_f = {}^tGQG$ with $1 < m' \mid m$.

1. Compute the SNF of Q with the algorithm described in [5].
2. Set $G := V$ and $Q := {}^tGQG$.
3. Let Q_3 be the 3×3 bottom-right submatrix of Q.
4. Apply a modified Gram-Schmidt orthogonalization process (see below) to Q_3 and m.
5. If the Gram-Schmidt process returns a vector, store it in S and go to step 10. If it returns an integer m', go back to step 4 with $m = m'$.
6. Denote by D_3 the returned matrix and by G_3 the corresponding change of basis.

7. Let $d = \gcd(D_3[1,1], m)$. If $d \neq 1$, go back to step 5 with $m = d$.
8. Use the Pollard-Schnorr algorithm [8] to solve

$$X^2 + \frac{D_3[2,2]}{D_3[1,1]} Y^2 \equiv -\frac{D_3[3,3]}{D_3[1,1]} \pmod{m}.$$

   Let S be a solution.
9. Set $S := [S, 1]$.
10. Let H be a 3×3 matrix whose first column is equal to S and whose columns form a $\mathbb{Z}^3$ basis. This can be done using the Hermite normal form algorithm.
11. Set $G_3 := G_3 \times H$.
12. Let $\widetilde{G}$ be the block-diagonal 5×5 matrix such that the 2×2 upper-left block is the identity and the 3×3 bottom-right block is equal to G_3.
13. Set $G := G \times \widetilde{G}$ and $Q' := (1/m)\, {}^tGQG$.
14. Apply the LLL algorithm to reduce Q', and denote by Q_f the returned form and by G' the corresponding change of basis.
15. Set $G := G \times G'$.
16. Return Q_f, G.

The minimization algorithm. We can now give the complete algorithm that minimizes an integral quadratic form of dimension 5.

Algorithm 2.12 (Minimization).

Input: $Q \in \mathrm{Sym}(5, \mathbb{Z})$.

Output: $Q_t \in \mathrm{Sym}^*(5, \mathbb{Z})$ equivalent to Q;
B : the corresponding change of basis.

1. Set $Q_t := Q$.
2. Compute the SNF D of Q.
3. If $d_1 = \det Q$, go to step 8.
4. If $d_5 \neq 1$ set $i := 5$.
5. Let $i \leq 5$ be such that $d_i \neq 1$ and $d_{i+1} = 1$ or $d_i = d_5$ if $d_5 \neq 1$.
6. Set $B := \mathrm{Id}_5$.
7. While $d_1 \neq \det Q_t$:
   (a) Switch according to i:
   Case $i = 5$: apply Algorithm 2.5 to Q_t and d_i.
   Case $i = 4$: apply Algorithm 2.7 to Q_t and d_i.
   Case $i = 3$: apply Algorithm 2.9 to Q_t and d_i.

Case $i = 2$: apply Algorithm 2.11 to Q_t and d_i.

(b) Let Q_f and G be the returned matrices.

(c) Set $Q_t := Q_f$ and $B := B \times G$.

(d) Compute the SNF D of Q_t.

(e) Let d_i be the diagonal coefficient of the SNF of Q_t such that $d_i \neq 1$ with $d_{i+1} = 1$ and $d_i = d_5$ if $d_5 \neq 1$.

8. Return Q_t, B.

Remark. This algorithm computes the Smith normal form at any step. To do this, it is strongly recommended to use the method described in [5] which is optimized and also gives the corresponding matrices U and V.

Remark. In this algorithm, we do not use a divisor m of d_i, but d_i itself. Using a divisor would force the algorithm to use factorization.

Remark. Algorithms 2.7, 2.9, and 2.11 include a reduction step using an LLL algorithm for indefinite quadratic forms given in [11]. This reduction is done to have concrete bounds for the size of the coefficients at the end of the algorithm.

2C. ***Reducing the even part of the determinant.*** After performing the *minimization step*, we get a form whose coefficient d_2 is equal to 1. We now need to have an equivalent form whose determinant is odd. This is performed by what we call the *reducing the even part step*.

Lemma 2.13. *Let $Q \in \mathrm{Sym}^*(5, \mathbb{Z})$ be indefinite. Let v be the quotient in the Euclidean division of the 2-adic valuation of* $\det Q$ *by 2. There exist two matrices Q' and G such that*

$$\begin{aligned} \det G &= \frac{1}{2^v}, \\ Q' &= {}^tGQG, \\ v_2(\det Q') &= 0 \text{ or } 1, \\ Q' &\in \mathrm{Sym}^*(5, \mathbb{Z}). \end{aligned}$$

Proof. If $\det Q$ is odd, we simply take $G = \mathrm{Id}_5$ and $Q' = Q$. Thus assume that $v_2(\det Q) \neq 0$. We compute the SNF of Q and obtain unimodular integer matrices U, V and a diagonal matrix D such that $D = UQV$, and $d_{1,1} = |\det Q|$. Since $d_2(Q) = 1$ the other diagonal coefficients of D are all equal to 1. We apply to Q the change of basis given by the matrix V. The first row and the first column of $Q'' = {}^tVQV$ are divisible by $2^{v_2(\det Q)}$. Let v be the quotient in the Euclidean division of the 2-adic valuation of $\det Q$ by 2, F be the diagonal matrix whose upper-left entry is equal to $1/2^v$ and the others equal to 1. If $v_2(\det Q)$ is even, the determinant of ${}^tFQ''F = Q'$ is odd. Otherwise the determinant of Q'

is divisible by 2 but not by 4. So we take $G = V \times F$. It remains to show that $Q' \in \mathrm{Sym}^*(5, \mathbb{Z})$. We know that $Q \in \mathrm{Sym}^*(5, \mathbb{Z})$. Since the change of basis given by the SNF is unimodular the invariant factors have not changed during the process. The last operation is done on the first column and only with a power of 2, so it also does not change the invariant factors, and so we have $Q' \in \mathrm{Sym}^*(5, \mathbb{Z})$. □

The corresponding algorithm is as follows.

Algorithm 2.14 (Reduction of the even part — I).

Input: $Q \in \mathrm{Sym}^*(5, \mathbb{Z})$ indefinite, of dimension 5, of determinant Δ.

Output: $Q' \in \mathrm{Sym}^*(5, \mathbb{Z})$ indefinite, of determinant $2^k n$ with n odd and $k \equiv v_2(\det Q) \pmod 2$, Q' equivalent to Q;
G the corresponding change of basis.

1. If $\Delta \equiv 1 \pmod 2$, return Q, Id_5.
2. Set $G := \mathrm{Id}_5$.
3. Let v_2 be the 2-adic valuation of Δ.
4. Let v be the quotient in the Euclidean division of v_2 by 2.
5. Let U, V and D be the matrices given by the SNF of Q such that $D = UQV$.
6. Set $Q' := {}^tVQV$ and $G := G \times V$.
7. Let H be the diagonal matrix such that $H_{1,1} = 1/2^v$ and $H_{i,i} = 1$ otherwise.
8. Set $Q' := {}^tHQ'H$ and $G := G \times H$.
9. Return Q', G.

Lemma 2.15. *Let $Q \in \mathrm{Sym}^*(5, \mathbb{Z})$ indefinite and such that $\det Q = 2k$, $k \in \mathbb{Z}$, odd. There exist two matrices Q' and G such that*

$$\det G = \frac{1}{2^3},$$
$$Q' = 2 \times {}^tGQG,$$
$$\det Q' \equiv k \pmod 2.$$

Proof. As in proof of the previous lemma, we begin by computing the Smith normal form of Q to obtain integer matrices U, V unimodular and D diagonal such that $D = UQV$ and $d_{1,1} = |\det Q|$. We apply to Q the change of basis given by the matrix V and obtain Q' which has the following form:

$$Q' = {}^tVQV = \begin{bmatrix} 2* & 2* & & \\ & Q_1 & * & * \\ 2* & & * & * \\ & & * & * \end{bmatrix}.$$

We are now interested in the form Q_1 which is the restriction of the form Q to the subspace generated by the second and third vectors of the basis. Denote this form by the following matrix: $\left[\begin{smallmatrix} a & b \\ b & c \end{smallmatrix}\right]$. We are looking for a change of basis such that the coefficient a in the new basis will be even. This means that we want a pair (x, y) such that $ax^2 + cy^2 \equiv 0 \pmod 2$. We solve this equation, apply the corresponding change of basis to Q_1, and we multiply the whole matrix by 2. The determinant of the form is now divisible by 2^6 but not by 2^7. We rescale the first two vectors by a factor 2. The determinant is now divisible by 2^2. We then compute the SNF of this matrix and apply the change of basis according to the matrix V. Since the determinant is divisible by 4, we have two possibilities: If the kernel modulo 2 has dimension 1, the first row and the first column are divisible by 2 and the upper left coefficient is divisible by 4. In this case, we rescale the first vector by 2. Otherwise, the kernel has dimension 2. In this case, the first two rows and columns are divisible by 2. Consider the upper-left 2×2 block of the matrix. This corresponds to the restriction of the form to the subspace generated by the first two vectors of the basis. We are going to apply a change of basis such that the upper-left coefficient will be divisible by 4. This corresponds to solving the equation $ax^2 + cy^2 \equiv 0 \pmod 2$ which can be done as explained above. Once the change of basis is done, we simply rescale the first vector by 2. In such a basis, the determinant of the form is now odd. It remains to show that this form belongs to $\mathrm{Sym}^*(5, \mathbb{Z})$. Indeed, since the determinants of the changes of basis that we have applied are all equal to a power of 2 they are invertible modulo the odd primes factors of the determinant of the form, and it follows that the rank of the form is unchanged, so we have $Q' \in \mathrm{Sym}^*(5, \mathbb{Z})$. □

The corresponding algorithm is as follows.

Algorithm 2.16 (Reduction of the even part — II).

Input: $Q \in \mathrm{Sym}^*(n, \mathbb{Z})$ indefinite, with $\det Q = \Delta = 2^k n$ with n odd and $k = 0$ or 1.

Output: Q', a form in $\mathrm{Sym}^*(5, \mathbb{Z})$ with odd determinant and same solutions as Q up to a change of basis;
G the corresponding change of basis.

1. If $\Delta \equiv 1 \pmod 2$ return Q, Id_5.
2. Set $G := \mathrm{Id}_5$.
3. Let v be the 2-adic valuation of Δ.
4. Let U, V and D be the matrices given by the SNF of Q such that $D = UQV$.
5. Set $Q' := {}^tVQV$ and $G := G \times V$.
6. If $(q'_{2,2}, q'_{3,3}) \equiv (1, 1) \pmod 2$,

(a) set $H := \mathrm{Id}_5$ and $H[3,2] := 1$,
(b) set $Q' := {}^tHQ'H$ and $G := G \times H$.

7. If $(q'_{2,2}, q'_{3,3}) \equiv (1,0) \pmod 2$,
(a) set $H := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$,
(b) set $Q' := {}^tHQ'H$ and $G := G \times H$.
8. Set $Q' := 2 \times Q'$.
9. Set $P := \mathrm{Id}_5$ and $P[2,2] := 1/2$.
10. Set $Q' := {}^tPQ'P$ and $G := G \times P$.
11. Let U', V' and D' be the matrices given by the SNF of Q' such that $D = UQ'V$.
12. Set $Q' := {}^tV'Q'V'$ and $G := G \times V'$.
13. If $q'_{1,1} \equiv 0 \pmod 4$,
(a) set $R := \mathrm{Id}_5$ and $R[1,1] := 1/2$,
(b) set $Q' := {}^tRQ'R$ and $G := G \times R$,
(c) return Q', G.
14. Repeat steps 6 to 2 with $(q'_{1,1}, q'_{2,2})$.
15. Set $R := \mathrm{Id}_5$ and $R[1,1] := 1/2$.
16. Set $Q' := {}^tRQ'R$ and $G := G \times R$.
17. Return Q', G.

2D. ***Completion.*** We now explain how to complete the form to a form of dimension 6 in the way announced in Section 2A, and in particular how to choose the value of z. Controlling this value will allow us to change the signature of the completed form Q_6.

Lemma 2.17. *Let $Q \in \mathrm{Sym}(5,\mathbb{Z})$ be an indefinite form with signature (r,s) and determinant Δ. Let X be a 5-dimensional column vector with integral entries and $\bar{\beta}$ be a coset representative of the coset of ${}^tX\,\mathrm{Co}(Q)X$ modulo Δ. Let*

$$z := \frac{{}^tX\,\mathrm{Co}(Q)X - \bar{\beta}}{\Delta}$$

and

$$Q_6 = \begin{bmatrix} Q & X \\ {}^tX & z \end{bmatrix}.$$

The signature of Q_6 is determined by the signs of $\overline{\beta}$ and $\det Q$ *as follows*:

$$\textit{signature of } Q_6 = \begin{cases} (r, s+1) & \textit{if } \overline{\beta} \det Q > 0; \\ (r+1, s) & \textit{if } \overline{\beta} \det Q < 0. \end{cases}$$

Moreover we have $\overline{\beta} = -\det Q_6$.

Proof. As seen in Section 2A, the formula (2) gives us the determinant of the form Q_6:

$$\det Q_6 = z \det Q - {}^tX \operatorname{Co}(Q) X.$$

We also have defined the quantities: $\beta = {}^tX \operatorname{Co}(Q) X$ and $\overline{\beta}$ a coset representative of the coset of β modulo Δ which is also equal to $\beta - z\Delta = -\det Q_6$. Since the link between Q and Q_6 is the addition of a row and a column, if we consider the restriction of Q to the subspace generated by the first 5 vectors of the basis, we get back exactly the form Q. Thus if we add a row and a column, we do not change its signature on this subspace. It follows that we can deduce the signature of Q_6 from the signature of Q by simply considering the sign of their determinant. Indeed, we know that $\operatorname{sgn}(\det Q) = (-1)^s$. If $\det Q > 0$, we have $s \equiv 0 \pmod 2$. We take $\overline{\beta} > 0$ and have $\det Q_6 < 0$. We have changed the sign of the determinant, so the signature of Q_6 is $(r, s+1)$. The others cases are done in the same way, and combining them gives the formula for the signature given in the lemma. □

In order to be able to compute a solution, we need the signature (u, v) of Q_6 to satisfy $u \geq 2$ and $v \geq 2$. The following algorithm will choose the value of $\overline{\beta}$ so that this is satisfied. The algorithm for completing the form and controlling the signature is the following.

Algorithm 2.18 (Completion).

Input: Q: an indefinite, nondegenerate dimension 5 integral quadratic form; $k \geq 1$ an integer.

Output: Q_6: an indefinite, nondegenerate dimension 6 integral quadratic form with signature (r, s) such that $r \geq 2$ and $s \geq 2$, of the form: $\begin{bmatrix} Q & X \\ {}^tX & z \end{bmatrix}$, and such that $|\det Q_6| < k |\det Q|$.

1. Compute the signature (r, s) of Q.
2. Choose an integer vector X whose coordinates are nonnegative integers less than $|\det Q|^5$.
3. Set $\beta := {}^tX \operatorname{Co}(Q) X$ and $\overline{\beta} := \beta \pmod{\det Q}$ with $0 \leq \overline{\beta} < |\det Q|$.
4. If $r = 1$ and $\det Q > 0$, set $\overline{\beta} := \overline{\beta} - |\det Q|$.
5. If $s = 1$, set $\overline{\beta} := \overline{\beta} - \det Q$.
6. Set $z := \dfrac{\beta - \overline{\beta}}{\det Q}$.

7. Add a random multiple of $|\det Q|$ to $\bar{\beta}$ so that $|\det Q_6| < k\,|\det Q|$ while respecting the signature condition, and update the value of z.
8. Return $Q_6 = \begin{bmatrix} Q & X \\ {}^tX & z \end{bmatrix}$.

Remark. The bounds on X in step 2 are chosen in this way since everything is then reduced modulo $\det Q$. Changing the bounds would not change the complexity of the whole algorithm.

Remark. At the end of the algorithm, the determinant of Q_6 is always equal to $\bar{\beta}$. This is a consequence of the choice of the value of z.

Remark. We will use this algorithm until we obtain a $\bar{\beta}$ of the form $2 \times p$ with p an odd prime number. This choice will be explained in Section 2E.

2E. ***Computing a solution.*** The complete algorithm for finding a nonzero isotropic vector for a quadratic form dimension 5 without factoring the determinant is as follows.

Algorithm 2.19 (Solving).

Input: Q, an integral indefinite, nondegenerate quadratic form of dimension 5.

Output: X, a nonzero integral isotropic vector for Q.

1. Apply the minimization Algorithm 2.12 to Q.
2. Apply Algorithms 2.14 and 2.16 to the result of step 1.
3. Apply the completion Algorithm 2.18 to the result of step 2 until the determinant of the returned form Q_6 is equal to $\pm 2p$ where p is an odd prime number.
4. Solve the equation ${}^tXQ_6X = 0$.
5. Write $Q_6 = H \oplus Q_4$ where H is a hyperbolic plane.
6. Solve the equation ${}^tXQ_4X = 0$.
7. Write $Q_4 = H' \oplus Q_2$ where H' is a hyperbolic plane.
8. Deduce from the previous steps a solution S to the equation ${}^tXQX = 0$.
9. Return S.

Theorem 2.20. *Let Q be an integral indefinite, nondegenerate quadratic form of dimension 5. Then Algorithm 2.19, applied to Q, outputs a nonzero integral vector S that is a solution to the equation ${}^tXQX = 0$ without factorizing any integer.*

Remark. The above algorithm is based on the fact that the method developed by Simon in [11] is very efficient as soon as the factorization of the determinant of the form is known. This theorem shows that there exists an efficient algorithm even when the factorization is not known or when it is not possible to factor the determinant in a reasonable amount of time.

Proof. This proof follows the steps of the algorithm. We are going to divide the proof in the same way as the algorithm is divided:

1: Minimizations

2: Reducing the even part

3: Choice of the signature and completion of Q while imposing the form of the determinant

4: Computing a solution for Q_6

5: Decomposition in a sum with a hyperbolic plane

6: Computing a solution for Q_4

7: Decomposition in a sum with a hyperbolic plane

8: Computing a solution for Q

Step 1: We apply Algorithm 2.12 to Q. At the end of this step, we have a form $Q^{(2)} \in \mathrm{Sym}^*(5,\mathbb{Z})$ equivalent to Q, an invertible matrix G_2, and a nonzero rational number $\lambda^{(2)}$ such that $Q^{(2)} = \lambda^{(2)}\,{}^tG_2 Q G_2$.

Step 2: We successively apply Algorithms 2.14 and 2.16 to $Q^{(2)}$ in order to have a form with an odd determinant. At the end of this step, we obtain a form $Q^{(3)} \in \mathrm{Sym}^*(5,\mathbb{Z})$ equivalent to Q, an invertible matrix G_3, and a nonzero rational number $\lambda^{(3)}$ such that $Q^{(3)} = \lambda^{(3)}\,{}^tG_3 Q^{(2)} G_3$ and the determinant Δ of $Q^{(3)}$ is odd.

Step 3: We apply Algorithm 2.18 and choose $k = 10^6$ (the value of k will be detailed in a further paper) until the determinant of the returned form is equal to $\pm 2p$ with p an odd prime number; the condition $2 \times p$ is necessary because of some conditions on local solubility at 2. It is possible to show that a vector X verifying these conditions can always be found efficiently by using an effective version of the Cebotarev density theorem [6]. At the end of this step, we have a form Q_6 whose restriction to the subspace generated by the first 5 vectors of the basis is equal to $Q^{(3)}$, whose determinant is equal to $\pm 2p$ with p an odd prime number, and whose signature (r,s) is such that $r \geq 2$ and $s \geq 2$.

Step 4: We use the algorithm described in [11], and obtain a nonzero integral vector T such that ${}^tTQ_6T = 0$. We divide T by the GCD of its coordinates in order to have T primitive.

Step 5: This step consists in finding a hyperbolic plane containing the vector T. The existence of such a plane is given by the result in [9, p.55, Proposition 3.]. We first write the form Q_6 in a unimodular basis whose first vector is the vector T (the basis can be found by using the HNF of a primitive vector), we denote by G_4 such a change of basis. We then have $Q_6^{(1)} = {}^tG_4 Q_6 G_4$ and the upper-left coefficient is 0. Let $R = \left(Q_6^{(1)}[1,2], Q_6^{(1)}[1,3], Q_6^{(1)}[1,4], Q_6^{(1)}[1,5], Q_6^{(1)}[1,6]\right)$, and let G_5

be a unimodular matrix such that $RG_5 = (a, 0, 0, 0, 0)$, where a is the GCD of the coefficients of the vector R. Since a divides the first row and the first column of the matrix $Q_6^{(1)}$ we have $a^2 \mid \det Q_6^{(1)}$, but since $\det Q_6^{(1)} = \pm 2p$ with p prime, we must therefore have $a = 1$. Such a G_5 matrix is given by the HNF of the vector R. We can now set $G_6 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & G_5 \end{smallmatrix}\right]$, and we then have

$$Q_6^{(2)} = {}^tG_6 Q_6^{(1)} G_6 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ 0 & b_3 & * & * & * & * \\ 0 & b_4 & * & * & * & * \\ 0 & b_5 & * & * & * & * \\ 0 & b_6 & * & * & * & * \end{bmatrix}.$$

Now let G_7 be the following matrix:

$$G_7 = \begin{bmatrix} 1 & \left[\frac{-b_2}{2}\right] & -b_3 & -b_4 & -b_5 & -b_6 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We have $\det G_7 = 1$, and

$$Q_6^{(3)} = {}^tG_7 Q_6^{(2)} G_7 = \left[\begin{array}{cc:cccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & \alpha & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & \multicolumn{4}{c}{\multirow{3}{*}{Q_4}} \\ 0 & 0 & & & & \\ 0 & 0 & & & & \\ 0 & 0 & & & & \end{array}\right],$$

where $Q_4 \in \mathrm{Sym}(4, \mathbb{Z})$. We also have $\det Q_4 = -\det Q_6$. The coefficient in this matrix is either 0 or 1 according to the parity of the coefficient b_2, but it will not change anything in the rest of the algorithm. We regroup all the changes of basis and set $G_8 = G_4 \times G_6 \times G_7$. We then have $Q_6^{(3)} = {}^tG_8 Q_6 G_8$. This step ends with the computation of the matrices $Q_6^{(3)}$ and G_8.

Step 6: We now work on the quadratic form Q_4 defined above. Its determinant is $-\det Q_6$, which is still equal to $\mp 2p$ with p a prime number. We are going to show that the equation ${}^tXQ_4X = 0$ has a nontrivial solution: We know that Q_4 is indefinite; indeed, the form $Q^{(3)}$ has been completed in order to have $r \geq 2$ and $s \geq 2$. We have decomposed this form into the sum of a hyperbolic plane and a dimension 4 quadratic form Q_4, but the signature of a quadratic form on a

hyperbolic plane is $(1,1)$, and $Q_6^{(3)}$ has the same signature as Q_6, so the signature Q_4 is $(r-1, s-1)$ and we have $r-1 \geq 1$, $s-1 \geq 1$, showing that Q_4 is indefinite hence that there exists real solutions. We now need to show the existence of a solution over $\mathbb{Q}_\ell$ for every prime number ℓ. If ℓ is an odd prime number not dividing $\det Q_4$, the consideration of Hilbert symbols shows that solutions always exist. Two cases remain: $\ell = 2$ and $\ell \mid \det Q_4$. We know that $\det Q_4 = \pm 2p$ is not a square neither in $\mathbb{Q}_2$ nor in $\mathbb{Q}_p$ since the valuations are odd and $p \neq 2$, so there exist local solutions, and using the local-global principle allows us to conclude. Since solutions exist, we can now use Simon's algorithm to compute such a solution, and since the determinant is equal to $\pm 2p$ with p prime, we do not need to use any factorization. We denote by R a primitive solution.

Step 7: This step is the same as the step 5, but the work is done over the form $Q_4^{(1)}$. Let B be the corresponding change of basis.

Step 8: We have to recall the changes of basis done on the matrix Q_4. We set

$$G_9 = \left[\begin{array}{cc|cccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & \multicolumn{4}{c}{\multirow{4}{*}{B}} \\ 0 & 0 & & & & \\ 0 & 0 & & & & \\ 0 & 0 & & & & \end{array}\right]$$

and

$$P = G_8 \times G_9.$$

We thus have a matrix P such that

$${}^tPQ_6P = \left[\begin{array}{cc|cc|c} 0 & 1 & \multicolumn{2}{c|}{\multirow{2}{*}{0}} & \multirow{2}{*}{0} \\ 1 & \alpha & \multicolumn{2}{c|}{} & \\ \hline \multicolumn{2}{c|}{\multirow{2}{*}{0}} & 0 & 1 & \multirow{2}{*}{0} \\ \multicolumn{2}{c|}{} & 1 & \beta & \\ \hline 0 & & 0 & & Q_2 \end{array}\right]$$

with $\alpha, \beta = 0$ or 1. We note that the first and the third columns of P are solutions of the equation ${}^tXQ_6X = 0$. But they also are orthogonal vectors for Q_6. It follows that every linear combination of these vectors still is a solution for Q_6. We now consider a combination such that the last coordinate is 0, denote it by J. We then have

$$J = \begin{bmatrix} U \\ \hline 0 \end{bmatrix} \quad \text{with } U \in \mathbb{Z}^5.$$

We know that ${}^tJQ_6J = 0$, but we give the computation in detail:

$$\begin{aligned} {}^tJQ_6J &= \begin{bmatrix} {}^tU & 0 \end{bmatrix} \left[\begin{array}{c|c} Q^{(3)} & X \\ \hline {}^tX & z \end{array}\right] \begin{bmatrix} U \\ \hline 0 \end{bmatrix} \\ &= {}^tUQ^{(3)}U \\ &= 0. \end{aligned}$$

Thus U is a nonzero solution to the equation ${}^tXQ^{(3)}X = 0$. We then set $S = G_2G_3U$, and we have ${}^tSQS = 0$. We are finally done. □

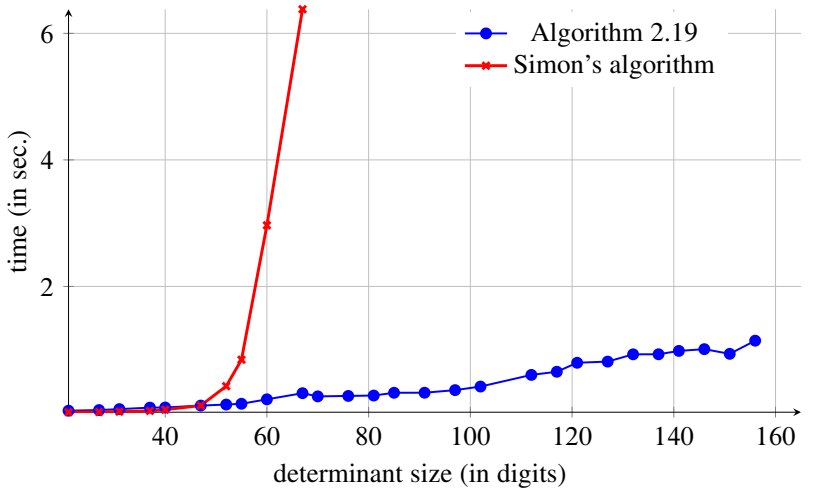
Remark. The condition of having the determinant equal to $\pm 2 \times p$ with p an odd prime is necessary due to the condition of local solubility over $\mathbb{Q}_2$. The 2 can be replaced by 2^{2k+1} with $k \in \mathbb{N}$, but the analysis is much more complicated in this case and it practically does not affect the running time of the algorithm.

Remark. The complexity of the algorithm is not done here, but the number of vectors X that we need to try in step 3 until we have a determinant of the desired shape is $\mathcal{O}(\log|\det Q|)$.

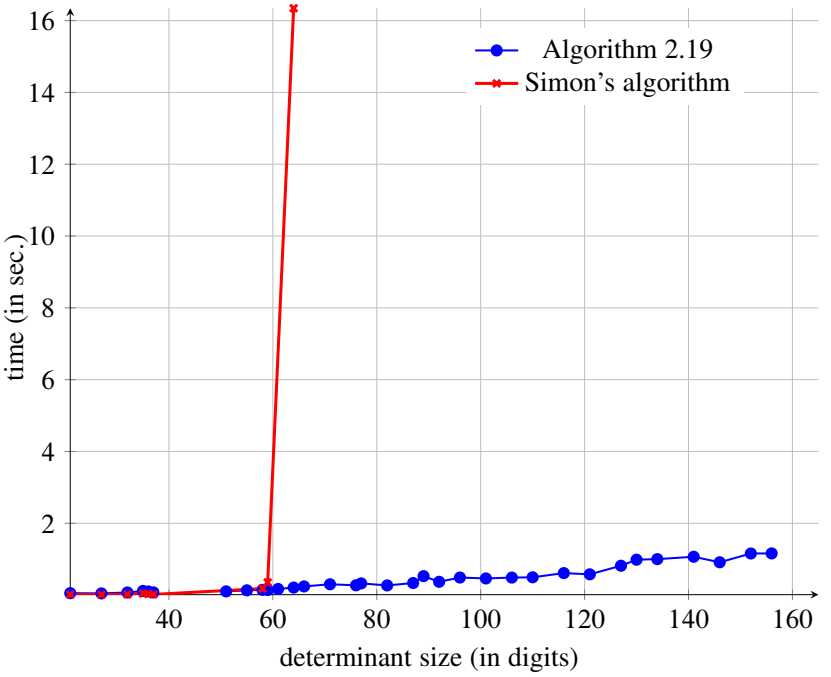
2F. ***Generalization to higher dimensions.*** The algorithm given above is for quadratic forms of dimension 5. It is easy to generalize it to higher dimensions: Indeed, since the algorithm needs a form of dimension 5 as an input, if the given form has a larger dimension, we simply need to restrict the form to a subspace of dimension 5. The only condition required is that the restriction of the form must have a signature (r, s) that verifies $r \geq 1$ and $s \geq 1$ so that the decomposition as the sum of two hyperbolic planes is possible. When a solution to the restriction is found, we simply lift the solution to the original space by setting the remaining coordinates to 0.

3. Overview of performance

This algorithm has been implemented in the PARI/GP language, see [7]. Since the proof of the complexity of this algorithm requires a considerable amount of additional work it will not be detailed here, but will be explained in a further work. However, we give an overview of the global performances of the algorithm with the two following figures. The comparisons are made with the method given by Simon in [11] and [10]. These algorithms have also been implemented in the PARI/GP language and can downloaded from the author's webpage (http://www.math.unicaen.fr/~simon).



These values have been computed by averaging over 100 random forms for each point. The forms are the same for each algorithm. We can clearly observe the fact that the factorization of the determinant makes Simon's algorithm very slow for determinants with size larger than 50 digits. The graph below shows the same comparison, but this time, the method used for building the forms is made in such a way that the algorithm often needs to do minimizations. We still can see the "wall" due to the factorization of the determinant in Simon's method.



Acknowledgments

I thank Henri Cohen for his help with translation and the anonymous reviewers for their helpful comments.

References

[1] Pierre Castel, *Un algorithme de résolution des équations quadratiques en dimension* 5 *sans factorisation*, Ph.D. thesis, Laboratoire de Mathématiques Nicolas Oresme, 2011. http://www.math.unicaen.fr/~castel/production/these_castel.pdf

[2] Henri Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics, no. 138, Springer, Berlin, 1993. MR 94i:11105

[3] J. E. Cremona and D. Rusin, *Efficient solution of rational conics*, Math. Comp. **72** (2003), no. 243, 1417–1441. MR 2004a:11137

[4] A. Fröhlich (ed.), *Algebraic number fields*: *L-functions and Galois properties*: *Proceedings of a Symposium held at the University of Durham*, *Sept.* $2--12$, 1975, Academic Press, London, 1977. MR 55 #10416

[5] Costas S. Iliopoulos, *Worst-case complexity bounds on algorithms for computing the canonical structure of finite abelian groups and the Hermite and Smith normal forms of an integer matrix*, SIAM J. Comput. **18** (1989), no. 4, 658–669. MR 91a:20065

[6] J. C. Lagarias and A. M. Odlyzko, *Effective versions of the Chebotarev density theorem*, in Fröhlich [4], 1977, pp. 409–464. MR 56 #5506

[7] PARI Group, Bordeaux, France, *PARI/GP (version* 2.5.0*)*, 2011. http://pari.math.u-bordeaux.fr/

[8] John M. Pollard and Claus-P. Schnorr, *An efficient solution of the congruence* $x^2 + ky^2 = m \pmod{n}$, IEEE Trans. Inform. Theory **33** (1987), no. 5, 702–709. MR 89e:11080

[9] Jean-Pierre Serre, *Cours d'arithmétique*, Presses Universitaires de France, Paris, 1988.

[10] Denis Simon, *Quadratic equations in dimension* 4, 5 *and more*, preprint, 2005. http://www.math.unicaen.fr/~simon/maths/Dim4.pdf

[11] ———, *Solving quadratic equations using reduced unimodular quadratic forms*, Math. Comp. **74** (2005), no. 251, 1531–1543. MR 2005k:11246

PIERRE CASTEL: pierre.castel@unicaen.fr
Laboratoire de Mathématiques Nicolas Oresme, Université de Caen Basse-Normandie, UMR CNRS 6139, 14032 Caen, France

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.235

msp

Counting value sets: algorithm and complexity

Qi Cheng, Joshua E. Hill, and Daqing Wan

Let p be a prime. Given a polynomial in $\mathbb{F}_{p^m}[x]$ of degree d over the finite field $\mathbb{F}_{p^m}$, one can view it as a map from $\mathbb{F}_{p^m}$ to $\mathbb{F}_{p^m}$, and examine the image of this map, also known as the *value set* of the polynomial. In this paper, we present the first nontrivial algorithm and the first complexity result on explicitly computing the cardinality of this value set. We show an elementary connection between this cardinality and the number of points on a family of varieties in affine space. We then apply Lauder and Wan's p-adic point-counting algorithm to count these points, resulting in a nontrivial algorithm for calculating the cardinality of the value set. The running time of our algorithm is $(pmd)^{O(d)}$. In particular, this is a polynomial-time algorithm for fixed d if p is reasonably small. We also show that the problem is #P-hard when the polynomial is given in a sparse representation, $p = 2$, and m is allowed to vary, or when the polynomial is given as a straight-line program, $m = 1$ and p is allowed to vary. Additionally, we prove that it is NP-hard to decide whether a polynomial represented by a straight-line program has a root in a prime-order finite field, thus resolving an open problem proposed by Kaltofen and Koiran.

1. Introduction

Let $f \in \mathbb{F}_q[x]$ be a polynomial of degree d with coefficients in a finite field having $q = p^m$ elements, where p is prime. Denote the image set of this polynomial by

$$V_f = \{f(\alpha) \mid \alpha \in \mathbb{F}_q\}$$

and denote the cardinality of this set by $\#(V_f)$.

There are a few trivial bounds on $\#(V_f)$ that can be immediately established. There are only q elements in the field, so $\#(V_f) \leq q$. Additionally, any polynomial

MSC2010: primary 11Y16; secondary 11Y40, 68Q17.
Keywords: finite field, polynomial value set cardinality, point counting, polynomial time, randomized polynomial time, RP-reduction, NP-hard, #P-hard, straight-line program, sparse polynomial, subset sum problem.

of degree d can have at most d roots, thus for all $a \in V_f$, $f(x) = a$ is satisfied at most d times. This is true for every element in V_f, so $\#(V_f)d \geq q$, whence

$$\left\lceil \frac{q}{d} \right\rceil \leq \#(V_f) \leq q,$$

where $\lceil \cdot \rceil$ is the ceiling function.

Both of these bounds can be achieved: If $\#(V_f) = q$, then f is called a *permutation polynomial*, and if $\#(V_f) = \lceil q/d \rceil$, then f is said to have a *minimal value set*.

The problem of computing $\#(V_f)$ has been studied in various forms for at least the last 115 years, but exact formulas for $\#(V_f)$ are known only for polynomials of very specific forms. Results that apply to general polynomials are asymptotic in nature, or provide estimates whose errors have reasonable bounds only on average [14].

The fundamental problem of determining the value set cardinality $\#(V_f)$ can be thought of as a much more general version of the problem of determining whether a particular polynomial is a permutation polynomial. Shparlinski [17] provides a baby-step giant-step type test that determines if a given polynomial is a permutation polynomial by extending the ideas in [20] to an algorithm that runs in time $\tilde{O}((dq)^{6/7})$. This is still fully exponential in $\log q$. Ma and von zur Gathen [13] provide a ZPP (zero-error probabilistic polynomial-time) algorithm for testing if a given polynomial is a permutation polynomial. According to [10], the first deterministic polynomial-time algorithm for testing permutation polynomials was obtained by Lenstra using the classification of exceptional polynomials, which in turn depends on the classification of finite simple groups. Subsequently, an elementary approach based on the Gao-Kaltofen-Lauder factorization algorithm was given by Kayal [10].

Essentially nothing is known about the complexity of the more general problem of exactly computing $\#(V_f)$, and no nontrivial algorithms for this problem are known. For instance, no baby-step giant-step type algorithm for computing $\#(V_f)$ is known, and no probabilistic polynomial-time algorithm for this problem is known. Finding a nontrivial algorithm and proving a nontrivial complexity result for the value counting problem were raised as open problems in [13], where a probabilistic approximation algorithm is given. In this paper, we provide the first nontrivial algorithm and the first nontrivial complexity result for the exact counting of the value set problem.

1A. ***Our results.*** Perhaps the most obvious method to calculate $\#(V_f)$ is to evaluate the polynomial at each point in $\mathbb{F}_q$ and count how many distinct images result. This algorithm has a time and space complexity $(dq)^{O(1)}$. One can also approach this problem by operating on points in the codomain. One has $f(x) = a$ for some

$x \in \mathbb{F}_q$ if and only if $f_a(X) = f(X) - a$ has a zero in $\mathbb{F}_q$; this algorithm again has a time complexity $(dq)^{O(1)}$, but the space complexity is improved considerably to $(d \log q)^{O(1)}$.

In this paper we present several results on determining the cardinality of value sets. On the algorithmic side, we show an elementary connection between this cardinality and the number of points on a family of varieties in affine space. We then apply Lauder and Wan's p-adic point-counting algorithm [12], resulting in a nontrivial algorithm for calculating the image set cardinality in the case that p is sufficiently small (that is, $p = O((d \log q)^C)$ for some positive constant C). Precisely, we have the following.

Theorem 5.2. *There exists an explicit deterministic algorithm and an explicit polynomial R such that for any $f \in \mathbb{F}_q[x]$ of degree d, where $q = p^m$ (p prime), the algorithm computes $\#(V_f)$, the cardinality of the image set, in a number of bit operations bounded by $R(m^d d^d p^d)$.*

The running time of this algorithm is polynomial in both p and m, but is exponential in d. In particular, this is a polynomial-time algorithm for fixed d if the characteristic p is small: $q = p^m$ can be large, but $p = O((d \log q)^C)$.

On the complexity side, we have several hardness results on the value set problem. We frame these results using some standard classes in complexity theory, which we outline here. NP is the complexity class of decision problem whose positive solutions can be verified in polynomial time. NP-hard is the computational class of decision problems that all NP problems can be reduced to using a polynomial-time reduction. NP-complete is the complexity class of all NP-hard problems whose solution can be verified in polynomial time (that is, NP-complete is the intersection of NP-hard and NP). Co-NP-complete is the complexity class of problems where answering the logical complement of the decision problem is NP-complete.

The corresponding counting complexity theory classes that we use are as follows. #P (read "sharp-P") is the set of counting problems whose corresponding decision problem is in NP. #P-hard is the computational class of counting problems that all #P problems can be reduced to using a polynomial-time counting reduction. #P-complete is the intersection of #P-hard and #P.

With a field of characteristic 2, we have the following.

Theorem 4.3. *The problem of counting the value set of a sparse polynomial over a finite field of characteristic* 2 *is #P-hard.*

The central approach in our proof of this theorem is to reduce the problem of counting satisfying assignments for a 3SAT formula to the problem of value set counting.

Over a prime-order finite field, we have the following.

Theorem 4.6. *Over a prime-order finite field $\mathbb{F}_p$, the problem of counting the value set is #P-hard under RP-reduction (randomized polynomial-time reduction) if the polynomial is given as a straight-line program.*

Additionally, we prove that it is NP-hard to decide whether a polynomial in $\mathbb{Z}[x]$ represented by a straight-line program has a root in a prime-order finite field, thus resolving an open problem proposed in [7; 8]. We accomplish the complexity results over prime-order finite fields by reducing the prime-order finite field subset sum problem (PFFSSP) to these problems.

In the PFFSSP, given a prime p, an integer b, and a set of integers $S = \{a_1, a_2, \ldots, a_t\}$, we want to decide the solvability of the equation

$$a_1x_1 + a_2x_2 + \cdots + a_tx_t \equiv b \pmod{p}$$

with $x_i \in \{0, 1\}$ for $1 \le i \le t$. The main idea comes from the observation that if $t < \log p/3$, there is a sparse polynomial $\alpha(x) \in \mathbb{F}_p[x]$ such that as x runs over $\mathbb{F}_p$, the vector

$$\big(\alpha(x), \alpha(x+1), \ldots, \alpha(x+t-1)\big)$$

runs over all the elements in $\{0, 1\}^t$. In fact, a lightly modified version of the quadratic character $\alpha(x) = (x^{(p-1)/2} + x^{p-1})/2$ suffices. So the PFFSSP can be reduced to deciding whether the sparse shift polynomial $\sum_{i=0}^{t-1} a_{i+1}\alpha(x+i) - b = 0$ has a solution in $\mathbb{F}_p$.

2. Background

2A. ***The subset sum problem.*** To prove the complexity results, we use the subset sum problem (SSP) extensively. The SSP is a well-known problem in computer science; we describe three versions of it. Let an integer b and a set of positive integers $S = \{a_1, a_2, \ldots, a_t\}$ be given.

(1) Decision version: The goal is to decide whether there exists a subset $T \subseteq S$ such that the sum of all the integers in T equals b.

(2) Search version: The goal is to find a subset $T \subseteq S$ such that the sum of all the integers in T equals b.

(3) Counting version: The goal is to count the number of subsets $T \subseteq S$ such that the sum of all the integers in T equals b.

The decision version of the SSP is a classical NP-complete problem. The counting version of the SSP is #P-complete, which can be easily derived from proofs of the NP-completeness of the decision version, for example [5, Theorem 34.15].

One can view the SSP as a problem of solving the linear equation

$$a_1x_1 + a_2x_2 + \cdots + a_tx_t = b$$

with $x_i \in \{0, 1\}$ for $1 \le i \le t$. The prime-order finite field subset sum problem is a similar problem where in addition to b and S, one is given a prime p, and the goal is to decide the solvability of the equation

$$a_1 x_1 + a_2 x_2 + \cdots + a_t x_t \equiv b \pmod{p}$$

with $x_i \in \{0, 1\}$ for $1 \le i \le t$.

Proposition 2.1. *The prime-order finite field subset sum problem is NP-hard under RP-reduction.*

Proof. To reduce the subset sum problem to the prime-order finite field subset sum problem, one finds a prime $p > \sum_{i=1}^{t} a_i$, which can be done in randomized polynomial time. □

Remark. To make the reduction deterministic, one needs to derandomize the problem of finding a large prime, which appears to be difficult [18].

2B. ***Polynomial representations.*** There are different ways to represent a polynomial over a field $\mathbb{F}$. The dense representation lists all the coefficients of a polynomial, including the zero coefficients. The sparse representation lists only the nonzero coefficients, along with the degrees of the corresponding terms. If most of the coefficients of a polynomial are zero, then the sparse representation is much shorter than the dense representation. A sparse shift representation of a polynomial in $\mathbb{F}[x]$ is a list of n triples $(a_i, b_i, e_i) \in \mathbb{F} \times \mathbb{F} \times \mathbb{Z}_{\ge 0}$ which represents the polynomial

$$\sum_{1 \le i \le n} a_i (x + b_i)^{e_i}.$$

More generally, a straight-line program for a univariate polynomial in $\mathbb{Z}[x]$ or $\mathbb{F}_p[x]$ is a sequence of assignments, starting from $x_1 = 1$ and $x_2 = x$. After that, the i-th assignment has the form

$$x_i = x_j \odot x_k$$

where $0 \le j, k < i$ and $\odot$ is one of the three operations $+$, $-$, $\times$. We first let α be an element in $\mathbb{F}_{p^m}$ such that $\mathbb{F}_{p^m} = \mathbb{F}_p[\alpha]$. A straight-line program for a univariate polynomial in $\mathbb{F}_{p^m}[x]$ can be defined similarly, except that the sequence starts from $x_1 = \alpha$ and $x_2 = x$. One can verify that a straight-line program computes a univariate polynomial, and that sparse polynomials and sparse shift polynomials have short straight-line programs. A polynomial produced by a short straight-line program may have very high degree, and most of its coefficients may be nonzero, so it may be costly to write it in either a dense form or a sparse form.

3. Hardness of solving straight-line polynomials

It is known that deciding whether there is a root in a finite field for a sparse polynomial is NP-hard [11]. In a related work, it was shown that deciding whether there is a p-adic rational root for a sparse polynomial is NP-hard [1]. However, the complexity of deciding the solvability of a straight-line polynomial in $\mathbb{Z}[x]$ within a prime-order finite field was not known. This open problem was proposed in [7] and [8]. We resolve this problem within this section, and this same idea will be used later on to prove the hardness result of the value set counting problem.

Let p be an odd prime. Let χ be the quadratic character modulo p; that is, $\chi(x)$ equals 1, -1, or 0, depending on whether x is a quadratic residue, a quadratic nonresidue, or is congruent to 0 modulo p. For $x \in \mathbb{F}_p$, we have $\chi(x) = x^{(p-1)/2}$. Consider the list

$$\chi(1), \chi(2), \ldots, \chi(p-1). \tag{1}$$

It is a sequence in $\{1, -1\}^{p-1}$. The following bound is a standard consequence of the celebrated Weil bound for character sums; see [16] for a detailed proof.

Proposition 3.1. *Let $(b_1, b_2, \ldots, b_t)$ be a sequence in $\{1, -1\}^t$. Then the number of $x \in \mathbb{F}_p$ such that*

$$\chi(x) = b_1, \chi(x+1) = b_2, \ldots, \chi(x+t-1) = b_t$$

lies between $p/2^t - t(3+\sqrt{p})$ and $p/2^t + t(3+\sqrt{p})$.

The proposition implies that if $t < (\log p)/3$, then every possible sequence in $\{-1, 1\}^t$ occurs as a consecutive subsequence in expression (1). In many situations it is more convenient to use binary 0/1 sequences, which suggests instead using the polynomial $(x^{(p-1)/2} + 1)/2$, but this results in a small problem at $x = 0$. We instead use the sparse polynomial

$$\alpha(x) = (x^{(p-1)/2} + x^{p-1})/2. \tag{2}$$

The polynomial $\alpha(x)$ takes values in $\{0, 1\}$ if $x \in \mathbb{F}_p$, and $\alpha(x) = 1$ if and only if $\chi(x) = 1$.

Corollary 3.2. *If $t < (\log p)/3$, then for any binary sequence $(b_1, b_2, \ldots, b_t) \in \{0, 1\}^t$ there exists an $x \in \mathbb{F}_p$ such that*

$$\alpha(x) = b_1,\ \alpha(x+1) = b_2, \ldots, \alpha(x+t-1) = b_t.$$

In other words, if $t < (\log p)/3$, the map

$$x \mapsto \big(\alpha(x), \alpha(x+1), \ldots, \alpha(x+t-1)\big)$$

is a *surjective* map from $\mathbb{F}_p$ to $\{0, 1\}^t$; one can view this map as sending an algebraic object to a combinatorial object.

Given a straight-line polynomial $f(x) \in \mathbb{Z}[x]$ and a prime p, how hard is it to decide whether the polynomial has a solution in $\mathbb{F}_p$? We now prove that this problem is NP-hard.

Theorem 3.3. *Given a sparse shift polynomial $f(x) \in \mathbb{Z}[x]$ and a large prime p, it is NP-hard to decide whether $f(x)$ has a root in $\mathbb{F}_p$ under RP-reduction.*

Proof. We reduce the (decision version of the) subset sum problem to this problem. Given $b \in \mathbb{Z}_{\geq 0}$ and $S = \{a_1, a_2, \ldots, a_t\} \subseteq \mathbb{Z}_{\geq 0}$, one can find a prime p such that $p > \max(2^{3t}, \sum_{i=1}^{t} a_i)$ and construct a sparse shift polynomial

$$\beta(x) = \sum_{i=0}^{t-1} a_i \alpha(x+i) - b. \tag{3}$$

If the polynomial has a solution modulo p, then the answer to the subset sum problem is "yes", since for every $x \in \mathbb{F}_p$ we have $\alpha(x+i) \in \{0, 1\}$.

In the other direction, if the answer to the subset sum problem is "yes", then according to Corollary 3.2, the polynomial has a solution in $\mathbb{F}_p$. Note that the reduction can be computed in randomized polynomial time. □

4. Complexity of the value set counting problem

In this section, we prove several results about the complexity of the value set counting problem.

4A. ***Finite fields of characteristic 2.*** We will use a problem about NC^0_5 circuits to prove that counting the value set of a sparse polynomial in a field of characteristic 2 is #P-hard. A Boolean circuit is in NC^0_5 if every output bit of the circuit depends only on at most 5 input bits. We can view a circuit with n input bits and m output bits as a map from $\{0,1\}^n$ to $\{0,1\}^m$ and call the image of the map the *value set* of the circuit. The following proposition is implied in [6]; we provide a sketch of the proof.

Proposition 4.1. *Given a* 3SAT *formula with n variables and m clauses, one can construct in polynomial time an NC^0_5 circuit with $n+m$ input bits and $n+m$ outputs bits, such that if there are M satisfying assignments for the* 3SAT *formula, then the cardinality of the value set of the NC^0_5 circuit is $2^{n+m} - 2^{m-1}M$. In particular, if the* 3SAT *formula can not be satisfied, then the circuit computes a permutation from $\{0,1\}^{n+m}$ to $\{0,1\}^{n+m}$.*

Proof. Denote the variables and the clauses of the 3SAT formula by $x_1, x_2, \ldots, x_n$ and $C_1, C_2, \ldots, C_m$, respectively. Build a circuit with $n + m$ input bits and $n + m$ output bits as follows. The input bits will be denoted by $x_1, x_2, \ldots, x_n$

and $y_1, y_2, \ldots, y_m$, and the output bits will be denoted by $z_1, z_2, \ldots, z_n$ and $w_1, w_2, \ldots, w_m$. Set $z_i = x_i$ for $1 \le i \le n$, and set

$$w_i = \big(C_i \wedge (y_i \oplus y_{(i+1 \,(\mathrm{mod}\, m))})\big) \vee (\neg C_i \wedge y_i)$$

for $1 \le i \le m$. In other words, if C_i is evaluated to be TRUE, then output $y_i \oplus y_{(i+1 \,(\mathrm{mod}\, m))}$ as w_i, and otherwise output y_i as w_i. Note that C_i depends only on 3 variables from $\{x_1, x_2, \ldots, x_n\}$, so we obtain an NC^0_5 circuit. After fixing an assignment to the x_i, the z_i are also fixed, and the transformation from $(y_1, y_2, \ldots, y_m)$ to $(w_1, w_2, \ldots, w_m)$ is linear over $\mathbb{F}_2$. One can verify that the linear transformation has rank $m-1$ if the assignment satisfies all the clauses, and it has rank m (that is, it has full rank) if some of the clauses are not satisfied. So the cardinality of the value set of the circuit is

$$M2^{m-1} + (2^n - M)2^m = 2^{n+m} - 2^{m-1}M. \qquad \square$$

If we replace the Boolean gates in the NC^0_5 circuit by algebraic gates over $\mathbb{F}_2$, we obtain an algebraic circuit that computes a polynomial map from $\mathbb{F}_2^{n+m}$ to itself, where each polynomial depends only on 5 variables and has degree equal to or less than 5. There is an $\mathbb{F}_2$-basis for $\mathbb{F}_{2^{n+m}}$, say $\omega_1, \omega_2, \ldots, \omega_{n+m}$, which induces a bijection from $\mathbb{F}_2^{n+m}$ to $\mathbb{F}_{2^{n+m}}$ given by

$$(x_1, x_2, \ldots, x_{n+m}) \mapsto x = \sum_{i=1}^{n+m} x_i \omega_i;$$

the inverse of this map can be represented by sparse polynomials in $\mathbb{F}_{2^{n+m}}[x]$. Using this fact, we can replace the input bits of the algebraic circuit by sparse polynomials, and collect the output bits together using the base to form a single element in $\mathbb{F}_{2^{n+m}}$. We thus obtain a sparse univariate polynomial in $\mathbb{F}_{2^{n+m}}[x]$ from the NC^0_5 circuit such that their value sets have the same cardinality. We thus have the following theorem.

Theorem 4.2. *Given a* 3SAT *formula with n variables and m clauses, one can construct in polynomial time a sparse polynomial $\gamma(x)$ over $\mathbb{F}_{2^{n+m}}$ such that the value set of $\gamma(x)$ has cardinality $2^{n+m} - 2^{m-1}M$, where M is the number of satisfying assignments of the* 3SAT *formula.*

Since counting the number of satisfying assignments for a 3SAT formula is known to be #P-complete, we have our main theorem.

Theorem 4.3. *The problem of counting the value set of a sparse polynomial over a finite field of characteristic* 2 *is #P-hard.*

Corollary 4.4. *The set of sparse permutation polynomials over finite fields of characteristic* 2 *is co-NP-complete.*

4B. ***Prime-order finite fields.*** The construction in Theorem 4.2 relies on building extensions over $\mathbb{F}_2$. The technique cannot be adopted easily to the prime-order finite field case. We will prove that counting the value set of a straight-line polynomial over a prime-order finite field is #P-hard. We reduce the counting version of the subset sum problem to the value set counting problem.

Theorem 4.5. *Given access to an oracle that solves the value set counting problem for straight-line polynomials over prime-order finite fields, there is a randomized polynomial-time algorithm solving the counting version of the* SSP.

Proof. Suppose we are given an instance of the counting subset sum problem, say b with the set $S = \{a_1, a_2, \ldots, a_n\}$. If $b > \sum_{i=1}^{n} a_i$ we answer 0, while if $b = 0$ we answer 1. Otherwise, we find a prime $p > \max(2^{3t}, 2\sum_{i=1}^{n} a_i)$ and ask the oracle to count the value set of the sparse shift polynomial

$$f(x) := \left(1 - \beta(x)^{p-1}\right)\left(\sum_{i=0}^{t-1} \alpha(x+i)2^i\right)$$

over the prime-order field $\mathbb{F}_p$, where $\alpha(x)$ and $\beta(x)$ are as defined in (2) and (3), respectively. We output the answer $\#(V_f) - 1$, which is easily seen to be exactly the number of subsets of $\{a_1, \ldots, a_n\}$ that sum to b. □

Since the counting version of the SSP is #P-complete, this theorem yields the following.

Theorem 4.6. *Over a prime-order finite field $\mathbb{F}_p$, the problem of counting the value set is #P-hard under RP-reduction, if the polynomial is given as a straight-line program.*

5. The image set and point counting

Proposition 5.1. *If $f \in \mathbb{F}_q[x]$ is a polynomial of degree $d > 0$, then the cardinality of its image set is*

$$\#(V_f) = \sum_{i=1}^{d} (-1)^{i-1} N_i \sigma_i\left(1, \frac{1}{2}, \ldots, \frac{1}{d}\right) \tag{4}$$

where $N_k = \#(\{(x_1, \ldots, x_k) \in \mathbb{F}_q^k \mid f(x_1) = \cdots = f(x_k)\})$ and σ_i denotes the i-th elementary symmetric function on d elements.

Proof. For any $y \in V_f$, define

$$\tilde{N}_{k,y} = \left\{(x_1, \ldots, x_k) \in \mathbb{F}_q^k \mid f(x_1) = \cdots = f(x_k) = y\right\}$$

and denote the cardinalities of $\tilde{N}_{k,y}$ by $N_{k,y}$. We then see that

$$N_k = \sum_{y \in V_f} N_{k,y}. \tag{5}$$

Let us refer to the right-hand side of (4) as η; plugging (5) into this expression and rearranging, we get

$$\eta = \sum_{y \in V_f} \sum_{i=1}^{d} (-1)^{i-1} N_{i,y}\, \sigma_i\Big(1, \frac{1}{2}, \ldots, \frac{1}{d}\Big).$$

Let us call the inner sum ω_y; that is,

$$\omega_y = \sum_{i=1}^{d} (-1)^{i-1} N_{i,y}\, \sigma_i\left(1, \frac{1}{2}, \ldots, \frac{1}{d}\right).$$

If we can show that for all $y \in V_f$ we have $\omega_y = 1$, then we clearly have $\eta = \#(V_f)$.

Let $y \in V_f$ be fixed. Let $k = \#(f^{-1}(y))$. It is clear that $1 \le k \le d$ and $N_{i,y} = k^i$ for $0 \le i \le d$. Substituting this in, our expression mercifully becomes somewhat nicer:

$$\begin{aligned}
\omega_y &= 1 - \sum_{i=0}^{d} (-1)^i k^i \sigma_i\Big(1, \frac{1}{2}, \ldots, \frac{1}{d}\Big) \\
&= 1 - \sum_{i=0}^{d} (-1)^i \sigma_i\Big(k, \frac{k}{2}, \ldots, \frac{k}{d}\Big) \qquad (6) \\
&= 1 - \Big[\Big(1-k\Big)\Big(1-\frac{k}{2}\Big)\cdots\Big(1-\frac{k}{d}\Big)\Big] \qquad (7) \\
&= 1.
\end{aligned}$$

From step (6) to step (7), we are using the identity

$$\prod_{j=1}^{n} (\lambda - X_j) = \sum_{j=0}^{n} (-1)^j\, \lambda^{n-j} \sigma_j\,(X_1, \ldots, X_n).$$

Note that the bracketed term of (7) is 0, as k must be an integer such that $1 \le k \le d$, so one term in the product will be 0. Thus, we have $\eta = \#(V_f)$, as desired. □

Proposition 5.1 gives us a way to express $\#(V_f)$ in terms of the numbers of rational points on a sequence of curves over $\mathbb{F}_q$. If we had a way of getting N_k for $1 \le k \le d$, then it would be easy to calculate $\#(V_f)$.

We proceed by examining a family of related spaces,

$$\tilde{N}_k = \big\{(x_1, \ldots, x_k) \in \mathbb{F}_q^k \mid f(x_1) = \cdots = f(x_k)\big\}.$$

We immediately note that $N_k = \#(\tilde{N}_k)$.

Spaces similar to our $\tilde{N}_k$ have been used several times [19; 2] to establish various asymptotic results for $\#(V_f)$. The spaces used in these earlier papers require that $x_i \neq x_j$ for $i \neq j$. We will see that our work would have been dramatically harder had we imposed these additional restrictions.

The spaces $\tilde{N}_k$ are not of any nice form (in particular, we cannot assume they are nonsingular projective, abelian varieties, and so on), so we proceed by using the p-adic point counting method described in [12], which runs in polynomial time for any variety over a field of small characteristic (that is, $p = O((d \log q)^C)$ for some positive constant C).

Theorem 5.2. *There exist an explicit deterministic algorithm and an explicit polynomial R such that for any $f \in \mathbb{F}_q[x]$ of degree d, where $q = p^m$ and p is prime, the algorithm computes the cardinality of the image set $\#(V_f)$ in a number of bit operations bounded by $R(m^d d^d p^d)$.*

Proof. We first note that

$$\tilde{N}_k = \left\{ (x_1, \ldots, x_k) \in \mathbb{F}_q^k \mid f(x_1) = \cdots = f(x_k) \right\}$$

$$= \left\{ (x_1, \ldots, x_k) \in \mathbb{F}_q^k \;\middle|\; \begin{array}{c} f(x_1) - f(x_2) = 0 \\ f(x_1) - f(x_3) = 0 \\ \vdots \\ f(x_1) - f(x_k) = 0 \end{array} \right\}.$$

For reasons soon to become clear, we need to represent this as the solution set of a single polynomial. Let us introduce additional variables z_1 to z_{k-1}, and set $x = (x_1, \ldots, x_k)$ and $z = (z_1, \ldots, z_{k-1})$. Now examine the auxiliary function

$$F_k(x, z) = z_1\big(f(x_1) - f(x_2)\big) + \cdots + z_{k-1}\big(f(x_1) - f(x_k)\big). \tag{8}$$

Clearly, if $\gamma \in \tilde{N}_k$, then $F_k(\gamma, z)$ is the zero function. If $\gamma \in \mathbb{F}_q^k \setminus \tilde{N}_k$, then the solutions of $F_k(\gamma, z) = 0$ specify a $(k-2)$-dimensional $\mathbb{F}_q$-linear subspace of $\mathbb{F}_q^{k-1}$. Thus, if we denote the cardinality of the solution set to $F_k(x, z) = 0$ as $\#(F_k)$, then we see that

$$\begin{aligned} \#(F_k) &= q^{k-1} N_k + q^{k-2}(q^k - N_k) \\ &= N_k q^{k-2}(q-1) + q^{2k-2}. \end{aligned}$$

Solving for N_k, we find that

$$N_k = \frac{\#(F_k) - q^{2k-2}}{q^{k-2}(q-1)}. \tag{9}$$

Thus we have an easy way to determine N_k, if we know the number of points on the hypersurface defined by the single polynomial equation $F_k = 0$.

The main theorem in [12] yields an algorithm for toric point counting in $\mathbb{F}_{q^\ell}$ that is polynomial time when the characteristic is small (that is, $p = O((d \log q)^C)$ for some positive constant C) that works for general varieties. In [12, §6.4], this theorem is adapted to be a generic point counting algorithm.

To apply this result to our problem, we note that F_k is a polynomial in $2k-1$ variables with total degree $d+1$, and that we only care about the case where $\ell = 1$. Thus, the running time for this algorithm is $\tilde{O}(2^{8k+1} m^{6k+4} k^{6k+2} d^{6k-3} p^{4k+2})$ bit operations. In order to calculate $\#(V_f)$ using (4), we calculate N_k for $1 \le k \le d$, scaled by an elementary symmetric polynomial. All of the necessary elementary symmetric polynomials can be evaluated using Newton's identities (see [15]) in $O(d^2 \log d)$ multiplications. Therefore, the entire calculation has a running time of $\tilde{O}(2^{8d+1} m^{6d+4} d^{12d-1} p^{4d+2})$ bit operations. For consistency with [12], we can then note that as $d > 1$, we can write $2^{8d+1} = d^{(\log_d 2)(8d+1)}$. Thus, there is a polynomial R in one variable such that the running time of this algorithm is bounded by $R(m^d d^d p^d)$ bit operations. In the dense polynomial model, the polynomial f has input size $O(d \log q)$, so this algorithm does not have polynomial running time with respect to the input length. This algorithm has running time that is exponential in the degree d of the polynomial, and polynomial in m and p. □

Note that if we had adopted the spaces constructed in prior works [19; 2], we would have then required $x_i \neq x_j$ for $i \neq j$. The standard approach to representing such inequalities is the "Rabinovich trick". To use this trick, we would have introduced an additional variable, say y, and the additional equation

$$y \prod_{i<j} (x_j - x_i) = 1.$$

This is a polynomial of degree $\binom{k}{2} + 1$, which would have led to an equation corresponding to (8) of degree at least $\binom{k}{2} + 2$ with $2k+1$ variables; this would have increased the work factor of the algorithm significantly.

6. Open problems

The algorithm we have presented relies on the result of Lauder and Wan, which is intended to calculate the number of $\mathbb{F}_q$-rational points on a general variety. We use this algorithm on a polynomial of a very special form. As such, it may be possible to get a considerably more efficient algorithm by exploiting symmetry in the resulting Newton polytope.

Though value sets of polynomials appear to be closely related to zero sets, they are not as well studied. There are many interesting open problems about value sets. The most important one is to find a counting algorithm with running time $(d \log q)^{O(1)}$, that is, a deterministic polynomial-time algorithm in the dense

model. It is not clear if this is always possible. Our result affirmatively solves this problem for fixed d if the characteristic p is reasonably small. We conjecture that the same result is true for fixed d and all characteristic p.

For the complexity side, can one prove that the counting problem for sparse polynomials in prime-order finite fields is hard? Can one prove that the counting problem for the dense input model is hard for general degree d?

Acknowledgments

We thank Dr. Tsuyoshi Ito for pointing out reference [6] to us. All three authors are partially supported by the NSF.

References

[1] Martín Avendaño, Ashraf Ibrahim, J. Maurice Rojas, and Korben Rusek, *Randomized NP-completeness for p-adic rational roots of sparse polynomials in one variable*, in Watt [21], 2010, pp. 331–338. MR 2920572

[2] B. J. Birch and H. P. F. Swinnerton-Dyer, *Note on a problem of Chowla*, Acta Arith. **5** (1959), 417–423. MR 22 #4675

[3] J. P. Buhler and P. Stevenhagen (eds.), *Algorithmic number theory*: *lattices*, *number fields*, *curves and cryptography*, Mathematical Sciences Research Institute Publications, no. 44, Cambridge University Press, 2008. MR 2009h:11003

[4] Luís Caires, Giuseppe F. Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung (eds.), *Automata*, *languages and programming*: *Proceedings of the* 32*nd International Colloquium* (*ICALP* 2005) *held in Lisbon*, *July* 11–15, 2005, Lecture Notes in Computer Science, no. 3580, Berlin, Springer, 2005. MR 2006f:68001

[5] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein, *Introduction to algorithms*, 2nd ed., MIT Press, Cambridge, MA, 2001. MR 2002e:68001

[6] B. Durand, *Inversion of* 2*D cellular automata*: *some complexity results*, Theoret. Comput. Sci. **134** (1994), no. 2, 387–401. MR 96b:68131

[7] Erich Kaltofen, *Polynomial factorization*: *a success story*, slides presented at the International Symposium on Symbolic and Algebraic Computation (ISSAC '03), Philadelphia, August 3–6, 2003. http://www4.ncsu.edu/∼kaltofen/bibliography/lectures/lectures.html#issacphiladelphia

[8] Erich Kaltofen and Pascal Koiran, *On the complexity of factoring bivariate supersparse* (*lacunary*) *polynomials*, in Kauers [9], 2005, pp. 208–215. MR 2280549

[9] Manuel Kauers (ed.), *ISSAC'*05: *Proceedings of the* 30*th International Symposium on Symbolic and Algebraic Computation held in Beijing*, *July* 24–27, 2005, New York, ACM Press, 2005, held in Beijing, July 24–27, 2005. MR 2007g:68005

[10] Neeraj Kayal, *Solvability of a system of bivariate polynomial equations over a finite field* (*extended abstract*), in Caires et al. [4], 2005, pp. 551–562. MR 2184660

[11] Aviad Kipnis and Adi Shamir, *Cryptanalysis of the HFE public key cryptosystem by relinearization*, in Wiener [22], 1999, pp. 19–30. MR 2000i:94052

[12] Alan G. B. Lauder and Daqing Wan, *Counting points on varieties over finite fields of small characteristic*, in Buhler and Stevenhagen [3], 2008, pp. 579–612. MR 2009j:14029

[13] Keju Ma and Joachim von zur Gathen, *The computational complexity of recognizing permutation functions*, Comput. Complexity **5** (1995), no. 1, 76–97. MR 96c:68066

[14] ———, *Tests for permutation functions*, Finite Fields Appl. **1** (1995), 31–56. MR 96a:11137

[15] D. G. Mead, *Newton's identities*, Amer. Math. Monthly **99** (1992), no. 8, 749–751. MR 93h:05011

[16] René Peralta, *On the distribution of quadratic residues and nonresidues modulo a prime number*, Math. Comp. **58** (1992), no. 197, 433–440. MR 93c:11115

[17] I. E. Shparlinski, *A deterministic test for permutation polynomials*, Comput. Complexity **2** (1992), no. 2, 129–132. MR 93h:11136

[18] Terence Tao, Ernest Croot, III, and Harald Helfgott, *Deterministic methods to find primes*, Math. Comp. **81** (2012), no. 278, 1233–1246. MR 2869058

[19] Saburô Uchiyama, *Note on the mean value of $V(f)$*, Proc. Japan Acad. **31** (1955), 199–201. MR 17,130f

[20] Joachim von zur Gathen, *Tests for permutation polynomials*, SIAM J. Comput. **20** (1991), no. 3, 591–602. MR 92g:11117

[21] Stephen M. Watt (ed.), *ISSAC* 2010*—Proceedings of the* 2010 *International Symposium on Symbolic and Algebraic Computation*, Association for Computing Machinery, New York, 2010. MR 2920530

[22] Michael Wiener (ed.), *Advances in cryptology—CRYPTO '*99: *Proceedings of the* 19*th Annual International Cryptology Conference held in Santa Barbara, CA, August* 15–19, 1999, Lecture Notes in Computer Science, no. 1666, Berlin, Springer, 1999. MR 2000h:94003

QI CHENG: qcheng@cs.ou.edu
School of Computer Science, The University of Oklahoma, Norman, OK 73019, United States

JOSHUA E. HILL: hillje@math.uci.edu
Department of Mathematics, University of California, Irvine, Irvine, CA 92697, United States

DAQING WAN: dwan@math.uci.edu
Department of Mathematics, University of California, Irvine, Irvine, CA 92697, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.249

msp

Haberland's formula and numerical computation of Petersson scalar products

Henri Cohen

We study several methods for the numerical computation of Petersson scalar products, and in particular we prove a generalization of Haberland's formula to any subgroup of finite index G of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, which gives a fast method to compute these scalar products when a Hecke eigenbasis is not necessarily available.

1. Introduction

Let G be a subgroup of $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ of finite index $r = [\Gamma : G]$. Recall that Γ acts on the upper half-plane $\mathcal{H}$ via linear fractional transformations and that we have an invariant measure $d\mu = dx\,dy/y^2$. We will denote by $D(G)$ a "reasonable" fundamental domain for the action of G on $\mathcal{H}$; see Definition 4.1 below.

Given two modular forms f_1 and f_2 having the same weight k and the same multiplier system v on G, we recall that one defines the *Petersson scalar product* $\langle f_1, f_2\rangle_G$ (abbreviated PSP), when it exists, by the formula

$$\langle f_1, f_2\rangle_G = \frac{1}{[\Gamma:G]}\int_{G\backslash\mathcal{H}} f_1(\tau)\overline{f_2(\tau)}y^k\,\frac{dx\,dy}{y^2} = \frac{1}{r}\int_{D(G)} f_1(\tau)\overline{f_2(\tau)}y^k\,d\mu.$$

This is a fundamental quantity which enters almost everywhere in the theory of modular forms, and the aim of the present paper is to study how to compute it numerically in practice. The normalizing factor $1/r$ is included so that the result does not depend on which group is taken with respect to which both f_1 and f_2 are modular.

The absolute convergence of the above integral is assured if either f_1 or f_2 is a cusp form, or if we are in weight $1/2$. Note however that it can also converge in other cases. We will always consider the case where one of f_1 and f_2 is a cusp

MSC2010: primary 11F11; secondary 11Y35.
Keywords: Petersson product.

form and we will assume that $k \geq 2$ and that k is integral. It is an interesting and nontrivial question to ask what can be done when $k = 1$.

When the space $S_k(G, v)$ of cusp forms of weight k and multiplier system v is known explicitly, and in particular when the decomposition into Hecke eigenforms is known (when $G = \Gamma_0(N)$ or $\Gamma_1(N)$ for instance), there are specific methods for computing the PSP if the decomposition of f_1 and f_2 on the eigenbasis can be easily computed; we will mention these methods below. But we are more interested in the general context where one does not need to know either $S_k(G, v)$ or the eigenbasis decompositions, but where we assume that for any $\tau \in \mathcal{H}$ one can rapidly compute $f_1(\tau)$ and $f_2(\tau)$ to reasonably high accuracy.

In the sequel we will let $(\gamma_j)_{1 \leq j \leq r}$ be a system of representatives of right cosets of $G\backslash\Gamma$, so that $\Gamma = \bigsqcup_{1 \leq j \leq r} G\gamma_j$. In particular, if $\mathfrak{F}$ is a fundamental domain for the full modular group Γ (for instance the standard one), then $\bigcup_{1 \leq j \leq r} \gamma_j(\mathfrak{F})$ is a fundamental domain for G, where the union is essentially disjoint, with the only possible intersections being on the boundaries.

Recall that if $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ we write $f|_k\gamma$ to mean

$$f|_k\gamma(\tau) = (c\tau + d)^{-k} f\left(\frac{a\tau + b}{c\tau + d}\right),$$

so that f is an element of $M_k(G, v)$ if and only if $f|_k\gamma = v(\gamma)f$ for all $\gamma \in G$ and f is holomorphic on $\mathcal{H}$ and at the cusps; also, f lies in $S_k(G, v)$ if in addition f vanishes at the cusps.

It is clear that $f|_k g\gamma_j = v(g) f|_k \gamma_j$, so up to the factor $v(g)$ the function $f_j = f|_k\gamma_j$ is independent of the chosen representative of the right coset $G\gamma_j$. In addition, for any $\alpha \in \Gamma$ we have by definition $\gamma_j\alpha = g_j\gamma_{a(j)}$ for some $g_j \in G$, the map $j \mapsto a(j)$ being a *permutation* of $[1, r]$, so up to the factors $v(g_j)$, the family of $f_j|_k\alpha$ is simply a permutation of the f_j.

2. Some standard methods

Before coming to the more original part of the paper, where we explain how to compute PSP's in a quite general setting, we recall with some detail some well-known methods.

Throughout the paper we will use three test examples, even though they are not completely general:

$$f_1 = f_2 = \Delta(\tau) = \eta(\tau)^{24} \in S_{12}(\Gamma),$$

$$f_1 = f_2 = \Delta_5(\tau) = \big(\eta(\tau)\eta(5\tau)\big)^4 \in S_4(\Gamma_0(5)),$$

and

$$f_1 = f_2 = \Delta_{11}(\tau) = \big(\eta(\tau)\eta(11\tau)\big)^2 \in S_2(\Gamma_0(11)),$$

the last of these being the cusp form associated to the elliptic curve $X_0(11)$. To 47 decimals, we have

$$\begin{aligned}\langle \Delta, \Delta\rangle_\Gamma &= 0.00000103536205680432092234781681222516459322491\ldots,\\ \langle \Delta_5, \Delta_5\rangle_{\Gamma_0(5)} &= 0.00014513335082978187614092680220909259631066600\ldots,\\ \langle \Delta_{11}, \Delta_{11}\rangle_{\Gamma_0(11)} &= 0.00390834565612459898524738548138211386179054941\ldots.\end{aligned}$$

In most cases, we assume for simplicity that $G = \Gamma$, but we will of course state the necessary modifications for a general subgroup of finite index G.

2A. ***Computing from the definition.*** A first method for computing PSP's is to use the definition directly: Assuming for instance that $G = \Gamma$, we have

$$\begin{aligned}\langle f_1, f_2\rangle &= \int_{\mathfrak{F}} f_1(\tau)\overline{f_2(\tau)}y^{k-2}\,dx\,dy\\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}}\left(\int_{\sqrt{1-x^2}}^{\infty} f_1(x+iy)\overline{f_2(x+iy)}y^{k-2}\,dy\right)dx.\end{aligned}$$

Since the functions f_i are holomorphic, to compute the integrals numerically one can use the *doubly exponential integration method* (see for instance [2, §9.3]). This little-known but remarkable method is especially efficient for holomorphic functions, and it can be shown that to obtain an accuracy of N decimals the method requires $O(N \log N)$ evaluations of the function to be integrated.

However, we have here a double integral, so the method requires $O(N^2 \log^2 N)$ evaluations of the functions, which can be rather expensive. Of course this can be generalized to any subgroup G by using a natural choice of fundamental domain $D(G) = \bigcup_{1\le j\le r}\gamma_j(\mathfrak{F})$ and making the obvious changes of variable. Table 1 gives a selection of timings to compute $\langle f, f\rangle_G$ to a given number N of decimals using this method. The timings are in seconds, and those not given (as indicated by a dash) are greater than 30 minutes. The present timings have been made on a single processor of a standard 1.8 GHz Intel core i7 CPU, but they are highly dependent on the implementation, so this table is only indicative.

f	$N = 19$	38	57	96	250	500
Δ	11	16	87	143	—	—
Δ_5	154	219	1185	—	—	—
Δ_{11}	327	468	—	—	—	—

Table 1. Timings (in seconds, on one processor of a 1.8 GHz Intel core i7 CPU) to compute $\langle f, f\rangle_G$ to N decimal places using the definition of the pairing. Timings greater than 30 minutes are indicated with a dash.

To summarize: The advantages of this method are its complete generality and simplicity, while its main disadvantage is that it is quite slow, especially at high accuracy and/or for a subgroup of large index.

2B. ***Using Kloosterman sums.*** Thanks to the computation of the Fourier expansion of Poincaré series for Γ, it is easy to show that

$$\frac{1}{\langle \Delta, \Delta \rangle} = \frac{(4\pi)^{11}}{10!\,\tau(n)} \left(\delta_{n,1} + 2\pi \cdot n^{11/2} \sum_{c \geq 1} \frac{K(n,1;c)}{c} J_{11}\left(\frac{4\pi n^{1/2}}{c} \right) \right),$$

and similar formulas exist in higher weight and for congruence subgroups.

The convergence of this type of series is essentially of the order of $O(1/c^{k-2})$ (here with $k = 12$). This shows that, although useful, the above formula has severe limitations. First, even in the case of Δ, the convergence in $O(1/c^{10})$ and the necessity of computing Kloosterman sums and Bessel functions implies that one can reasonably compute perhaps 10^6 terms if one is patient, giving an accuracy of 60 decimals. A more important limitation occurs for subgroups of Γ, for which there exist forms of lower weight than 12. For instance, in weight 2 the absolute convergence is not even clear, and in weight 4 the convergence is in $O(1/c^2)$, which is too slow to obtain any reasonable accuracy.

Table 2 presents some timings for this method, but limited to Δ since the convergence for Δ_5 would be too slow.

To summarize: The advantage of this method is its speed for high weight and reasonably low accuracy such as 19 or 38 decimals, but the method is essentially useless in all other cases. In addition, its use is restricted to congruence subgroups.

2C. ***Using symmetric square L-functions.*** Once again for simplicity we restrict to $G = \Gamma$, but there is no difficulty in generalizing.

Since there exists an explicit orthogonal basis of eigenfunctions in $M_k(\Gamma)$, computing Petersson scalar products of two arbitrary forms can easily be reduced to the computation of $\langle f, f \rangle$ for f a normalized eigenform. If

$$L(f,s) = \sum_{n \geq 1} \frac{a(n)}{n^s} = \prod_p \frac{1}{1 - a(p)p^{-s} + p^{k-1-2s}} = \prod_p \frac{1}{(1-\alpha_p p^{-s})(1-\beta_p p^{-s})}$$

f	$N = 19$	38	57	96	250	500
Δ	0.01	3	900	—	—	—

Table 2. Timings (in seconds) to compute $\langle f, f \rangle_G$ to N decimal places using Kloosterman sums.

with $\alpha_p + \beta_p = a(p)$ and $\alpha_p\beta_p = p^{k-1}$, recall that we define the *symmetric square L-function* $L(\mathrm{Sym}^2(f), s)$ for $\Re(s) > k$ by the formula

$$L(\mathrm{Sym}^2(f), s) = \prod_p \frac{1}{(1-\alpha_p^2 p^{-s})(1-\alpha_p\beta_p p^{-s})(1-\beta_p^2 p^{-s})}.$$

The main properties of this function are summarized in the following result.

Theorem 2.1. *Let* $f = \sum_{n\geq 1} a(n)q^n \in S_k(\Gamma)$ *be a normalized Hecke eigenform.*

(1) (*Fourier expansion.*) *If we set*

$$A(n) = \sum_{m|n} (-1)^{\Omega(m)} m^{k-1} a(n/m)^2,$$

where $\Omega(m)$ *is the number of prime divisors of* m, *counted with multiplicity*, *then*

$$L(\mathrm{Sym}^2(f), s) = \sum_{n\geq 1} \frac{A(n)}{n^s}.$$

(2) (*Functional equation.*) *The function* $L(\mathrm{Sym}^2(f), s)$ *can be extended* holomorphically *to the whole of* $\mathbb{C}$, *and the completed L-function*

$$\Lambda\big(\mathrm{Sym}^2(f), s\big) = \pi^{-3s/2}\Gamma(s/2)\Gamma\big((s+1)/2\big)\Gamma\big((s-k)/2+1\big)L\big(\mathrm{Sym}^2(f), s\big)$$

satisfies the functional equation

$$\Lambda\big(\mathrm{Sym}^2(f), 2k-1-s\big) = \Lambda\big(\mathrm{Sym}^2(f), s\big).$$

(3) (*Special value.*) *We have*

$$L\big(\mathrm{Sym}^2(f), k\big) = \frac{\pi}{2}\frac{(4\pi)^k}{(k-1)!}\langle f, f\rangle.$$

Proof. The meromorphic continuation, functional equation, and special value are very classical and immediate consequences of the Rankin-Selberg method. The holomorphy is more difficult, and was proved independently by Shimura and Zagier in 1975. □

Note that similar results are of course valid for subgroups.

The last statement of the theorem allows us to reduce the computation of $\langle f, f\rangle$ to that of $L(\mathrm{Sym}^2(f), k)$. For this, the direct use of the definition is of little help, since it is not even clear that the series or product defining this L-function converge, and even if they do, the convergence will be extremely slow. However, the crucial point is the following: Any Dirichlet series satisfying a functional equation of standard type can be evaluated numerically very efficiently using *exponentially convergent* series, see for instance [1, §10.3]. Specializing to our case, it is easy to show the following theorem.

Theorem 2.2. *Let $f=\sum_{n\ge1}a(n)q^n\in S_k(\Gamma)$ be a Hecke eigenform. Set $C=2\cdot\pi^{\frac{3}{2}}$ and $\gamma(s)=C^{-s}\Gamma(s)\Gamma\big((s-k)/2+1\big)$, and as usual let H_n denote the harmonic number $\sum_{1\le j\le n}1/j$ and let γ denote Euler's constant. Let*

$$F_{1,k}(s,x)=\sum_{1\le m\le(k-2)/2}(-1)^{k/2-m-1}\frac{(2m-1)!}{(k/2-m-1)!}\frac{(Cx)^{-2m}}{s-2m},$$

$$F_{2,k}(s,x)=\sum_{m\ge0}(-1)^{k/2-m-1}\frac{2^{2m+k}(m+k/2)!}{(2m+1)!(2m+k)!}\frac{(Cx)^{2m+1}}{s+2m+1},\quad\text{and}$$

$$F_{3,k}(s,x)=\sum_{m\ge0}(-1)^{k/2-m-1}\frac{1}{(2m)!(m+k/2-1)!}\frac{(Cx)^{2m}}{2m+s}\,G_m(s,x),$$

where

$$G_m(s,x)=2H_{2m}+H_{m+k/2-1}-3\gamma-2\log(Cx)+\frac{2}{2m+s},$$

and set

$$F_k(s,x)=\gamma(s)-x^s\big(2F_{1,k}(s,x)+\pi^{1/2}F_{2,k}(s,x)+F_{3,k}(s,x)\big).$$

Then for every $s\in\mathbb{C}$ with $\Re(s)>k-2$ and every $t_0>0$, we have

$$\gamma(s)L\big(\mathrm{Sym}^2(f),s\big)=\sum_{n\ge1}\frac{A(n)}{n^s}F_k(s,nt_0)+\sum_{n\ge1}\frac{A(n)}{n^{2k-1-s}}F_k(2k-1-s,n/t_0)$$

where the $A(n)$ are the coefficients given in part (1) *of Theorem 2.1. In particular*,

$$\langle f,f\rangle=2^{1-k}\pi^{k/2-1}\left(\sum_{n\ge1}\frac{A(n)}{n^k}\big(F_k(k,n)+nF_k(k-1,n)\big)\right).$$

Note that even though there is cancellation for large x, the series for $F_k(s,x)$ are sufficient for practical computation. One can also compute asymptotic expansions for large x, if desired, showing in particular that $F_k(s,x)$ tends to 0 exponentially.

Table 3 presents a few timings; for simplicity of implementation, we again limit the table to the case $f=\Delta$.

The advantages of this method are that it is general and fast; its main disadvantage is that its implementation requires great care in writing the correct formulas,

f	$N=19$	38	57	96	250	500
Δ	0.03	0.09	0.2	0.8	11	97

Table 3. Timings (in seconds) to compute $\langle f,f\rangle_G$ to N decimal places using symmetric-square L-functions.

especially for subgroups, and in dealing with cancellation and accuracy problems. But once these hurdles have been overcome, it is the best method that we have seen up to now, and most experts in the field would agree that it is the best available. However, as already mentioned, it assumes that the eigenfunction decomposition of f is known, and this is not always easy or possible. This lead us now to a different method, which is completely general.

3. Basic lemmas

The main computational difficulty related to Petersson products is that they are truly *double* integrals. In the first naïve approach, we have explained that nonetheless these integrals can be computed, somewhat slowly, by using doubly exponential integration techniques. A remarkable fact however, discovered by Haberland [4] (see also [7]) some time ago, is that PSP's can be reduced to the computation of a reasonably small finite number of *simple* integrals, which can now be evaluated very rapidly using doubly exponential integration.

Haberland's result was given for general weights k but only for the full modular group. In a slightly different form it was generalized long ago to $\Gamma_0(N)$ but only in weight $k = 2$ and trivial character, first by Cremona [3] and Zagier [10] in the context of computing the degree of modular parametrizations of elliptic curves (see the more recent paper of Watkins [9] on this subject), and much more recently by Merel [5] in connection with Manin symbols. It was realized that a complete generalization should not be difficult to obtain, and it is one of the purposes of this paper to give it. Note that in [6] the authors also give such a generalization, in a slightly different form, and also for noncuspforms. In what follows, we will assume that f_1 and f_2 are both cuspforms; if one of the f_i is not a cuspform we can either find its decomposition into its Eisenstein and cuspidal part, which can usually be done with ease, or use the generalization due to [6].

Our goal in this section, which is the main step toward Haberland's formulas, is to show that PSP's are related to other double integrals, which are not "true" double integrals in the sense that they can easily be expressed in terms of simple integrals. For this, we need some preliminary definitions and results. We assume G, $(\gamma_j)_{1\le j\le r}$, k, v, f_1, and f_2 as above, and we will set $f_{1,j} = f_1|_k\gamma_j$ and $f_{2,j} = f_2|_k\gamma_j$ for $1 \le j \le r$. As mentioned above, for simplicity we assume that f_1 and f_2 are *both* cuspforms.

3A. *The differentials ε and δ.*

Definition 3.1. We set

$$\varepsilon(f_1, f_2)(\tau_1, \tau_2) = f_1(\tau_1)\overline{f_2(\tau_2)}(\tau_1 - \overline{\tau_2})^{k-2}\, d\tau_1\, d\overline{\tau_2}$$

and

$$\delta(f_1, f_2) = \sum_{1 \le j \le r} \varepsilon(f_{1,j}, f_{2,j}).$$

Lemma 3.2. *Let* $\alpha \in \Gamma$.

(1) *We have*

$$\varepsilon(f_1, f_2)(\alpha\tau_1, \alpha\tau_2) = \varepsilon\big(f_1|_k\alpha, f_2|_k\alpha\big)(\tau_1, \tau_2).$$

(2) *The expression* $\varepsilon(f_{1,j}, f_{2,j})$ *does not depend on the choice of the right coset representative* γ_j.

(3) *If* $\gamma_j\alpha = g_j\gamma_{a(j)}$ *with* $g_j \in G$ *we have*

$$\varepsilon(f_{1,j}, f_{2,j})(\alpha\tau_1, \alpha\tau_2) = \varepsilon(f_{1,a(j)}, f_{2,a(j)}).$$

(4) *We have* $\delta(f_1, f_2)(\alpha\tau_1, \alpha\tau_2) = \delta(f_1, f_2)$; *in other words,* $\delta(f_1, f_2)$ *is invariant under* Γ.

Proof. Writing $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we have

$$\begin{aligned}
&\varepsilon(f_1, f_2)(\alpha\tau_1, \alpha\tau_2)\\
&\quad = f_1|_k\alpha(\tau_1)\ \overline{f_2|_k\alpha}(\tau_2)\cdot(c\tau_1+d)^k\overline{(c\tau_2+d)}^k(\alpha\tau_1-\overline{\alpha\tau_2})^{k-2}\,d\alpha\tau_1\,d\overline{\alpha\tau_2}\\
&\quad = f_1|_k\alpha(\tau_1)\ \overline{f_2|_k\alpha}(\tau_2)(\tau_1-\overline{\tau_2})^{k-2}\,d\tau_1\,d\overline{\tau_2}\\
&\quad = \varepsilon\left(f_1|_k\alpha, f_2|_k\alpha\right)(\tau_1, \tau_2),
\end{aligned}$$

using the immediate but fundamental identity

$$(c\tau_1+d)^k\overline{(c\tau_2+d)}^k(\alpha\tau_1-\overline{\alpha\tau_2})^{k-2}\,d\alpha\tau_1\,d\overline{\alpha\tau_2} = (\tau_1-\overline{\tau_2})^{k-2}\,d\tau_1\,d\overline{\tau_2}.$$

Statement (1) follows.

If $g \in G$ we have $f_1|_k g\gamma_j = v(g)f_{1,j}$, and similarly for f_2, so statement (2) follows from $v(g)\overline{v(g)} = 1$.

By definition we have

$$f_{1,j}|_k\alpha = f_1|_k\gamma_j\alpha = f_1|_k g_j\gamma_{a(j)} = v(g_j)f_{1,a(j)}$$

since $f_1 \in M_k(G, v)$, and similarly for $f_{2,j}$. Again using $v(g_j)\overline{v(g_j)} = 1$, we obtain statement (3). Statement (4) follows by summing on j since the map $j \mapsto a(j)$ is a permutation. □

3B. ***The simple integral $F_{2,j}$.***

Definition 3.3. Let $Z \in \overline{\mathcal{H}}$ be fixed, and set

$$F_{2,j}(Z;\tau) = F_{2,j}(\tau) = \int_Z^\tau \overline{f_{2,j}(\tau_2)}(\tau - \overline{\tau_2})^{k-2}\, d\overline{\tau_2}.$$

Remarks. (1) We could also define $F_{1,j}$ in a similar manner, but we will only need $F_{2,j}$ since we temporarily treat f_1 and f_2 in a nonsymmetric manner.

(2) Note that $F_{2,j}$ is in general not holomorphic, so must be considered as a function of τ and $\overline{\tau}$.

(3) We have

$$F_{2,j}(Z_1;\tau) - F_{2,j}(Z_2;\tau) = \int_{Z_1}^{Z_2} \overline{f_{2,j}(\tau_2)}(\tau - \overline{\tau_2})^{k-2}\, d\overline{\tau_2},$$

which is a *polynomial* (hence in particular a holomorphic function) in τ.

Lemma 3.4. (1) *We have*

$$\frac{\partial F_{2,j}}{\partial \overline{\tau}} = \overline{f_{2,j}(\tau)}(\tau - \overline{\tau})^{k-2}.$$

(2) *For every $\alpha \in \Gamma$ we have*

$$F_{2,j}\,|_{2-k}\,\alpha(\tau) = \int_{\alpha^{-1}(Z)}^{\tau} \overline{f_{2,j}\,|_k\,\alpha(\tau_2)}(\tau - \overline{\tau_2})^{k-2}\, d\overline{\tau_2}.$$

(3) *In particular, if we write $\gamma_j \alpha = g_j \gamma_{a(j)}$ with $g_j \in G$, we have*

$$F_{2,j}\,|_{2-k}\,\alpha(\tau) = \overline{v(g_j)}\big(F_{2,a(j)}(\tau) - P_{a(j)}(\alpha;\tau)\big),$$

where

$$P_{a(j)}(\alpha;\tau) = \int_Z^{\alpha^{-1}(Z)} \overline{f_{2,a(j)}(\tau_2)}(\tau - \overline{\tau_2})^{k-2}\, d\overline{\tau_2}$$

is a polynomial *in τ of degree less than or equal to $k-2$ (recall once again that we assume $k \geq 2$).*

(4) *We have*

$$\left(\int_A^B - \int_{\alpha(A)}^{\alpha(B)}\right) \sum_{1 \leq j \leq r} f_{1,j}(\tau) F_{2,j}(\tau)\, d\tau = \int_A^B \sum_{1 \leq j \leq r} f_{1,j}(\tau) P_j(\alpha;\tau)\, d\tau.$$

Proof. We have $\overline{F_{2,j}(\tau)} = \int_Z^\tau f_{2,j}(\tau_2)(\overline{\tau} - \tau_2)^{k-2}\, d\tau_2$, so

$$\frac{\partial \overline{F_{2,j}(\tau)}}{\partial \tau} = f_{2,j}(\tau)(\overline{\tau} - \tau)^{k-2}.$$

Conjugating this equality proves statement (1).

Setting $\tau_2 = \alpha z$ and writing $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we have

$$\begin{aligned} F_{2,j}\,|_{2-k}\,\alpha(\tau) &= (c\tau+d)^{k-2}\int_Z^{\alpha\tau} \overline{f_{2,j}(\tau_2)}(\alpha\tau-\overline{\tau_2})^{k-2}\,d\overline{\tau_2} \\ &= (c\tau+d)^{k-2}\int_{\alpha^{-1}Z}^{\tau} (c\overline{z}+d)^{k-2}\overline{f_{2,j}\,|_k\,\alpha(z)}(\alpha\tau-\alpha\overline{z})^{k-2}\,d\overline{z} \\ &= \int_{\alpha^{-1}Z}^{\tau} \overline{f_{2,j}\,|_k\,\alpha(z)}(\tau-\overline{z})^{k-2}\,d\overline{z}, \end{aligned}$$

since $\alpha u-\alpha v = (u-v)\big/\big((cu+d)(cv+d)\big)$; this proves statement (2).

Since we have $f_{2,j}\,|_k\,\alpha = v(g_j)f_{2,a(j)}$, it follows from statement (2) that

$$F_{2,j}\,|_{2-k}\,\alpha(\tau) = \overline{v(g_j)}\int_{\alpha^{-1}(Z)}^{\tau} \overline{f_{2,a(j)}(\tau_2)}(\tau-\overline{\tau_2})^{k-2}\,d\overline{\tau_2},$$

proving statement (3).

Setting $\tau = \alpha z$ with $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and as before $\gamma_j\alpha = g_j\gamma_{a(j)}$, we have

$$\begin{aligned} \int_{\alpha(A)}^{\alpha(B)} f_{1,j}(\tau)F_{2,j}(\tau)\,d\tau &= \int_A^B f_{1,j}(\alpha z)F_{2,j}(\alpha z)(cz+d)^{-2}\,dz \\ &= \int_A^B f_{1,j}\,|_k\,\alpha(z)F_{2,j}\,|_{2-k}\,\alpha(z)\,dz \\ &= v(g_j)\overline{v(g_j)}\int_A^B f_{1,a(j)}(\tau)\big(F_{2,a(j)}(\tau)-P_{a(j)}(\alpha;\tau)\big)\,d\tau, \end{aligned}$$

and since $j \mapsto a(j)$ is a bijection, we obtain

$$\int_{\alpha(A)}^{\alpha(B)} \sum_{1\le j\le r} f_{1,j}(\tau)F_{2,j}(\tau)\,d\tau = \int_A^B \sum_{1\le j\le r} f_{1,j}(\tau)\big(F_{2,j}(\tau)-P_j(\alpha;\tau)\big)\,d\tau,$$

proving statement (4). □

Corollary 3.5. *Let f_1 and f_2 be in $M_k(G,v)$, one of them being a cusp form. For every subgroup H of Γ of finite index $s = [\Gamma : H]$ we have*

$$(2i)^{k-1}rs\langle f_1, f_2\rangle_G = \int_{\partial(D(H))} \sum_{1\le j\le r} f_{1,j}(\tau)F_{2,j}(\tau)\,d\tau,$$

where $\partial(D(H))$ denotes the boundary of a reasonable fundamental domain $D(H)$ of H.

Note that the subgroup H need not have anything to do with the subgroup G.

Proof. By definition we have

$$\begin{aligned}(2i)^{k-1} r\langle f_1, f_2\rangle_G &= \int_{D(G)} f_1(\tau)\overline{f_2(\tau)}(\tau-\overline{\tau})^{k-2}\,d\tau\,d\overline{\tau}\\ &= \sum_{1\le j\le r}\int_{\gamma_j(D(\Gamma))} f_1(\tau)\overline{f_2(\tau)}(\tau-\overline{\tau})^{k-2}\,d\tau\,d\overline{\tau}\\ &= \int_{D(\Gamma)}\sum_{1\le j\le r} f_{1,j}(\tau)\overline{f_{2,j}(\tau)}(\tau-\overline{\tau})^{k-2}\,d\tau\,d\overline{\tau}\\ &= \int_{D(\Gamma)}\delta(f_1,f_2)(\tau,\tau)\\ &= \frac{1}{s}\int_{D(H)}\delta(f_1,f_2)(\tau,\tau),\end{aligned}$$

after an evident change of variable, and since δ is invariant by Γ by Lemma 3.2. Now since $f_{1,j}$ is holomorphic, we have $\partial f_{1,j}/\partial\overline{\tau} = 0$, so by Stokes's theorem and the above lemma we have

$$\begin{aligned}(2i)^{k-1} rs\langle f_1, f_2\rangle_G &= \int_{D(H)}\sum_{1\le j\le r}\frac{\partial(f_{1,j}F_{2,j})}{\partial\overline{\tau}}\,d\tau\,d\overline{\tau}\\ &= \int_{\partial(D(H))}\sum_{1\le j\le r} f_{1,j}(\tau)F_{2,j}(\tau)\,d\tau,\end{aligned}$$

as claimed. □

3C. ***The basic double integral*** $\mathcal{J}$**.** We make the following definition.

Definition 3.6. Let f_1 and f_2 be modular forms. If A_1, B_1, A_2, B_2 are in $\overline{\mathcal{H}}$, we set, when defined,

$$\begin{aligned}\mathcal{J}(A_1,B_1;A_2,B_2) &= \int_{A_1}^{B_1}\int_{A_2}^{B_2}\delta(f_1,f_2)\\ &= \sum_{1\le j\le r}\int_{A_1}^{B_1}\int_{A_2}^{B_2} f_{1,j}(\tau_1)\overline{f_{2,j}(\tau_2)}(\tau_1-\overline{\tau_2})^{k-2}\,d\tau_1\,d\overline{\tau_2},\end{aligned}$$

where $f_{1,j} = f_1|_k\gamma_j$ and $f_{2,j} = f_2|_k\gamma_j$.

When we need to emphasize the dependence in f_1 and f_2 we will of course write $\mathcal{J}(f_1, f_2; A_1, B_1; A_2, B_2)$ instead of $\mathcal{J}(A_1, B_1; A_2, B_2)$. Also, as usual when integrating on $\mathcal{H}$ it is understood that integrals having a cusp as an endpoint must end with a hyperbolic circle. The following properties are immediate.

Lemma 3.7. (1) *The above definition does not depend on the paths of integration, as long as the conditions at the cusps are satisfied.*

(2) *The above definition does not depend on the right coset representatives* γ_j.

(3) *The function* $\mathcal{J}$ *is transitive separately on* (A_1, B_1) *and on* (A_2, B_2); *in other words*,

$$\mathcal{J}(A_1, C_1; A_2, B_2) + \mathcal{J}(C_1, B_1; A_2, B_2) = \mathcal{J}(A_1, B_1; A_2, B_2),$$

and similarly for (A_2, B_2).

(4) *We have*

$$\mathcal{J}(f_1, f_2; A_2, B_2; A_1, B_1) = (-1)^{k-2}\overline{\mathcal{J}(f_2, f_1; A_1, B_1; A_2, B_2)}.$$

(5) *We have*

$$\mathcal{J}(A_1, B_1; A_2, B_2) = \sum_{1\le j\le r}\sum_{0\le n\le k-2}(-1)^n\binom{k-2}{n}\int_{A_1}^{B_1}\tau^{k-2-n}f_{1,j}(\tau)\,d\tau\overline{\int_{A_2}^{B_2}\tau^n f_{2,j}(\tau)\,d\tau},$$

where we must assume that f_1 *and* f_2 *are both cusp forms if at least one of the* A_i *or* B_i *is a cusp.*

In particular, this last statement shows that $\mathcal{J}$ is much easier to compute than a PSP, and it is in this sense that we said above that it is not a "true" double integral.

Proposition 3.8. *For any* $\alpha \in \Gamma$ *we have*

$$\mathcal{J}(\alpha A_1, \alpha B_1; \alpha A_2, \alpha B_2) = \mathcal{J}(A_1, B_1; A_2, B_2).$$

Proof. This follows immediately from the Γ-invariance of δ, proved in Lemma 3.2. □

4. The main result

4A. ***Fundamental domains.*** Before stating and proving the main result, we must discuss fundamental domains of subgroups of Γ. We first set the following definition.

Definition 4.1. Let $G \subset \Gamma$ be a subgroup of finite index r. A subset $D(G)$ of $\mathcal{H}$ is called a *reasonable fundamental domain* (or simply a *fundamental domain*) for G if the following conditions are satisfied:

(1) $D(G)$ is a finite union of connected and simply connected open subsets of $\mathcal{H}$.

(2) The boundary $\partial(D(G)) = \overline{D(G)} \setminus D(G)$ has measure 0.

(3) For any $\tau \in \mathcal{H}$ there exists $g \in G$ such that $g\tau \in \overline{D(G)}$. In addition, if $g\tau \in D(G)$ then g is unique, or equivalently, if g_1 and $g_2 \in G$ are such that $g_1(\tau)$ and $g_2(\tau)$ are in $D(G)$, then $g_i(\tau) \in \partial(D(G))$.

If $\mathfrak{F}$ is the standard fundamental domain for the full modular group Γ, it is clear that $D(G) = \bigcup \gamma_j(\mathfrak{F}^\circ)$ is a reasonable fundamental domain. The following results are well-known.

Proposition 4.2. *The fundamental domain $D(G)$ can be chosen so that its boundary $\partial(D(G))$ is the union of an even number of oriented hyperbolic circles, say $[A_i, A_{i+1}[$ with $1 \le i \le 2n$ (where the indices are taken modulo $2n$), such that there exists a family $(\alpha_i)_{1\le i\le 2n}$ of elements of Γ and a permutation τ of $[1, 2n]$ satisfying the following properties*:

(1) *τ is an involution without fixed points (that is, $\tau^2 = 1$ and $\tau(i) \ne i$ for all i); equivalently, τ is a product of n disjoint transpositions $(i_m, j_m)_{1\le m\le n}$.*

(2) $\alpha_{\tau(i)} = \alpha_i^{-1}$.

(3) *$\alpha_i(A_i) = A_{\tau(i)+1}$ and $\alpha_i(A_{i+1}) = A_{\tau(i)}$, so that α_i gives a bijection from $[A_i, A_{i+1}[$ to $[A_{\tau(i)+1}, A_{\tau(i)}[$.*

Corollary 4.3. *If τ is the product of the n disjoint transpositions $(i_m, j_m)_{1\le m\le n}$, then α_{i_m} gives a bijection from $[A_{i_m}, A_{i_m+1}[$ to the reverse of $[A_{j_m}, A_{j_m+1}[$, and*

$$\partial(D(H)) = \bigsqcup_{1\le m\le n} \big([A_{i_m}, A_{i_m+1}[\sqcup [A_{j_m}, A_{j_m+1}[\big).$$

Proof. Clear. □

4B. ***Examples of fundamental domains.*** For simplicity, we will choose subgroups G having a fundamental domain whose boundary has only 4 sides, and τ will always be the product $(1,2)(3,4)$ of the two transpositions exchanging 1 and 2, and 3 and 4, so $i_1 = 1$ and $i_2 = 3$. The fundamental domain is thus a hyperbolic quadrilateral given by its vertices A_1, A_2, A_3, and A_4, and α_1 sends $[A_1, A_2[$ bijectively to the reverse of $[A_2, A_3[$, and α_3 sends $[A_3, A_4[$ bijectively to the reverse of $[A_4, A_1[$.

We consider a number of different subgroups H of Γ, and give one or more fundamental domains of the above type for each, where as usual $\rho = e^{2i\pi/3}$:

(1) $H = \Gamma$, with $A_1 = \rho + 1$, $A_2 = i\infty$, $A_3 = \rho$, $A_4 = i$, $\alpha_1 = T^{-1}$, and $\alpha_3 = S$, which corresponds to the standard fundamental domain $\mathfrak{F}$, where as usual $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

(2) $H = \Gamma$, with $A_1 = 0$, $A_2 = i$, $A_3 = i\infty$, $A_4 = \rho$, $\alpha_1 = S$, and $\alpha_3 = ST$.

(3) $H = \Gamma_2$ the unique subgroup of index 2 in Γ, with $A_1 = \rho + 1$, $A_2 = i\infty$, $A_3 = \rho$, $A_4 = 0$, $\alpha_1 = T^{-1}$, and $\alpha_3 = TST = ST^{-1}S = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$.

(4) $H = \Gamma_2$ the unique subgroup of index 2 in Γ, with $A_1 = 0$, $A_2 = i\infty$, $A_3 = -1$, $A_4 = \rho$, $\alpha_1 = T^{-1}$ and $\alpha_3 = T^{-1}S = \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$.

(5) $H = \Gamma_3$ one of the subgroups of index 3 in Γ, with $A_1 = 1$, $A_2 = i\infty$, $A_3 = -1$, $A_4 = I$, $\alpha_1 = T^{-2}$, and $\alpha_3 = S$.

(6) $H = \Gamma_0(3)$, which has index 4 in Γ, with $A_1 = (\rho+2)/3$, $A_2 = i\infty$, $A_3 = (\rho-1)/3$, $A_4 = 0$, $\alpha_1 = T^{-1}$, and $\alpha_3 = ST^{-3}S = \left(\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right)$.

(7) $H = \Gamma(2)$ the principal congruence subgroup of level 2, which has index 6 in Γ and is a free group, with $A_1 = 1$, $A_2 = i\infty$, $A_3 = -1$, $A_4 = 0$, $\alpha_1 = T^{-2}$, and $\alpha_3 = ST^{-2}S = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$.

Proof. The domain (1) is of course completely classical, and the others, which can all be found somewhere in the literature, can be usually deduced by splitting the standard fundamental domain of (1) into a finite number of pieces and then applying to those a suitable finite number of elements of Γ. One can also prove the results directly in the same way as the classical proofs of (1). □

4C. *The main result.*

Proposition 4.4. *Keep the above notation and let H be a subgroup of finite index s in Γ. For every $Z \in \overline{\mathcal{H}}$ we have*

$$(2i)^{k-1} rs \langle f_1, f_2 \rangle_G = \sum_{1 \le m \le n} \mathcal{J}\big(A_{i_m}, A_{i_m+1}; Z, \alpha_{i_m}^{-1}(Z)\big).$$

Proof. By Corollary 3.5 and Lemma 3.4(4), we have

$$\begin{aligned}(2i)^{k-1} rs \langle f_1, f_2 \rangle_G &= \sum_{1 \le m \le n} \left(\int_{A_{i_m}}^{A_{i_m+1}} - \int_{\alpha_{i_m}(A_{i_m})}^{\alpha_{i_m}(A_{i_m+1})} \right) \sum_{1 \le j \le r} f_{1,j}(\tau) F_{2,j}(\tau)\, d\tau \\ &= \sum_{1 \le m \le n} \int_{A_{i_m}}^{A_{i_m+1}} \sum_{1 \le j \le r} f_{1,j}(\tau) P_j(\alpha_{i_m}; \tau)\, d\tau,\end{aligned}$$

proving the proposition using the definition of P_j and $\mathcal{J}$. □

Since we have seen that $\mathcal{J}$ is not a "true" double integral but an explicit finite linear combination of products of two simple integrals, we see that we have achieved our goal of expressing PSP's in terms of simple integrals. In the next section, we will specialize this formula to the fundamental domains given above.

5. The main corollaries

5A. ***General formulas.*** From the above proposition, we can deduce infinitely many expressions of PSP's in terms of simple integrals. We give a few here.

Theorem 5.1. *Assume that f_1 and f_2 are in $M_k(G, v)$, one of them being a cusp form. Then for all Z, we have*

$$\begin{aligned}
(2i)^{k-1} r\langle f_1, f_2\rangle_G &= \mathcal{J}\big(\rho, i\infty; Z-1, \quad Z\big) + \mathcal{J}\big(\ \rho, \quad i; \quad Z, \quad -\tfrac{1}{Z}\big) \\
&= \mathcal{J}\big(i, i\infty; \quad Z, \ -\tfrac{1}{Z}\big) + \mathcal{J}\big(\ \rho, i\infty; -\tfrac{Z+1}{Z}, \quad Z\big) \\
&= \Big(\mathcal{J}\big(\rho, i\infty; Z-1, \quad Z\big) + \mathcal{J}\big(\ \rho, i\infty; -\tfrac{Z+1}{Z}, \quad -\tfrac{1}{Z}\big)\Big)/2 \\
&= \Big(\mathcal{J}\big(0, i\infty; \quad Z, Z+1\big) + \mathcal{J}\big(-1, \quad \rho; \quad Z, -\tfrac{1}{Z+1}\big)\Big)/2 \\
&= \Big(\mathcal{J}\big(0, i\infty; \quad Z, \ \tfrac{Z}{Z+1}\big) + \mathcal{J}\big(\ \rho, i\infty; -\tfrac{1}{Z+1}, \quad Z\big)\Big)/2 \\
&= \Big(\mathcal{J}\big(0, i\infty; Z-1, Z+1\big) + \mathcal{J}\big(-1, \quad i; \quad Z, \quad -\tfrac{1}{Z}\big)\Big)/3 \\
&= \Big(\mathcal{J}\big(\rho, \quad 0; \ \tfrac{Z}{Z+1}, \tfrac{Z}{1-2Z}\big) + \mathcal{J}\big(\ \rho, \quad 1; \quad \tfrac{Z-1}{Z}, \quad \tfrac{Z}{Z+1}\big)\Big)/4 \\
&= \Big(\mathcal{J}\big(0, i\infty; Z-1, Z+1\big) + \mathcal{J}\big(-1, \quad 0; \quad Z, \ \tfrac{Z}{1-2Z}\big)\Big)/6 \\
&= \Big(\mathcal{J}\big(0, i\infty; Z-1, Z+1\big) + \mathcal{J}\big(\ 0, i\infty; -\tfrac{Z+1}{Z}, \quad \tfrac{Z-1}{Z}\big)\Big)/6.
\end{aligned}$$

In particular, we have

$$\begin{aligned}
(2i)^{k-1} r\langle f_1, f_2\rangle_G &= \mathcal{J}(i, \quad \rho; \quad 0, \quad i\infty) \quad = \mathcal{J}(i, i\infty; \quad \rho, \rho+1) \\
&= \mathcal{J}(\rho, i\infty; i-1, \quad i) \quad = \mathcal{J}(\rho, i\infty; \quad -1, \quad 0)/2 \\
&= \mathcal{J}(\rho, i\infty; \rho-1, \rho+1)/2 = \mathcal{J}(0, i\infty; \quad \rho, \rho+1)/2 \\
&= \mathcal{J}(0, i\infty; \quad -1, \quad \rho)/2 = \mathcal{J}(0, i\infty; \quad -1, \rho+1)/4 \\
&= \mathcal{J}(0, i\infty; \quad -1, \quad i)/3 = \mathcal{J}(0, i\infty; i-1, \ i+1)/3 \\
&= \mathcal{J}(0, i\infty; \quad -1, \quad 1)/6
\end{aligned}$$

as well as

$$(2i)^{k-1} r\langle f_1, f_2\rangle_G = \big(\mathcal{J}(0, i\infty; -1, 0) - \mathcal{J}(-1, 0; 0, i\infty)\big)/6.$$

Proof. The first collection of formulas follows from the different subgroups H and corresponding fundamental domains given in the preceding section, together with Proposition 3.8, which expresses the Γ-invariance of $\mathcal{J}$. The formulas in the second collection are obtained from those in the first by specializing to specific values of Z and using Proposition 3.8 and transitivity of the function $\mathcal{J}$. The details are left to the reader. □

Note that even though the final formula in the theorem involves two evaluations of the function $\mathcal{J}$ instead of one, and so takes longer to compute, we have included

it because it is the only formula which is symmetrical in f_1 and f_2, and because it leads directly to Haberland's formulas, given below.

5B. ***Haberland's formulas for subgroups.*** Even though the above theorem is sufficient for computational needs, we now reach our goal of generalizing Haberland's formulas to general subgroups of finite index of Γ. Recall that for any cusp form f we let $r_n(f) = \int_0^{i\infty} \tau^n f(\tau)\,d\tau$ denote the n-th period of f, and that $T = \left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$.

Theorem 5.2. *If f_1 and f_2 are in $S_k(G,v)$, we have the formula*

$$6r(-2i)^{k-1}\langle f_1, f_2\rangle_G = \sum_{m+n\le k-2} \binom{k-2}{m+n}\binom{m+n}{m} M_{m,n}(f_1, f_2),$$

where

$$M_{m,n}(f_1, f_2) = \sum_{1\le j\le r} \left((-1)^m r_m(f_{1,j})\overline{r_n(f_{2,j}\,|_k\,T)} - (-1)^n r_m(f_{1,j}\,|_k\,T)\overline{r_n(f_{2,j})}\right),$$

and where we recall that $f_{i,j} = f_i\,|_k\,\gamma_j$. In particular, we have

$$-6r(-2i)^{k-2}\langle f, f\rangle_G = \sum_{m+n\le k-2} \binom{k-2}{m+n}\binom{m+n}{m} \sum_{1\le j\le r} (-1)^m \Im\left(r_m(f_{1,j})\overline{r_n(f_{2,j}\,|_k\,T)}\right).$$

Proof. As already mentioned, by the binomial theorem we have

$$\mathcal{J}(-1,0;0,i\infty) = \sum_{1\le j\le r}\ \sum_{0\le n\le k-2} (-1)^n \binom{k-2}{n}\overline{r_n(f_{2,j})} \int_{-1}^{0} \tau^{k-2-n} f_{1,j}(\tau)\,d\tau.$$

Setting $\tau = -1/(z+1) = ST(z) = U(z)$, we have

$$\begin{aligned}\int_{-1}^{0} \tau^{k-2-n} f_{1,j}(\tau)\,d\tau &= (-1)^{k-2-n}\int_0^{i\infty} (z+1)^n f_{1,j}\,|_k\,U(z)\,dz\\ &= (-1)^{k-2-n}\sum_{0\le m\le n}\binom{n}{m} r_m(f_{1,j}\,|_k\,U),\end{aligned}$$

so using the trivial equality $r_{k-2-n}(f) = (-1)^{k-1-n} r_n(f\,|_k\,S)$, we obtain

$$\begin{aligned}&\mathcal{J}(-1,0;\,0,i\infty)\\ &\quad= (-1)^{k-2}\sum_{0\le m\le n\le k-2}\binom{k-2}{n}\binom{n}{m}\sum_{1\le j\le r} r_m(f_{1,j}\,|_k\,U)\overline{r_n(f_{2,j})}\\ &\quad= \sum_{0\le m\le n\le k-2}(-1)^{n+1}\binom{k-2}{n}\binom{n}{m}\sum_{1\le j\le r} r_m(f_{1,j}\,|_k\,U)\overline{r_{k-2-n}(f_{2,j}\,|_k\,S)}.\end{aligned}$$

By Lemma 3.7(2), $\mathcal{J}$ does not depend on the chosen representatives of right cosets, so replacing γ_j by $\gamma_j S$ and then changing n into $k-2-n$ gives

$$\mathcal{J}(-1,0;0,i\infty) = \sum_{m+n\le k-2} (-1)^{k-1-n}\binom{k-2}{m+n}\binom{m+n}{m} \sum_{1\le j\le r} r_m(f_{1,j}\,|_k\,T)\,\overline{r_n(f_{2,j})}.$$

By symmetry, we have

$$\mathcal{J}(0,i\infty;-1,0) = \sum_{m+n\le k-2} (-1)^{k-1-m}\binom{k-2}{m+n}\binom{m+n}{m} \sum_{1\le j\le r} r_m(f_{1,j})\,\overline{r_n(f_{2,j}\,|_k\,T)},$$

so the last formula of Theorem 5.1 gives us the first formula of Theorem 5.2. The second formula of Theorem 5.2 follows immediately. □

Even though we will not need the following proposition, note that it can be proved in the same way.

Proposition 5.3. *Under the same assumptions as above, we have*

$$\sum_{m+n\le k-2}\binom{k-2}{m+n}\binom{m+n}{m} \cdot \sum_{1\le j\le r}\left((-1)^m r_m(f_{1,j})\,\overline{r_n(f_{2,j}\,|_k\,T)} + (-1)^n r_m(f_{1,j}\,|_k\,T)\,\overline{r_n(f_{2,j})}\right)$$
$$= \sum_{1\le j\le r}\ \sum_{m+n=k-2} (-1)^m\binom{k-2}{m} r_m(f_{1,j})\,\overline{r_n(f_{2,j})}.$$

Proof. Simply expand as above the identity

$$\mathcal{J}(-1,0;0,i\infty) + \mathcal{J}(0,i\infty;-1,0) = -\mathcal{J}(0,i\infty;0,i\infty).$$ □

Corollary 5.4 (Haberland). *Assume that $G=\Gamma$, so that $r=1$, $v=1$, and k is even. We have*

$$3(-2i)^{k-1}\langle f_1, f_2\rangle = \sum_{\substack{m+n\le k-2\\ m+n\equiv 1 \pmod 2}} \binom{k-2}{m+n}\binom{m+n}{m}(-1)^m r_m(f_1)\,\overline{r_n(f_2)},$$

and

$$\sum_{\substack{m+n\le k-2\\ m+n\equiv 0 \pmod 2}}{}' \binom{k-2}{m+n}\binom{m+n}{m}(-1)^m r_m(f_1)\,\overline{r_n(f_2)} = 0,$$

where $\sum'$ means that the term $m+n=k-2$ occurs with coefficient $1/2$. □

6. Using Theorem 5.1

We now consider methods for computing PSP's based on the results obtained above. First, let us consider one of the formulas of Theorem 5.1, for instance the formula

$$6r(2i)^{k-1}\langle f_1, f_2\rangle_G = \mathcal{J}(0, i\infty; -1, 1).$$

Once again we will assume for simplicity that $G = \Gamma$ but the reasoning is completely general. We have

$$\mathcal{J}(0, i\infty; -1, 1) = \sum_{0\le n\le k-2} (-1)^n \binom{k-2}{n} \int_0^{i\infty} \tau^{k-2-n} f_1(\tau)\, d\tau \overline{\int_{-1}^{1} \tau^n f_2(\tau)\, d\tau},$$

so the problem boils down to the computation of $k-1$ integrals involving f_1 and $k-1$ integrals involving f_2 (in the general case, this becomes $r(k-1)$ integrals).

The computation of $\int_0^{i\infty} \tau^{k-2-n} f(\tau)\, d\tau = r_{k-2-n}(f)$ can be done in two quite different ways. On the one hand, we can apply the above-mentioned theory of double-exponential integration, which here works very well since it is only a *simple* and not a double integral.

An important implementation remark must be noted here: Since $f(\tau)$ may be costly to compute, it is preferable to use the integration method on the *vector-valued* function $(1, \tau, \ldots, \tau^{k-2}) f(\tau)$ or on the *polynomial-valued* function $(X-\tau)^{k-2} f(\tau)$, instead of on each component individually, since this only requires one evaluation of f instead of $k-1$.

On the other hand, we can use the elementary link between this integral and the value of the Λ-function attached to f: Indeed, we have trivially

$$r_j(f) = i^{j+1}\Lambda(f, j+1),$$

where $\Lambda(f, s) = (2\pi)^{-s}\Gamma(s)L(f, s)$ satisfies the functional equation

$$\Lambda(f, k-s) = (-1)^{k/2}\Lambda(f, s).$$

Thus, using the standard method explained above, but here in a *much* simpler context because the inverse Mellin transform of $(2\pi)^{-s}\Gamma(s)$ is simply $e^{-2\pi x}$, we obtain the formula

$$\Lambda(f, s) = \sum_{n\ge 1} \frac{a(n)}{(2\pi n)^s} \Gamma(s, 2\pi n t_0) + (-1)^{k/2} \sum_{n\ge 1} \frac{a(n)}{(2\pi n)^{k-s}} \Gamma(k-s, 2\pi n/t_0),$$

where

$$\Gamma(s, x) = \int_x^{\infty} e^{-t} t^{s-1}\, dt$$

is the incomplete gamma function, which can be computed in many efficient ways.

The computation of $\int_{-1}^{1} \tau^n f(\tau)\, d\tau$ poses slightly different problems. We can

of course still use double-exponential integration. On the other hand, the link with L-functions still exists but is slightly more subtle (unless $G = \Gamma$). Indeed, we first write $\int_{-1}^{1} = \int_{-1}^{0} + \int_{0}^{1}$, and then set $\tau = ST(z) = -1/(z+1)$ in the first integral and $\tau = z/(z+1)$ in the second integral. We obtain

$$\int_{-1}^{1} \tau^n f(\tau)\, d\tau = (-1)^n \int_{0}^{i\infty} (z+1)^{k-2-n} f\big(-1/(z+1)\big)\, dz + \int_{0}^{i\infty} z^n (z+1)^{k-2-n} f\big(z/(z+1)\big)\, dz.$$

If $G = \Gamma$ then the transforms of f are equal to f, so by using the binomial theorem we reduce the computation to that of at most $k-1$ periods of f. If desired we can in fact directly use Haberland's formula; see below.

If $G \neq \Gamma$, a new difficulty appears: Since the transforms of f by Γ are not in general equal to f, we have to compute their periods. The doubly exponential integration method is of course always available, but the use of the L-function explained above now requires the knowledge of the Fourier expansions at infinity of the functions $f_j = f|_k \gamma_j$, using the notation of the beginning of this section; equivalently, given $f \in M_k(G, v)$ in some way, we need to compute the Fourier expansion of f at the *cusps* of G, not only at infinity. This is still another computational problem which we do not consider here.

Table 4 presents some timings to compute $\langle f, f\rangle_G$ to the given number N of decimals using this method, without using at all the functional equation but only double-exponential integration, so as to keep it as general as possible. Note that in my implementation, the fastest among the formulas given by Theorem 5.1 for Δ, Δ_5, and Δ_{11} is the one given above involving $\mathcal{J}(0, i\infty; -1, 1)$, but this may not be the case for other implementations.

As an illustration of the power of double-exponential integration, note that for instance to compute $\langle \Delta, \Delta\rangle$ to 500 decimal digits, we only need 500 sample points, so only 1000 evaluations of Δ (which is of course efficiently computed using the equality $\Delta(\tau) = \eta^{24}(\tau)$).

To summarize, in order to use Theorem 5.1 in the simplest possible manner, I suggest using the doubly exponential integration methods, since here they only apply to simple integrals.

f	$N = 19$	38	57	96	250	500
Δ	0.06	0.06	0.14	0.19	2.02	11.3
Δ_5	0.35	0.46	1.16	1.60	17.1	94.3
Δ_{11}	0.67	0.89	2.24	3.11	33.7	188

Table 4. Timings (in seconds) to compute $\langle f, f\rangle_G$ to N decimal places using Theorem 5.1.

f	$N = 19$	38	57	96	250	500
Δ	0.02	0.02	0.06	0.08	0.86	4.96
Δ_5	0.23	0.29	0.72	1.00	10.5	58.4
Δ_{11}	0.48	0.61	1.48	2.12	22.0	122.5

Table 5. Timings (in seconds) to compute $\langle f, f\rangle_G$ to N decimal places using Theorem 5.2.

7. Using Theorem 5.2

As mentioned above, a variant is to directly use Theorem 5.2. This should be done in the following way: Using either double-exponential integration or the L-function method if available, we compute the $(k-1)r$ periods $r_m(f_{1,j})$, as well as the $(k-1)r$ periods $r_n(f_{2,j})$ if $f_1 \neq f_2$ (as mentioned above, these should be computed as r *vectors* with $k-1$ components). It is *not* necessary to compute the periods of $f_{1,j}|_k T$ and $f_{2,j}|_k T$. Indeed, we can write $\gamma_j T = g_j \gamma_{t(j)}$, where $g_j \in G$ and $j \mapsto t(j)$ is a permutation of $[1, r]$. Thus, since $f_1 \in M_k(G, v)$, we have

$$r_m(f_{1,j}|_k T) = r_m(f_1|_k \gamma_j T) = v(g_j) r_m(f_{1,t(j)}),$$

so no additional computation is necessary. Table 5 gives the corresponding timings.

Note that the main gain compared to the use of Theorem 5.1 comes from the fact that since $f_2 = f_1$, the periods have to be computed only once.

8. Using rationality theorems

There is a more subtle way of using periods to compute Petersson scalar products, but only in the special case of Hecke eigenforms: It is a well-known theorem of Manin that in the case of $G = \Gamma$, if f is a normalized eigenform there exist positive real numbers ω^+ and ω^- such that the even (respectively, odd) periods are algebraic multiples of ω^+ (respectively, of ω^-), and that ω^+ and ω^- can be chosen such that $\langle f, f\rangle = \omega^+\omega^-$. Since ω^+ and ω^- are essentially periods, they are thus very easy to compute as explained above, so this gives a very efficient way of computing $\langle f, f\rangle$. For instance, once one knows that

$$\langle \Delta, \Delta\rangle = \frac{225}{2048i} r_1(\Delta) r_2(\Delta),$$

without using any tricks and computing the periods using the doubly exponential integration method, we obtain the result to 500 decimals in only 9 seconds, while using the L-function method we obtain the result in 1 second, so there is no special advantage in this case.

However, in the case of congruence subgroups G of Γ, similar results hold, and here we may use rationality to our advantage. I thank N. Skoruppa for the precise statement of this theorem.

Theorem 8.1. *Denote by*

$$\gamma_j^+ = \begin{pmatrix} a_j & b_j \\ c_j & d_j \end{pmatrix}$$

a system of representatives of right cosets of $G\backslash\Gamma$. *Set*

$$\gamma_j^- = \begin{pmatrix} -b_j & -a_j \\ d_j & c_j \end{pmatrix} = P^{-1}\gamma_j SP,$$

where $P = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, *and for* $f \in M_k(G, v)$ *write* $f_j^{\pm} = f\,|_k\,\gamma_j^{\pm}$. *Finally, let*

$$R_j^{\pm}(f)(X) = \int_0^{i\infty} (X \mp \tau)^{k-2} f_j^{\pm}\, d\tau,$$

and

$$P_j^{\pm}(f) = R_j^+(f) \pm R_j^-(f).$$

Assume that f *is a normalized eigenfunction of all Hecke operators, so that the Fourier coefficients of* f *at infinity are algebraic, and denote by* $K = \mathbb{Q}(f)$ *the number field generated by them. There exist complex numbers* $\omega^{\pm}$ *such that the coefficients of the polynomials* $P_j^{\pm}(f)(X)/\omega^{\pm}$ *are in* K. *In addition,* $\omega^{\pm}$ *can be chosen so that* $\omega^+\omega^- = \langle f, f\rangle$.

Remarks. (1) I do not know if this theorem is stated explicitly in the literature, although it certainly is implicit.

(2) I thank an anonymous referee for pointing out that a similar theorem is valid with $\gamma_j^- = \begin{pmatrix} a_j & -b_j \\ -c_j & d_j \end{pmatrix} = P^{-1}\gamma_j P$ instead.

For $f = \Delta$, as mentioned above we choose for instance $\omega^+ = r_2(\Delta)/i$ and $\omega^- = r_1(\Delta)$, and we have

$$\langle \Delta, \Delta\rangle = (225/2048)\omega^+\omega^-.$$

For $f = \Delta_5$, we choose for instance $\omega^+ = r_0(\Delta_5)/i$ and $\omega^- = r_1(\Delta_5)$, and we have

$$\langle \Delta_5, \Delta_5\rangle = -(13/24)\omega^+\omega^-.$$

For $f = \Delta_{11}$, we choose for instance $\omega^+ = r_0(\Delta_{11})/i$ and $\omega^- = \Re\big(r_0\big(\Delta_{11}; \left(\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix}\right)\big)\big)$ (which is one of the simplest choices), and we have

$$\langle \Delta_{11}, \Delta_{11}\rangle = (5/12)\omega^+\omega^-.$$

Table 6 gives the timings.

f	$N = 19$	38	57	96	250	500
Δ	0.013	0.017	0.043	0.063	0.75	4.41
Δ_5	0.023	0.028	0.071	0.103	1.20	7.07
Δ_{11}	0.06	0.09	0.20	0.28	3.08	17.58

Table 6. Timings (in seconds) to compute $\langle f, f\rangle_G$ to N decimal places using rationality theorems.

We see that this is by far the fastest method, especially when the index $r = [\Gamma : G]$ is large, since we only need to compute two periods. Its main disadvantages are first that it is applicable only to Hecke eigenforms, and second that we need to compute the rational (or algebraic) constants which occur for each form f, which we do not know how to give in closed form, although such a formula may well exist.

References

[1] Henri Cohen, *Advanced topics in computational number theory*, Graduate Texts in Mathematics, no. 193, Springer, New York, 2000. MR 2000k:11144

[2] _____, *Number theory*, *II*: *Analytic and modern tools*, Graduate Texts in Mathematics, no. 240, Springer, New York, 2007. MR 2008e:11002

[3] J. E. Cremona, *Computing the degree of the modular parametrization of a modular elliptic curve*, Math. Comp. **64** (1995), no. 211, 1235–1250. MR 95j:11047

[4] Klaus Haberland, *Perioden von Modulformen einer Variabler and Gruppencohomologie*, *I*, Math. Nachr. **112** (1983), 245–282. MR 85k:11022

[5] Loïc Merel, *Symboles de Manin et valeurs de fonctions L*, in Tschinkel and Zarhin [8], 2009, pp. 283–309. MR 2011d:11115

[6] Vicentiu Pasol and Alexandru A. Popa, *Modular forms and period polynomials*, 2012. arXiv 1202.5802 [math.NT]

[7] Alexandru A. Popa, *Rational decomposition of modular forms*, Ramanujan J. **26** (2011), no. 3, 419–435. MR 2860697

[8] Yuri Tschinkel and Yuri Zarhin (eds.), *Algebra*, *arithmetic*, *and geometry*: *In honor of Yu. I. Manin*, vol. 2, Progress in Math., no. 270, Birkhäuser, Boston, 2009. MR 2010k:00009

[9] Mark Watkins, *Computing the modular degree of an elliptic curve*, Experiment. Math. **11** (2002), no. 4, 487–502. MR 2004c:11091

[10] D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. **28** (1985), no. 3, 372–384. MR 86m:11041

HENRI COHEN: Henri.Cohen@math.u-bordeaux1.fr
Université Bordeaux I, Institut de Mathématiques de Bordeaux, 351 Cours de la Libération, 33405 Talence Cedex, France

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.271

msp

Approximate common divisors via lattices

Henry Cohn and Nadia Heninger

We analyze the multivariate generalization of Howgrave-Graham's algorithm for the approximate common divisor problem. In the m-variable case with modulus N and approximate common divisor of size N^β, this improves the size of the error tolerated from N^{β^2} to $N^{\beta(m+1)/m}$, under a commonly used heuristic assumption. This gives a more detailed analysis of the hardness assumption underlying the recent fully homomorphic cryptosystem of van Dijk, Gentry, Halevi, and Vaikuntanathan. While these results do not challenge the suggested parameters, a 2^{n^ε} approximation algorithm with $\varepsilon < 2/3$ for lattice basis reduction in n dimensions could be used to break these parameters. We have implemented the algorithm, and it performs better in practice than the theoretical analysis suggests.

Our results fit into a broader context of analogies between cryptanalysis and coding theory. The multivariate approximate common divisor problem is the number-theoretic analogue of multivariate polynomial reconstruction, and we develop a corresponding lattice-based algorithm for the latter problem. In particular, it specializes to a lattice-based list decoding algorithm for Parvaresh-Vardy and Guruswami-Rudra codes, which are multivariate extensions of Reed-Solomon codes. This yields a new proof of the list decoding radii for these codes.

1. Introduction

Given two integers, we can compute their greatest common divisor efficiently using Euclid's algorithm. Howgrave-Graham [28] formulated and gave an algorithm to solve an approximate version of this problem, asking the question "What if instead

MSC2010: primary 11Y16; secondary 94A60, 94B35.

Keywords: Coppersmith's algorithm, lattice basis reduction, fully homomorphic encryption, approximate common divisors, list decoding, Parvaresh-Vardy codes, noisy polynomial reconstruction.

of exact multiples of some common divisor, we only know approximations?" In the simplest case, we are given one exact multiple $N = pq_0$ and one near multiple $a_1 = pq_1 + r_1$, and the goal is to learn p, or at least $p \gcd(q_0, q_1)$.

In this paper, we generalize Howgrave-Graham's approach to the case when one is given many near multiples of p. The hardness of solving this problem for small p (relative to the size of the near multiples) was recently proposed as the foundation for a fully homomorphic cryptosystem [21]. Specifically, we can show that improving the approximation of lattice basis reduction for the particular lattices L we are looking at from $2^{\dim L}$ to $2^{(\dim L)^\varepsilon}$ with $\varepsilon < 2/3$ would break the suggested parameters in the system. See Section 3 for the details. The approximate common divisor problem is also closely related to the problem of finding small solutions to multivariate polynomials, a problem first posed by Coppersmith [15], and whose various extensions have many applications in cryptanalysis [9].

The multivariate version of the problem allows us to improve the bounds for when the approximate common divisor problem is solvable. Given $N = pq_0$ and m randomly chosen approximate multiples $a_i = pq_i + r_i$ of $p = N^\beta$, as well as upper bounds X_i for each $|r_i|$, we can find the perturbations r_i when

$$\sqrt[m]{X_1 \cdots X_m} < N^{(1+o(1))\beta^{(m+1)/m}}.$$

In other words, we can compute approximate common divisors when r_i is as large as $N^{\beta^{(m+1)/m}}$. For $m = 1$, we recover Howgrave-Graham's theorem [28], which handles errors as large as N^{β^2}. As the number m of samples grows large, our bound approaches N^β, i.e., the size of the approximate common divisor p. The algorithm runs in polynomial time for fixed m. We cannot rigorously prove that it always works, but it is supported by a heuristic argument and works in practice.

There is an analogy between the ring of integers and the ring of polynomials over a field. Under this analogy, finding a large approximate common divisor of two integers is analogous to reconstructing a polynomial from noisy interpolation information, as we explain in Section 1.2.2. One of the most important applications of polynomial reconstruction is decoding of Reed-Solomon codes. Guruswami and Sudan [25] increased the feasible decoding radius of these codes by giving a list-decoding algorithm that outputs a list of polynomially many solutions to a polynomial reconstruction problem. The analogy between the integers and polynomials was used in [14] to give a proof of the Guruswami-Sudan algorithm inspired by Howgrave-Graham's approach, as well as a faster algorithm.

Parvaresh and Vardy [40] developed a related family of codes with a larger list-decoding radius than Reed-Solomon codes. The decoding algorithm corresponds to simultaneous reconstruction of several polynomials.

In this paper, we observe that the problem of simultaneous reconstruction of multiple polynomials is the exact analogue of the approximate common divisor

problem with many inputs, and the improved list-decoding radius of Parvaresh-Vardy codes corresponds to the improved error tolerance in the integer case. We adapt the algorithm for the integers to give a corresponding algorithm to solve the multiple polynomial reconstruction problem.

This algorithm has recently been applied to construct an optimally Byzantine-robust private information retrieval protocol [20]. The polynomial lattice methods we describe are extremely fast in practice, and they speed up the client-side calculations by a factor of several thousand compared with a related scheme that uses the Guruswami-Sudan algorithm. See [20] for more information and timings.

1.1. ***Related work.*** Howgrave-Graham first posed the problem of approximate integer common divisors in [28], and used it to address the problem of factoring when information is known about one of the factors. His algorithm gave a different viewpoint on Coppersmith's proof [15] that one can factor an RSA modulus $N = pq$ where $p \approx q \approx \sqrt{N}$ given the most significant half of the bits of one of the factors. This technique was applied by Boneh, Durfee, and Howgrave-Graham [10] to factor numbers of the form $p^r q$ with r large. Jochemsz and May [29] and Jutla [30] considered the problem of finding small solutions to multivariate polynomial equations, and showed how to do so by obtaining several equations satisfied by the desired roots using lattice basis reduction. Herrmann and May [26] gave a similar algorithm in the case of finding solutions to multivariate linear equations modulo divisors of a given integer. They applied their results to the case of factoring with bits known when those bits might be spread across $\log \log N$ chunks of p. Notably, their results display similar behavior to ours as the number of variables grows large. Sarkar and Maitra [45] studied the multivariate extension of Howgrave-Graham's method and applied it to the problem of implicit factorization.

Most relevantly, van Dijk, Gentry, Halevi, and Vaikuntanathan [21] discussed extensions of Howgrave-Graham's method to larger m and provided a rough heuristic analysis in Appendix B.2 of the longer version of their paper available on the Cryptology ePrint Archive. In particular, they carried out the calculation using the parameter settings $t = k = 2$ from Section 2 below and estimating the determinant by the product of row lengths. They briefly sketched how to extend it to $t = k = d$ for larger values of d. However, they did not optimize the choice of parameters or provide a detailed analysis. They concluded that including products of pairs of equations does worse than the original Howgrave-Graham attack and does not threaten their parameter choices.

Chen and Nguyen [13] gave an algorithm to find approximate common divisors which is not related to the Coppersmith/Howgrave-Graham lattice techniques and which provides an exponential speedup compared with exhaustive search over the possible perturbations.

In addition to the extensive work on polynomial reconstruction and noisy polynomial interpolation in the coding theory literature, the problem in both the single and multiple polynomial cases has been used as a cryptographic primitive, for example in [33], [32], and [3] (broken in [17]). Coppersmith and Sudan [16] gave an algorithm for simultaneous reconstruction of multiple polynomials, assuming random (rather than adversarially chosen) errors. Bleichenbacher, Kiayias, and Yung [7] gave a different algorithm for simultaneous reconstruction of multiple polynomials under a similar probabilistic model. Parvaresh and Vardy [40] were the first to beat the list-decoding performance of Reed-Solomon codes for adversarial errors, by combining multiple polynomial reconstruction with carefully chosen constraints on the polynomial solutions; this allowed them to prove that their algorithm ran in polynomial time, without requiring any heuristic assumptions. Finally, Guruswami and Rudra [24] combined the idea of multipolynomial reconstruction with an optimal choice of polynomials to construct codes that can be list-decoded up to the information-theoretic bound (for large alphabets).

1.2. *Problems and results.*

1.2.1. *Approximate common divisors.* Following Howgrave-Graham, we define the "partial" approximate common divisor problem to be the case when one has $N = pq_0$ and m approximate multiples $a_i = pq_i + r_i$ of p. We want to recover an approximate common divisor. To do so, we will compute $r_1, \dots, r_m$, after which we can simply compute the exact greatest common divisor of $N, a_1 - r_1, \dots, a_m - r_m$.

If the perturbations r_i are allowed to be as large as p, then it is clearly impossible to reconstruct p from this data. If they are sufficiently small, then one can easily find them by a brute force search. The following theorem interpolates between these extremes. As m grows, the bound on the size of r_i approaches the trivial upper bound of p.

Theorem 1 (Partial approximate common divisors). *Given positive integers N, $a_1, \dots, a_m$ and bounds $\beta \gg 1/\sqrt{\log N}$ and $X_1, \dots, X_m$, we can find all $r_1, \dots, r_m$ such that*

$$\gcd(N, a_1 - r_1, \dots, a_m - r_m) \geq N^{\beta}$$

and $|r_i| \leq X_i$, provided that

$$\sqrt[m]{X_1 \cdots X_m} < N^{(1+o(1))\beta^{(m+1)/m}}$$

and that the algebraic independence hypothesis discussed in Section 2 holds. The algorithm runs in polynomial time for fixed m, and the $\gg$ and $o(1)$ are as $N \to \infty$.

For $m = 1$, this theorem requires no algebraic independence hypothesis and is due to Howgrave-Graham [28]. For $m > 1$, not all inputs $N, a_1, \dots, a_m$ will satisfy the hypothesis. Specifically, we must rule out attempting to improve on

the $m = 1$ case by deriving $a_2, \ldots, a_m$ from a_1, for example by taking a_i to be a small multiple of a_1 plus an additional perturbation (or, worse yet, $a_1 = \cdots = a_m$). However, we believe that generic integers will work, for example integers chosen at random from a large range, or at least integers giving independent information in some sense.

We describe the algorithm to solve this problem in Section 2. We follow the general technique of Howgrave-Graham; that is, we use LLL lattice basis reduction to construct m polynomials for which $r_1, \ldots, r_m$ are roots, and then we solve the system of equations. The lattice basis reduction is for a lattice of dimension at most $\beta \log N$, regardless of what m is, but the root finding becomes difficult when m is large.

This algorithm is heuristic, because we assume we can obtain m short lattice vectors representing algebraically independent polynomials from the lattice that we will construct. This assumption is commonly made when applying multivariate versions of Coppersmith's method, and has generally been observed to hold in practice. See Section 2 for more details. This is where the restriction to generic inputs becomes necessary; if $a_1, \ldots, a_m$ are related in trivial ways, then the algorithm will simply recover the corresponding relations between $r_1, \ldots, r_m$, without providing enough information to solve for them.

Note that we are always able to find one nontrivial algebraic relation between $r_1, \ldots, r_m$, because LLL will always produce at least one short vector. If we were provided in advance with $m - 1$ additional relations, carefully chosen to ensure that they would be algebraically independent of the new one, then we would have no need for heuristic assumptions. We will see later in this section that this situation arises naturally in coding theory, namely in Parvaresh-Vardy codes [40].

The condition $\beta \gg 1/\sqrt{\log N}$ arises from the exponential approximation factor in LLL. It amounts to $N^{\beta^2} \gg 1$. An equivalent formulation is $\log p \gg \sqrt{\log N}$; i.e., the number of digits in the approximate common factor p must be more than the square root of the number of digits in N. When $m = 1$, this is not a restriction at all, because when p is small enough that N^{β^2} is bounded, there are only a bounded number of possibilities for r_1 and we can simply try all of them. When $m > 1$, the multivariate algorithm can handle much larger values of r_i for a given p, but the $\log p \gg \sqrt{\log N}$ condition dictates that p cannot be any smaller than when $m = 1$. Given a lattice basis reduction algorithm with approximation factor $2^{(\dim L)^\varepsilon}$, one could replace this condition with $\beta^{1+\varepsilon} \log N \gg 1$. If $\varepsilon = 1/m$, then the constraint could be removed entirely in the m-variable algorithm. See Section 2 for the details.

The $\log p \gg \sqrt{\log N}$ condition is the only thing keeping us from breaking the fully homomorphic encryption scheme from [21]. Specifically, improving the approximation of lattice basis reduction for the particular lattices L we are looking

at to $2^{(\dim L)^\varepsilon}$ with $\varepsilon < 2/3$ would break the suggested parameters in the system. See Section 3 for the details.

We get nearly the same bounds for the "general" approximate common divisor problem, in which we are not given the exact multiple N.

Theorem 2 (General approximate common divisors). *Given positive integers $a_1, \dots, a_m$ (with $a_i \approx N$ for all i) and bounds $\beta \gg 1/\sqrt{\log N}$ and X, we can find all $r_1, \dots, r_m$ such that*

$$\gcd(a_1 - r_1, \dots, a_m - r_m) \geq N^\beta$$

and $|r_i| \leq X$, provided that

$$X < N^{(C_m + o(1))\beta^{m/(m-1)}},$$

where

$$C_m = \frac{1 - 1/m^2}{m^{1/(m-1)}} \approx 1 - \frac{\log m}{m},$$

and that the algebraic independence hypothesis holds. The algorithm runs in polynomial time for fixed m, and the $\gg$ and $o(1)$ are as $N \to \infty$.

Again, for $m = 2$, this result is due to Howgrave-Graham [28], and no algebraic independence hypothesis is needed.

The proof is very similar to the case when N is known, but the calculations are more tedious because the determinant of the lattice is more difficult to bound. See Section 2.2 for the details.

In [28], Howgrave-Graham gave a more detailed analysis of the behavior for $m = 2$. Instead of our exponent $C_2\beta^2 = \frac{3}{8}\beta^2$, he obtained $1 - \beta/2 - \sqrt{1 - \beta - \beta^2/2}$, which is asymptotic to $\frac{3}{8}\beta^2$ for small β but is slightly better when β is large. We are interested primarily in the case when β is small, so we have opted for simplicity, but one could carry out a similar analysis for all m.

1.2.2. *Noisy multipolynomial reconstruction.* Let F be a field. Given m single-variable polynomials $g_1(z), \dots, g_m(z)$ over F and n distinct points $z_1, \dots, z_n$ in F, evaluating the polynomials at these points yields mn elements $y_{ij} = g_i(z_j)$ of F.

The noisy multipolynomial reconstruction problem asks for the recovery of $g_1, \dots, g_m$ given the evaluation points $z_1, \dots, z_n$, degree bounds ℓ_i on g_i, and possibly incorrect values y_{ij}. Stated more precisely, we wish to find all m-tuples of polynomials $(g_1, \dots, g_m)$ satisfying $\deg g_i \leq \ell_i$, for which there are at least βn values of j such that $g_i(z_j) = y_{ij}$ for all i. In other words, some of the data may have been corrupted, but we are guaranteed that there are at least βn points at which all the values are correct.

Bleichenbacher and Nguyen [8] distinguish the problem of "polynomial reconstruction" from the "noisy polynomial interpolation" problem. Their definition of "noisy polynomial interpolation" involves reconstructing a single polynomial when there are several possibilities for each value. The multivariate version of this problem can be solved using Theorem 5.

This problem is an important stepping stone between single-variable interpolation problems and full multivariate interpolation, in which we reconstruct polynomials of many variables. The multipolynomial reconstruction problem allows us to take advantage of multivariate techniques to prove much stronger bounds, without having to worry about issues such as whether our evaluation points are in general position.

We can restate the multipolynomial reconstruction problem slightly to make the analogy with the integer case clear. Given evaluation points z_j and values y_{ij}, define $N(z) = \prod_j (z - z_j)$, and use ordinary interpolation to find polynomials $f_i(z)$ such that $f_i(z_j) = y_{ij}$. Then we will see shortly that $g_1, \ldots, g_m$ solve the noisy multipolynomial reconstruction problem if and only if

$$\deg \gcd\big(f_1(z) - g_1(z), \ldots, f_m(z) - g_m(z), N(z)\big) \geq \beta n.$$

This is completely analogous to the approximate common divisor problem, with $N(z)$ as the exact multiple and $f_1(z), \ldots, f_m(z)$ as the approximate multiples.

To see why this works, observe that $g_i(z_j) = y_{ij}$ if and only if $g_i(z) - y_{ij}$ is divisible by $z - z_j$. Thus, $g_i(z_j) = f_i(z_j) = y_{ij}$ if and only if $f_i(z) - g_i(z)$ is divisible by $z - z_j$, and $\deg \gcd\big(f_i(z) - g_i(z), N(z)\big)$ counts how many j satisfy $g_i(z_j) = y_{ij}$. Finally, to count the j such that $g_i(z_j) = y_{ij}$ for all i, we use

$$\deg \gcd\big(f_1(z) - g_1(z), \ldots, f_m(z) - g_m(z), N(z)\big).$$

This leads us to our result in the polynomial case.

Theorem 3. *Given polynomials $N(z), f_1(z), \ldots, f_m(z)$, degree bounds $\ell_1, \ldots, \ell_m$, and $\beta \in [0, 1]$, we can find all $g_1(z), \ldots, g_m(z)$ such that*

$$\deg \gcd\big(f_1(z) - g_1(z), \ldots, f_m(z) - g_m(z), N(z)\big) \geq \beta \deg N(z)$$

and $\deg g_i \leq \ell_i$, *provided that*

$$\frac{\ell_1 + \cdots + \ell_m}{m} < \beta^{(m+1)/m} \deg N(z)$$

and that the algebraic independence hypothesis holds. The algorithm runs in polynomial time for fixed m.

As in the integer case, our analysis depends on an algebraic independence hypothesis, but it may be easier to resolve this issue in the polynomial case, because

lattice basis reduction is far more effective and easier to analyze over polynomial rings than it is over the integers.

Parvaresh-Vardy codes [40] are based on noisy multipolynomial reconstruction. A codeword is constructed by evaluating polynomials $f_1, \dots, f_m$ at points $z_1, \dots, z_n$ to obtain mn elements $f_i(z_j)$. In their construction, $f_1, \dots, f_m$ are chosen to satisfy $m-1$ polynomial relations, so that they only need to find one more algebraically independent relation to solve the decoding problem. Furthermore, the $m-1$ relations are constructed so that they must be algebraically independent from the relation constructed by the decoding algorithm. This avoids the need for the heuristic assumption discussed above in the integer case. Furthermore, the Guruswami-Rudra codes [24] achieve improved rates by constructing a system of polynomials so that only n symbols need to be transmitted, rather than mn.

Parvaresh and Vardy gave a list-decoding algorithm using the method of Guruswami and Sudan, which constructs a polynomial by solving a system of equations to determine the coefficients. In our terms, they proved the following theorem:

Theorem 4. *Given polynomials $N(z), f_1(z), \dots, f_m(z)$, degree bounds $\ell_1, \dots, \ell_m$, and $\beta \in [0,1]$ satisfying*

$$\frac{\ell_1 + \cdots + \ell_m}{m} < \beta^{(m+1)/m} \deg N(z),$$

we can find a nontrivial polynomial $Q(x_1, \dots, x_m)$ such that

$$Q(g_1(z), \dots, g_m(z)) = 0$$

for all $g_1(z), \dots, g_m(z)$ satisfying $\deg g_i \le \ell_i$ and

$$\deg \gcd\bigl(f_1(z) - g_1(z), \dots, f_m(z) - g_m(z), N(z)\bigr) \ge \beta \deg N(z).$$

The algorithm runs in polynomial time.

In Section 4, we give an alternative proof of this theorem using the analogue of lattice basis reduction over polynomial rings. This algorithm requires neither heuristic assumptions nor conditions on β.

2. Computing approximate common divisors

In this section, we describe the algorithm to solve the approximate common divisor problem over the integers.

To derive Theorem 1, we will use the following approach:

(1) Construct polynomials $Q_1, \dots, Q_m$ of m variables such that

$$Q_i(r_1, \dots, r_m) = 0$$

for all $r_1, \dots, r_m$ satisfying the conditions of the theorem.

(2) Solve this system of equations to learn candidates for the roots $r_1, \ldots, r_m$.

(3) Test each of the polynomially many candidates to see if it is a solution to the original problem.

In the first step, we will construct polynomials Q satisfying

$$Q(r_1, \ldots, r_m) \equiv 0 \pmod{p^k}$$

(for a k to be chosen later) whenever $a_i \equiv r_i \pmod{p}$ for all i. We will furthermore arrange that

$$|Q(r_1, \ldots, r_m)| < N^{\beta k}.$$

These two facts together imply that $Q(r_1, \ldots, r_m) = 0$ whenever $p \geq N^{\beta}$.

To ensure that $Q(r_1, \ldots, r_m) \equiv 0 \pmod{p^k}$, we will construct Q as an integer linear combination of products

$$(x_1 - a_1)^{i_1} \cdots (x_m - a_m)^{i_m} N^{\ell}$$

with $i_1 + \cdots + i_m + \ell \geq k$. Alternatively, we can think of Q as being in the integer lattice generated by the coefficient vectors of these polynomials. To ensure that $|Q(r_1, \ldots, r_m)| < N^{\beta k}$, we will construct Q to have small coefficients; i.e., it will be a short vector in the lattice.

More precisely, we will use the lattice L generated by the coefficient vectors of the polynomials

$$(X_1 x_1 - a_1)^{i_1} \cdots (X_m x_m - a_m)^{i_m} N^{\ell}$$

with $i_1 + \cdots + i_m \leq t$ and $\ell = \max\left(k - \sum_j i_j, 0\right)$. Here t and k are parameters to be chosen later. Note that we have incorporated the bounds $X_1, \ldots, X_m$ on the desired roots $r_1, \ldots, r_m$ into the lattice. We define Q to be the corresponding integer linear combination of $(x_1 - a_1)^{i_1} \cdots (x_m - a_m)^{i_m} N^{\ell}$, without $X_1, \ldots, X_m$.

Given a polynomial $Q(x_1, \ldots, x_m)$ corresponding to a vector $v \in L$, we can bound $|Q(r_1, \ldots, r_m)|$ by the ℓ_1 norm $|v|_1$. Specifically, if

$$Q(x_1, \ldots, x_m) = \sum_{j_1, \ldots, j_m} q_{j_1 \ldots j_m} x_1^{j_1} \cdots x_m^{j_m},$$

then v has entries $q_{j_1 \ldots j_m} X_1^{j_1} \cdots X_m^{j_m}$, and

$$\begin{aligned} |Q(r_1, \ldots, r_m)| &\leq \sum_{j_1, \ldots, j_m} |q_{j_1 \ldots j_m}| |r_1|^{j_1} \cdots |r_m|^{j_m} \\ &\leq \sum_{j_1, \ldots, j_m} |q_{j_1 \ldots j_m}| X_1^{j_1} \cdots X_m^{j_m} \\ &= |v|_1. \end{aligned}$$

Thus, every vector $v \in L$ satisfying $|v|_1 < N^{\beta k}$ gives a polynomial relation between $r_1, \dots, r_m$.

It is straightforward to compute the dimension and determinant of the lattice:

$$\dim L = \binom{t+m}{m},$$

and

$$\det L = (X_1 \cdots X_m)^{\binom{t+m}{m}\frac{t}{m+1}} N^{\binom{k+m}{m}\frac{k}{m+1}}.$$

To compute the determinant, we can choose a monomial ordering so that the basis matrix for this lattice is upper triangular; then the determinant is simply the product of the terms on the diagonal.

Now we apply LLL lattice basis reduction to L. Because all the vectors in L are integral, the m shortest vectors $v_1, \dots, v_m$ in the LLL-reduced basis satisfy

$$|v_1| \le \cdots \le |v_m| \le 2^{(\dim L)/4} (\det L)^{1/(\dim L + 1 - m)}$$

(see Theorem 2 in [26]), and $|v|_1 \le \sqrt{\dim L}\,|v|$ by Cauchy-Schwarz, so we know that the corresponding polynomials Q satisfy

$$|Q(r_1, \dots, r_m)| \le \sqrt{\dim L}\, 2^{(\dim L)/4} (\det L)^{1/(\dim L + 1 - m)}.$$

If

$$\sqrt{\dim L}\, 2^{(\dim L)/4} \det L^{1/(\dim L + 1 - m)} < N^{\beta k}, \tag{1}$$

then we can conclude that $Q(r_1, \dots, r_m) = 0$.

If t and k are large, then we can approximate $\binom{t+m}{m}$ with $t^m/m!$ and $\binom{k+m}{m}$ with $k^m/m!$. The $\sqrt{\dim L}$ factor plays no significant role asymptotically, so we simply omit it (the omission is not difficult to justify). After taking a logarithm and simplifying slightly, our desired equation (1) becomes

$$\frac{t^m}{4km!} + \frac{1}{1 - \frac{(m-1)m!}{t^m}} \left(\frac{m \log_2 X}{m+1} \frac{t}{k} + \frac{\log_2 N}{m+1} \frac{k^m}{t^m} \right) < \beta \log_2 N,$$

where X denotes the geometric mean of $X_1, \dots, X_m$.

The $t^m/(4km!)$ and $(m-1)m!/t^m$ terms are nuisance factors, and once we optimize the parameters they will tend to zero asymptotically. We will take $t \approx \beta^{-1/m} k$ and $\log X \approx \beta^{(m+1)/m} \log N$. Then

$$\frac{m \log X}{m+1} \frac{t}{k} + \frac{\log N}{m+1} \frac{k^m}{t^m} \approx \frac{m}{m+1} \beta \log N + \frac{1}{m+1} \beta \log N = \beta \log N.$$

By setting $\log X$ slightly less than this bound (by a $1 + o(1)$ factor), we can achieve the desired inequality, assuming that the $1 - (m-1)!/t^m$ and $t^m/(4km!)$ terms

do not interfere. To ensure that they do not, we take $t \gg m$ and $t^m \ll \beta \log N$ as $N \to \infty$. Note that then $\dim L \leq \beta \log N$, which is bounded independently of m.

Specifically, when N is large we can take

$$t = \left\lfloor \frac{(\beta \log N)^{1/m}}{(\beta^2 \log N)^{1/(2m)}} \right\rfloor$$

and

$$k = \lfloor \beta^{1/m} t \rfloor \approx (\beta^2 \log N)^{1/(2m)}.$$

With these parameter settings, t and k both tend to infinity as $N \to \infty$, because $\beta^2 \log N \to \infty$, and they satisfy the necessary constraints. We do not recommend using these parameter settings in practice; instead, one should choose t and k more carefully. However, these choices work asymptotically. Notice that with this approach, $\beta^2 \log N$ must be large enough to allow t/k to approximate $\beta^{-1/m}$. This is a fundamental issue, and we discuss it in more detail in the next subsection.

The final step of the proof is to solve the system of equations defined by the m shortest vectors in the reduced basis to learn $r_1, \ldots, r_m$. One way to do this is to repeatedly use resultants to eliminate variables; alternatively, we can use Gröbner bases. See, for example, Chapter 3 of [19].

One obstacle is that the equations may be not algebraically independent, in which case we will not have enough information to complete the solution. In the experiments summarized in Section 6, we sometimes encountered cases when the m shortest vectors were algebraically dependent. However, in every case the vectors represented either (1) irreducible, algebraically independent polynomials, or (2) algebraically dependent polynomials that factored easily into polynomials which all had the desired properties. Thus when the assumption of algebraic dependence failed, it failed because there were fewer than m independent factors among the m shortest relations. In these cases, there were always more than m vectors of ℓ_1 norm less than $N^{\beta k}$, and we were able to complete the solution by using all these vectors. This behavior appears to depend sensitively on the optimization of the parameters t and k.

2.1. ***The*** $\boldsymbol{\beta^2 \log N \gg 1}$ ***requirement.*** The condition that $\beta^2 \log N \gg 1$ is not merely a convenient assumption for the analysis. Instead, it is a necessary hypothesis for our approach to work at all when using a lattice basis reduction algorithm with an exponential approximation factor. In previous papers on these lattice-based techniques, such as [15] or [28], this issue seemingly does not arise, but that is because it is hidden in a degenerate case. When $m = 1$, we are merely ruling out the cases when the bound N^{β^2} on the perturbations is itself bounded, and in those cases the problem can be solved by brute force.

To see why a lower bound on $\beta^2 \log N$ is necessary, we can start with (1). For that equation to hold, we must at least have $2^{(\dim L)/4} < N^{\beta k}$ and $(\det L)^{1/(\dim L)} < N^{\beta k}$, and these inequalities imply that

$$\frac{1}{4}\binom{t+m}{m} < \beta k \log_2 N$$

and

$$\frac{\binom{k+m}{m} \log_2 N}{\binom{t+m}{m}(m+1)} < \beta \log_2 N.$$

Combining them with $\binom{k+m}{m} > k$ yields

$$\frac{1}{4(m+1)} < \beta^2 \log_2 N,$$

so we have an absolute lower bound for $\beta^2 \log N$. Furthermore, one can check that in order for the $2^{(\dim L)/4}$ factor to become negligible compared with $N^{\beta k}$, we must have $\beta^2 \log N \gg 1$.

Given a lattice basis reduction algorithm with approximation factor $2^{(\dim L)^\varepsilon}$, we could replace t^m with $t^{\varepsilon m}$ in the nuisance term coming from the approximation factor. Then the condition $t^m \ll \beta \log N$ would become $t^{\varepsilon m} \ll \beta \log N$, and if we combine this with $k \approx \beta^{1/m} t$, we find that

$$k^{\varepsilon m} \approx \beta^\varepsilon t^{\varepsilon m} \ll \beta^{1+\varepsilon} \log N.$$

Because $k \geq 1$, the condition $\beta^{1+\varepsilon} \log N \gg 1$ is needed, and then we can take

$$t = \left\lfloor \frac{(\beta \log N)^{1/(\varepsilon m)}}{(\beta^{1+\varepsilon} \log N)^{1/(2\varepsilon m)}} \right\rfloor$$

and

$$k = \lfloor \beta^{1/m} t \rfloor \approx (\beta^{1+\varepsilon} \log N)^{1/(2\varepsilon m)}.$$

2.2. ***Theorem 2.*** The algorithm for Theorem 2 is identical to the above, except that we do not have an exact N, so we omit all vectors involving N from the construction of the lattice L.

The matrix of coefficients is no longer square, so we have to do more work to bound the determinant of the lattice. Howgrave-Graham [28] observed in the two-variable case that the determinant is preserved even under nonintegral row operations, and he used a nonintegral transformation to hand-reduce the matrix before bounding the determinant as the product of the ℓ_2 norms of the basis vectors; furthermore, the ℓ_2 norms are bounded by $\sqrt{\dim L}$ times the ℓ_∞ norms.

The nonintegral transformation that he uses is based on the relation

$$(x_i - a_i) - \frac{a_i}{a_1}(x_1 - a_1) = x_i - \frac{a_i}{a_1}x_1.$$

Adding a multiple of $f(x)(x_1 - a_1)$ reduces $f(x)(x_i - a_i)$ to $f(x)\big(x_i - \frac{a_i}{a_1}x_1\big)$. The advantage of this is that if $x_1 \approx x_i$ and $a_1 \approx a_i$, then $x_i - (a_i/a_1)x_1$ may be much smaller than $x_i - a_i$ was. The calculations are somewhat cumbersome, and we will omit the details (see [28] for more information).

When $a_1, \ldots, a_m$ are all roughly N (as in Theorem 2), we get the following values for the determinant and dimension in the m-variable case:

$$\det L \le (N/X)^{\binom{k+m-1}{m}(t-k+1)} X^{m\left(\binom{t+m}{m}\frac{t}{m+1} - \binom{k-1+m}{m}\frac{k-1}{m+1}\right)}$$

and

$$\dim L = \binom{t+m}{m} - \binom{k-1+m}{m}.$$

To optimize the resulting bound, we take $t \approx (m/\beta)^{1/(m-1)}k$.

3. Applications to fully homomorphic encryption

In [21], the authors build a fully homomorphic encryption system whose security relies on several assumptions, among them the hardness of computing an approximate common divisor of many integers. This assumption is used to build a simple "somewhat homomorphic" scheme, which is then transformed into a fully homomorphic system under additional hardness assumptions. In this section, we use Theorem 1 to provide a more precise theoretical understanding of the security assumption underlying this somewhat homomorphic scheme, as well as the related cryptosystem of [18].

For ease of comparison, we will use the notation from the above two papers (see Section 3 of [21]). Let γ be the bit length of N, η be the bit length of p, and ρ be the bit length of each r_i. Using Theorem 1, we can find $r_1, \ldots, r_m$ and the secret key p when

$$\rho \le \gamma\beta^{(m+1)/m}.$$

Substituting in $\beta = \eta/\gamma$, we obtain

$$\rho^m\gamma \le \eta^{m+1}.$$

The authors of [21] suggest as a "convenient parameter set to keep in mind" to set $\rho = \lambda$, $\eta = \lambda^2$, and $\gamma = \lambda^5$. Using $m > 3$ we would be able to solve this parameter set, if we did not have the barrier that η^2 must be much greater than γ.

As pointed out in Section 1.2.1, this barrier would no longer apply if we could improve the approximation factor for lattice basis reduction. If we could improve

the approximation factor to $2^{(\dim L)^\varepsilon}$, then the barrier would amount to $\beta^{1+\varepsilon}\lambda^5 \gg 1$, where $\beta = \eta/\gamma = \lambda^{-3}$. If $\varepsilon < 2/3$, this would no longer be an obstacle. Given a $2^{(\dim L)^{2/3}/\log\dim L}$ approximation factor, we could take $m = 4$, $k = 1$, and $t = \lfloor 3\lambda^{3/4} \rfloor$ in the notation of Section 2. Then (1) holds, and thus the algorithm works, for all $\lambda \geq 300$.

One might try to achieve these subexponential approximation factors by using blockwise lattice reduction techniques [22]. For an n-dimensional lattice, one can obtain an approximation factor of roughly $\kappa^{n/\kappa}$ in time exponential in κ. For the above parameter settings, the lattice will have dimension on the order of λ^3, and even a $2^{n^{2/3}}$ approximation will require $\kappa > n^{1/3} = \lambda$, for a running time that remains exponential in λ. (Note that for these parameters, using a subexponential-time factoring algorithm to factor the modulus in the "partial" approximate common divisor problem is super-exponential in the security parameter.)

In general, if we could achieve an approximation factor of $2^{(\dim L)^\varepsilon}$ for arbitrarily small ε, then we could solve the approximate common divisor problem for parameters given by any polynomials in λ. Furthermore, as we will see in Section 6, the LLL algorithm performs better in practice on these problems than the theoretical analysis suggests.

In [18], Coron, Mandal, Naccache, and Tibouchi suggest explicit parameter sizes for a modified version of the scheme from [21]. The parameters are well within the range for which the algorithm works, assuming typical LLL performance. However, although our attacks run in time polynomial in the input size, the running time is dependent on the largest input (the total key size) and for these parameters the performance of the lattice-based approach is not competitive with attacks such as Chen and Nguyen [13], which run in time subexponential in the size of the error.

4. Multipolynomial reconstruction

4.1. ***Polynomial lattices.*** For Theorem 3 and Theorem 4, we can use almost exactly the same technique, but with lattices over the polynomial ring $F[z]$ instead of the integers.

By a d-dimensional lattice L over $F[z]$, we mean the $F[z]$-span of d linearly independent vectors in $F[z]^d$. The degree $\deg v$ of a vector v in L is the maximum degree of any of its components, and the determinant $\det L$ is the determinant of a basis matrix (which is well-defined, up to scalar multiplication).

The polynomial analogue of lattice basis reduction produces a basis $b_1, \ldots, b_d$ for L such that

$$\deg b_1 + \cdots + \deg b_d = \deg \det L.$$

Such a basis is called a reduced basis (sometimes column or row-reduced, depending on how the vectors are written), and it can be found in polynomial time; see,

for example, Section 6.3 in [31]. If we order the basis so that $\deg b_1 \le \cdots \le \deg b_d$, then clearly

$$\deg b_1 \le \frac{\deg \det L}{d},$$

and more generally

$$\deg b_i \le \frac{\deg \det L}{d-(i-1)},$$

because

$$\deg \det L - (d-(i-1)) \deg b_i = \sum_{j=1}^{d} \deg b_j - \sum_{j=i}^{d} \deg b_i \ge 0.$$

These inequalities are the polynomial analogues of the vector length bounds in LLL-reduced lattices, but notice that the exponential approximation factor does not occur. See [14] for more information about this analogy, and [20] for applications that demonstrate the superior performance of these methods in practice.

4.2. ***Theorems 3 and 4.*** In the polynomial setting, we will choose $Q(x_1, \ldots, x_m)$ to be a linear combination (with coefficients from $F[z]$) of the polynomials

$$(x_1 - f_1(z))^{i_1} \cdots (x_m - f_m(z))^{i_m} N(z)^{\ell}$$

with $i_1 + \cdots + i_m \le t$ and $\ell = \max(k - \sum_j i_j, 0)$. We define the lattice L to be spanned by the coefficient vectors of these polynomials, but with x_i replaced with $z^{\ell_i} x_i$ to incorporate the bound on $\deg g_i$, much as we replaced x_i with $X_i x_i$ in Section 2.

As before, we can easily compute the dimension and determinant of L:

$$\dim L = \binom{t+m}{m}$$

and

$$\deg \det L = (\ell_1 + \cdots + \ell_m) \binom{t+m}{m} \frac{t}{m+1} + n \binom{k+m}{m} \frac{k}{m+1},$$

where $n = \deg N(z)$.

Given a polynomial $Q(x_1, \ldots, x_m)$ corresponding to a vector $v \in L$, we can bound $\deg Q(g_1(z), \ldots, g_m(z))$ by $\deg v$. Specifically, suppose

$$Q(x_1, \ldots, x_m) = \sum_{j_1, \ldots, j_m} q_{j_1 \ldots j_m}(z) x_1^{j_1} \cdots x_m^{j_m};$$

then v is the vector whose entries are $q_{j_1 \dots j_m}(z) z^{j_1 \ell_1 + \cdots + j_m \ell_m}$, and

$$\begin{aligned}\deg Q(g_1(z), \dots, g_m(z)) &\le \max_{j_1,\dots,j_m} (\deg q_{j_1 \dots j_m}(z) + j_1 \deg g_1(z) + \cdots + j_m \deg g_m(z)) \\ &\le \max_{j_1,\dots,j_m} (\deg q_{j_1 \dots j_m}(z) + j_1 \ell_1 + \cdots + j_m \ell_m) \\ &= \deg v.\end{aligned}$$

Let $v_1, \dots, v_{\dim L}$ be a reduced basis of L, arranged in increasing order by degree. If

$$\frac{\deg \det L}{\dim L - (m-1)} < \beta k n, \tag{2}$$

then each of $v_1, \dots, v_m$ yields a polynomial relation Q_i such that

$$Q_i(g_1(z), \dots, g_m(z)) = 0,$$

because by the construction of the lattice, $Q_i(g_1(z), \dots, g_m(z))$ is divisible by the k-th power of an approximate common divisor of degree βn, while

$$\deg Q_i(g_1(z), \dots, g_m(z)) \le \deg v_i < \beta k n.$$

Thus we must determine how large $\ell_1 + \cdots + \ell_m$ can be, subject to the inequality (2).

If we set $t \approx k\beta^{-1/m}$ and

$$\frac{\ell_1 + \cdots + \ell_m}{m} < n\beta^{(m+1)/m},$$

then inequality (2) is satisfied when t and k are sufficiently large. Because there is no analogue of the LLL approximation factor in this setting, we do not have to worry about t and k becoming too large (except for the obvious restriction that $\dim L$ must remain polynomially bounded), and there is no lower bound on β. Furthermore, we require no $1 + o(1)$ factors, because all degrees are integers and all the quantities we care about are rational numbers with bounded numerators and denominators; thus, any sufficiently close approximation might as well be exact, and we can achieve this when t and k are polynomially large. More precisely, without loss of generality we can take βn to be an integer. Then the inequality

$$\frac{\ell_1 + \cdots + \ell_m}{m} < n\beta^{(m+1)/m}$$

is equivalent to $n(\ell_1 + \cdots + \ell_m)^m < (n\beta)^{m+1} m^m$ and hence

$$n(\ell_1 + \cdots + \ell_m)^m \le (n\beta)^{m+1} m^m - 1$$

by integrality. Thus, $(\ell_1 + \cdots + \ell_m)/m$ is smaller than $n\beta^{(m+1)/m}$ by at least a factor of $(1 - n^{-(m+1)}m^{-m})^{1/m}$, and this factor is enough to ensure that inequality (2) holds when t and k are only polynomially large.

5. Higher-degree polynomials

It is possible to generalize the results in the previous sections to find solutions of a system of higher-degree polynomials modulo divisors of N.

Theorem 5. *Given a positive integer N and m monic polynomials $h_1(x), \ldots, h_m(x)$ over the integers, of degrees $d_1, \ldots, d_m$, and given any $\beta \gg 1/\sqrt{\log N}$ and bounds $X_1, \ldots, X_m$, we can find all $r_1, \ldots, r_m$ such that*

$$\gcd(N, h_1(r_1), \ldots, h_m(r_m)) \geq N^{\beta}$$

and $|r_i| \leq X_i$, provided that

$$\sqrt[m]{X_1^{d_1} \cdots X_m^{d_m}} < N^{(1+o(1))\beta^{(m+1)/m}}$$

and that the algebraic independence hypothesis holds. The algorithm runs in polynomial time for fixed m.

The $m = 1$ case does not require the algebraic independence hypothesis, and it encompasses both Howgrave-Graham and Coppersmith's theorems [28; 15]; it first appeared in [36].

In the case where $X_1 = \cdots = X_m$, the bound becomes $N^{\beta^{(m+1)/m}/\bar{d}}$, where $\bar{d} = (d_1 + \cdots + d_m)/m$ is the average degree.

Theorem 6. *Given a polynomial $N(z)$ and m monic polynomials $h_1(x), \ldots, h_m(x)$ over $F[z]$, of degrees $d_1, \ldots, d_m$ in x, and given degree bounds $\ell_1, \ldots, \ell_m$ and $\beta \in [0, 1]$, we can find all $g_1(z), \ldots, g_m(z)$ in $F[z]$ such that*

$$\deg \gcd\big(N(z), h_1(g_1(z)), \ldots, h_m(g_m(z))\big) \geq \beta \deg N(z)$$

and $\deg g_i(z) \leq \ell_i$, provided that

$$\frac{\ell_1 d_1 + \cdots + \ell_m d_m}{m} < \beta^{(m+1)/m} \deg N(z)$$

and that the algebraic independence hypothesis holds. The algorithm runs in polynomial time for fixed m.

The algorithms are exactly analogous to those for the degree-1 cases, except that $x_i - a_i$ (or $x_i - f_i(z)$) is replaced with $h_i(x_i)$.

6. Implementation

We implemented the number-theoretic version of the partial approximate common divisor algorithm using Sage [47]. We used Magma [11] to do the LLL and Gröbner basis calculations.

We solved the systems of equations by computing a Gröbner basis with respect to the lexicographic monomial ordering, to eliminate variables. Computing a Gröbner basis can be extremely slow, both in theory and in practice. We found that it was more efficient to solve the equations modulo a large prime, to limit the bit length of the coefficients in the intermediate and final results. Because $r_1, \dots, r_m$ are bounded in size, we can simply choose a prime larger than $2 \max_i |r_i|$.

We ran our experiments on a computer with a 3.30 GHz quad-core Intel Core i5 processor and 8 GB of RAM. Table 1 shows a selection of sample running times (in seconds) for various parameter settings. For comparison, the table includes the $m = 1$ case, which is Howgrave-Graham's algorithm. The rows for which no timing information is listed give example lattice dimensions for larger inputs, in order to illustrate the limiting behavior of the algorithm.

The performance of the algorithm depends on the ratio of t to k, which should be approximately $\beta^{-1/m}$. Incorrectly optimized parameters often perform much worse than correctly optimized parameters. For example, when $m=3$, $\log_2 N=1000$, and $\log_2 p = 200$, taking $(t, k) = (4, 2)$ can handle 84-bit perturbations r_i, as one can see in Table 1, but taking $(t, k) = (4, 3)$ cannot even handle 60 bits.

For large m, we experimented with using the nonoptimized parameters $(t, k) = (1, 1)$, as reported in Table 1. For the shortest vector only, the bounds would replace the exponent $\beta^{(m+1)/m}$ with $(m + 1)\beta/m - 1/m$, which is its tangent line at $\beta = 1$. This bound is always worse, and it is trivial when $\beta \le 1/(m + 1)$, but it still approaches the optimal exponent β for large m. Our analysis does not yield a strong enough bound for the m-th largest vector, but in our experiments the vectors found by LLL are much shorter than predicted by the worst-case bounds, as described below. Furthermore, the algorithm runs extremely quickly with these parameters, because the lattices have lower dimensions and the simultaneous equations are all linear.

The last column of the table gives the value of the "LLL factor" λ, which describes the approximation ratio obtained by LLL in the experiment. Specifically, the value of λ satisfies

$$|v_m| \approx \lambda^{\dim L} (\det L)^{1/(\dim L)},$$

where v_m is the m-th smallest vector in the LLL-reduced basis for L. Empirically, we find that all of the vectors in the reduced basis are generally quite close in size, so this estimate is more appropriate than using $1/(\dim L - (m-1))$ in the exponent

m	$\log_2 N$	$\log_2 p$	$\log_2 r$	t	k	$\dim L$	LLL	Gröbner	λ
1	1000	200	36	41	8	42	12.10	—	1.037
1	1000	200	39	190	38	191			
1	1000	400	154	40	16	41	34.60	—	1.023
1	1000	400	156	82	33	83	4554.49	—	1.029
1	1000	400	159	280	112	281			
2	1000	200	72	9	4	55	25.22	0.94	1.030
2	1000	200	85	36	16	703			
2	1000	400	232	10	6	66	126.27	5.95	1.038
2	1000	400	238	15	9	136	15720.95	25.86	1.019
2	1000	400	246	46	29	1128			
3	1000	200	87	5	3	56	18.57	1.20	1.038
3	1000	200	102	14	8	680			
3	1000	400	255	4	3	35	2.86	2.13	1.032
3	1000	400	268	7	5	120	1770.04	25.43	1.040
3	1000	400	281	19	14	1540			
4	1000	200	94	3	2	35	1.35	0.54	1.028
4	1000	200	111	8	5	495			
4	1000	400	279	4	3	70	38.32	9.33	1.035
4	1000	400	293	10	8	1001			
5	1000	200	108	3	2	56	7.35	1.42	1.035
5	1000	200	110	4	3	126	738.57	7.28	1.037
5	1000	400	278	3	2	56	1.86	0.90*	0.743
6	1000	200	115	3	2	84	31.51	3.16	1.038
6	1000	400	297	3	2	84	3.97	1.34*	0.586
7	1000	200	120	3	2	120	203.03	7.73	1.046
7	1000	400	311	3	2	120	12.99	2.23*	0.568
12	1000	400	347	1	1	13	0.01	0.52	1.013
18	1000	400	364	1	1	19	0.03	1.08	1.032
24	1000	400	372	1	1	25	0.04	1.93	1.024
48	1000	400	383	1	1	49	0.28	8.37	1.030
96	1000	400	387	1	1	97	1.71	27.94	1.040

Table 1. Timings, in seconds, of the LLL and Gröbner basis portions of our implementation of the integer partial approximate common divisor algorithm, for various choices of the parameters m, N, p, r, t, and k. Rows for which no timings are listed give sample parameters for more extreme calculations. The meanings of the final column and of the timings marked with an asterisk are explained in the text. We include results for the nonoptimized parameters $t = k = 1$, which perform well for a large number of samples but give a weaker result than Theorem 1.

(which we did in the theoretical analysis, in order to get a rigorous bound). The typical value is about 1.02, which matches the behavior one would expect from LLL on a randomly generated lattice [37], whose successive minima will all be close to $\det L^{1/(\dim L)}$.

Because of this, the reduced lattice bases in practice contain many more than m suitable polynomials, and we were able to speed up some of the Gröbner basis calculations in Table 1 by including all of them in the basis. Even using all the vectors with ℓ_1 norm less than $N^{\beta k}$ is overly conservative in many cases, because vectors that do not satisfy this constraint can still lead to valid relations. Our code initially tries using every vector in the reduced basis except the longest one; if that fails, we fall back on the m shortest vectors. We also experimented with using just those with ℓ_1 norm less than $N^{\beta k}$, but in our experiments this bound was often violated even for polynomials that did vanish. Including more polynomials in the Gröbner basis calculation in many cases leads to substantially better running times than using just m vectors.

A handful of our experimental parameters resulted in lattices whose shortest vectors were much shorter than the expected bounds; this tended to correlate with a small sublattice of algebraically dependent vectors. We marked cases where we encountered algebraically dependent relations with an asterisk in Table 1. In each case, we were still able to solve the system of equations by including more relations from the lattice and solving this larger system.

Acknowledgments

We thank Chris Umans and Alex Vardy for suggesting looking at Parvaresh-Vardy codes, and Martin Albrecht for advice on computing Gröbner bases in practice. N.H. would like to thank MIT, CSAIL, and Microsoft Research New England for their hospitality during the course of this research. This material is based upon work supported by an AT&T Labs Graduate Fellowship and by the National Science Foundation under Award No. DMS-1103803.

References

[1] ACM (ed.), *STOC'*03: *Proceedings of the* 35*th Annual ACM Symposium on Theory of Computing*, New York, ACM Press, 2003. MR 2005j:68009

[2] ACM (ed.), *STOC'*08: *Proceedings of the* 40*th Annual ACM Symposium on Theory of Computing*, New York, ACM Press, 2008. MR 2010m:68004

[3] Daniel Augot and Matthieu Finiasz, *A public key encryption scheme based on the polynomial reconstruction problem*, in Biham [6], 2003, pp. 229–240. MR 2005d:94075

[4] Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger (eds.), *Automata, languages and programming*, Lecture Notes in Computer Science, no. 2719, Berlin, Springer, 2003. MR 2005b:68008

[5] Feng Bao, Robert Deng, and Jianying Zhou (eds.), *Public Key Cryptography—PKC* 2004, Lecture Notes in Computer Science, no. 2947, Berlin, Springer, 2004.

[6] Eli Biham (ed.), *Advances in cryptology—EUROCRYPT* 2003, Lecture Notes in Computer Science, no. 2656, Berlin, Springer, 2003. MR 2005c:94003

[7] Daniel Bleichenbacher, Aggelos Kiayias, and Moti Yung, *Decoding of interleaved Reed Solomon codes over noisy data*, in Baeten et al. [4], 2003, pp. 97–108. MR 2005b:94062

[8] Daniel Bleichenbacher and Phong Q. Nguyen, *Noisy polynomial interpolation and noisy Chinese remaindering*, in Preneel [43], 2000, pp. 53–69. MR 2001b:94030

[9] Dan Boneh and Glenn Durfee, *Cryptanalysis of RSA with private key d less than* $N^{0.292}$, IEEE Trans. Inform. Theory **46** (2000), no. 4, 1339–1349. MR 2002g:94034

[10] Dan Boneh, Glenn Durfee, and Nick Howgrave-Graham, *Factoring* $N = p^r q$ *for large* r, in Wiener [49], 1999, pp. 326–337. MR 2000i:11188

[11] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[12] Bernard Chazelle (ed.), *Innovations in Computer Science—ICS* 2011, Beijing, Tsinghua University Press, 2011.

[13] Yuanmi Chen and Phong Q. Nguyen, *Faster algorithms for approximate common divisors: Breaking fully-homomorphic-encryption challenges over the integers*, in Pointcheval and Johansson [42], 2012, pp. 502–519.

[14] Henry Cohn and Nadia Heninger, *Ideal forms of Coppersmith's theorem and Guruswami-Sudan list decoding*, in Chazelle [12], 2011, pp. 298–308, full version at arXiv:1008.1284 [math.NT].

[15] Don Coppersmith, *Small solutions to polynomial equations, and low exponent RSA vulnerabilities*, J. Cryptology **10** (1997), no. 4, 233–260. MR 99b:94027

[16] Don Coppersmith and Madhu Sudan, *Reconstructing curves in three (and higher) dimensional space from noisy data*, in ACM [1], 2003, pp. 136–142. MR 2005k:94004

[17] Jean-Sébastien Coron, *Cryptanalysis of a public-key encryption scheme based on the polynomial reconstruction problem*, in Bao et al. [5], 2004, pp. 14–27.

[18] Jean-Sébastien Coron, Avradip Mandal, David Naccache, and Mehdi Tibouchi, *Fully homomorphic encryption over the integers with shorter public keys*, in Rogaway [44], 2011, pp. 487–504. MR 2874875

[19] David Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, Undergraduate Texts in Mathematics, Springer, New York, 1992. MR 93j:13031

[20] Casey Devet, Ian Goldberg, and Nadia Heninger, *Optimally robust private information retrieval*, in Kohno [34], 2012, pp. 269–283, extended version at http://eprint.iacr.org/2012/083.

[21] Marten van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan, *Fully homomorphic encryption over the integers*, in Gilbert [23], 2010, pp. 24–43, extended version at http://eprint.iacr.org/2009/616. MR 2660481

[22] Nicolas Gama and Phong Q. Nguyen, *Finding short lattice vectors within Mordell's inequality*, in ACM [2], 2008, pp. 207–216. MR 2011c:68214

[23] Henri Gilbert (ed.), *Advances in cryptology—EUROCRYPT* 2010, Lecture Notes in Computer Science, no. 6110, Berlin, Springer, 2010. MR 2011g:94001

[24] Venkatesan Guruswami and Atri Rudra, *Explicit codes achieving list decoding capacity: error-correction with optimal redundancy*, IEEE Trans. Inform. Theory **54** (2008), no. 1, 135–150. MR 2010b:94096

[25] Venkatesan Guruswami and Madhu Sudan, *Improved decoding of Reed-Solomon and algebraic-geometry codes*, IEEE Trans. Inform. Theory **45** (1999), no. 6, 1757–1767. MR 2000j:94033

[26] Mathias Herrmann and Alexander May, *Solving linear equations modulo divisors*: *on factoring given any bits*, in Pieprzyk [41], 2008, pp. 406–424. MR 2546108

[27] Florian Hess, Sebastian Pauli, and Michael Pohst (eds.), *Algorithmic number theory*, Lecture Notes in Computer Science, no. 4076, Berlin, Springer, 2006. MR 2007h:11001

[28] Nick Howgrave-Graham, *Approximate integer common divisors*, in Silverman [46], 2001, pp. 51–66. MR 2003h:11160

[29] Ellen Jochemsz and Alexander May, *A strategy for finding roots of multivariate polynomials with new applications in attacking RSA variants*, in Lai and Chen [35], 2006, pp. 267–282. MR 2009h:94128

[30] Charanjit S. Jutla, *On finding small solutions of modular multivariate polynomial equations*, in Nyberg [38], 1998, pp. 158–170. MR 2000k:94033

[31] Thomas Kailath, *Linear systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980. MR 82a:93001

[32] Aggelos Kiayias and Moti Yung, *Polynomial reconstruction based cryptography* (*a short survey*), in Vaudenay and Youssef [48], 2001, pp. 129–133. MR 2054433

[33] ———, *Secure games with polynomial expressions*, in Orejas et al. [39], 2001, pp. 939–950. MR 2066563

[34] Tadayoshi Kohno (ed.), 21*st USENIX Security Symposium*, Berkeley, The USENIX Association, 2012.

[35] Xuejia Lai and Kefei Chen (eds.), *Advances in cryptology—ASIACRYPT* 2006, Lecture Notes in Computer Science, no. 4284, Berlin, Springer, 2006. MR 2009e:94091

[36] Alexander May, *New RSA vulnerabilities using lattice reduction methods*, Ph.D. thesis, Universität Paderborn, 2003. http://nbn-resolving.de/urn:nbn:de:hbz:466-20030101205

[37] Phong Q. Nguyen and Damien Stehlé, *LLL on the average*, in Hess et al. [27], 2006, pp. 238–256. MR 2008a:11154

[38] Kaisa Nyberg (ed.), *Advances in cryptology—EUROCRYPT '*98, Lecture Notes in Computer Science, no. 1403, Berlin, Springer, 1998. MR 2000h:94004

[39] Fernando Orejas, Paul G. Spirakis, and Jan van Leeuwen (eds.), *Automata*, *languages and programming*, Lecture Notes in Computer Science, no. 2076, Berlin, Springer, 2001. MR 2005a:68007

[40] Farzad Parvaresh and Alexander Vardy, *Correcting errors beyond the Guruswami-Sudan radius in polynomial time*, Proceedings of the 46th IEEE Symposium on Foundations of Computer Science held in Pittsburgh, October 23–25, 2005 (Los Alamitos, CA), Institute of Electrical and Electronics Engineers, IEEE Computer Society, 2005, pp. 285–294.

[41] Josef Pieprzyk (ed.), *Advances in cryptology—ASIACRYPT* 2008, Lecture Notes in Computer Science, no. 5350, Berlin, Springer, 2008. MR 2010j:94005

[42] David Pointcheval and Thomas Johansson (eds.), *Advances in cryptology—EUROCRYPT* 2012, Lecture Notes in Computer Science, no. 7237, Berlin, Springer, 2012.

[43] Bart Preneel (ed.), *Advances in cryptology—EUROCRYPT* 2000, Lecture Notes in Computer Science, no. 1807, Berlin, Springer, 2000. MR 2001b:94028

[44] Phillip Rogaway (ed.), *Advances in cryptology—CRYPTO* 2011, Lecture Notes in Computer Science, no. 6841, Heidelberg, Springer, 2011. MR 2012i:94009

[45] Santanu Sarkar and Subhamoy Maitra, *Approximate integer common divisor problem relates to implicit factorization*, IEEE Trans. Inform. Theory **57** (2011), no. 6, 4002–4013. MR 2012f:94121

[46] Joseph H. Silverman (ed.), *Cryptography and lattices*, Lecture Notes in Computer Science, no. 2146, Berlin, Springer, 2001. MR 2002m:11002

[47] W. A. Stein et al., *Sage Mathematics Software* (*version* 4.6.2), 2011. http://www.sagemath.org

[48] Serge Vaudenay and Amr M. Youssef (eds.), *Selected areas in cryptography*: *Revised papers from the* 8*th Annual International Workshop* (*SAC* 2001) *held in Toronto*, *ON*, *August* 16–17, 2001, Lecture Notes in Computer Science, no. 2259, Springer, Berlin, 2001. MR 2004k:94066

[49] Michael Wiener (ed.), *Advances in cryptology—CRYPTO '*99, Lecture Notes in Computer Science, no. 1666, Berlin, Springer, 1999. MR 2000h:94003

HENRY COHN: cohn@microsoft.com
Microsoft Research New England, One Memorial Drive, Cambridge, MA 02142, United States

NADIA HENINGER: nadiah@cis.upenn.edu
Department of Computer Science, Princeton University, Princeton, NJ 08540, United States
Current address: *Department of Computer and Information Science, 3330 Walnut St., Philadelphia, PA 19104, United States*

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.295

msp

Explicit descent in the Picard group of a cyclic cover of the projective line

Brendan Creutz

Given a curve X of the form $y^p = h(x)$ over a number field, one can use descents to obtain explicit bounds on the Mordell-Weil rank of the Jacobian or to prove that the curve has no rational points. We show how, having performed such a descent, one can easily obtain additional information which may rule out the existence of rational divisors on X of degree prime to p. This can yield sharper bounds on the Mordell-Weil rank by demonstrating the existence of nontrivial elements in the Shafarevich-Tate group. As an example we compute the Mordell-Weil rank of the Jacobian of a genus 4 curve over $\mathbb{Q}$ by determining that the 3-primary part of the Shafarevich-Tate group is isomorphic to $\mathbb{Z}/3 \times \mathbb{Z}/3$.

1. Introduction

Let k be a global field and J/k an abelian variety. Any separable isogeny $\varphi : J \to J$ gives rise to a short exact sequence of finite abelian groups,

$$0 \longrightarrow J(k)/\varphi(J(k)) \longrightarrow \mathrm{Sel}^{\varphi}(J/k) \longrightarrow Ш(J/k)[\varphi] \longrightarrow 0,$$

relating the finitely generated Mordell-Weil group $J(k)$ and the conjecturally finite Shafarevich-Tate group $Ш(J/k)$. Computation of the middle term, the *φ-Selmer group of J*, is typically referred to as a *φ-descent on J*. This produces an explicit upper bound for the Mordell-Weil rank which will only be sharp when $Ш(J/k)[\varphi]$ is trivial.

While descents on elliptic curves have a history stretching back as far as Fermat, the first examples for abelian varieties of higher dimension appear to have been computed in the 1990s by Gordon and Grant [10], though Cassels had suggested a method using his so-called $(x - T)$ map a decade earlier [6]. These first examples concerned Jacobians of genus 2 curves with rational Weierstrass points.

MSC2010: primary 11G10; secondary 11Y50.
Keywords: abelian variety, Mordell-Weil group, explicit descent.

Schaefer [16; 17] and Poonen and Schaefer [13] later developed a cohomological interpretation of Cassels' $(x-T)$ map which allowed them to generalize the method to Jacobians of all cyclic covers of the projective line. More recently Bruin and Stoll [5] and Mourao [12] have used a similar $(x-T)$ map to do a descent on the cyclic cover itself. This computes a finite set of everywhere locally solvable coverings of the curve which may be of use in determining its set of rational points. In particular, when this set is empty there are no rational points on the curve.

We show how, having performed a descent on the Jacobian J of a cyclic cover X, one can easily obtain additional information which may rule out the existence of k-rational divisors of degree 1 on X. When X is everywhere locally solvable (for instance) the scheme $\mathbf{Pic}^1(X)$, whose k-rational points parametrize k-rational divisor classes of degree 1 on X, represents an element of $Ш(J/k)$. So this can be used to show that $Ш(J/k)$ is nontrivial, and consequently to deduce sharper bounds for the Mordell-Weil rank. We show that this new information can be interpreted as a set parametrizing certain everywhere locally solvable coverings of $\mathbf{Pic}^1(X)$, so one might refer to the method as a *descent on* $\mathbf{Pic}^1(X)$. This interpretation allows us to relate the set in question to the divisibility properties of $\mathbf{Pic}^1(X)$ in $Ш(J/k)$ (see Theorem 4.5 and Corollary 4.6). Well known properties of the Cassels-Tate pairing then allow us to deduce a better lower bound for the size of $Ш(J/k)$ (unconditionally). We give several examples. In one we compute the Mordell-Weil rank of the Jacobian of a genus 4 curve over $\mathbb{Q}$ by determining that the 3-primary part of the Shafarevich-Tate group is isomorphic to $\mathbb{Z}/3\times\mathbb{Z}/3$. We also present empirical data suggesting better bounds are thus obtained rather frequently for hyperelliptic curves.

While one gets additional information on k-rational divisors of degree 1, this is unlikely to be of much additional use for determining the set of rational points on X when the genus is at least 2. When $X(k)\neq\varnothing$, the descent on $\mathbf{Pic}^1(X)$ yields no new information on the Mordell-Weil rank since $\mathbf{Pic}^1(X)\simeq J$. The obstruction to the existence of rational points on X provided by the descent on $\mathbf{Pic}^1(X)$ is weaker than that given by the descent on X, and only provides any new information when the descent on X actually gives an obstruction. That being said, descents on $\mathbf{Pic}^1(X)$ could be useful for computing large generators of the Mordell-Weil group or for finding a k-rational embedding of X into the Jacobian (see [4, Section 3.2] for some examples with genus 2 curves), both of which are relevant for tools such as the Mordell-Weil sieve or Chabauty's method. However, such benefits can only be reaped by constructing explicit models for the coverings parametrized by the descent, which is a topic which we will not address here.

1A. ***Notation.*** Throughout the paper p will be a prime number and k a field of characteristic different from p containing the p-th roots of unity. We use $\bar{k}$ to

denote a separable closure of k and $\mathfrak{g}_k$ to denote the absolute Galois group of k. When k is a global field we denote its completion at a prime v by k_v.

If G is a group, a *principal homogeneous space for* G is a set H on which G acts simply transitively. We make the convention that $\varnothing$ is a principal homogeneous space for any group. Suppose H and H' are principal homogeneous spaces for groups G and G', respectively, and that $i_0 : G \to G'$ is a homomorphism of groups. Then a map $i : H \to H'$ is said to be *affine* (with respect to i_0) if $i(g \cdot h) = i_0(g) \cdot i(h)$ for all $h \in H$ and $g \in G$. An *affine isomorphism* is an affine bijection with respect to an isomorphism of groups. When G is an abelian group and n is an integer, we use $G[n]$ and $G(n)$ to denote the n-torsion subgroup and the subgroup of elements killed by some power of n, respectively.

If L is a k-algebra we use $\bar{L}$ to denote $L \otimes_k \bar{k}$. If V is a projective variety over k and L is a commutative k-algebra we use V_L or $V \otimes_k L$ to denote the extension of scalars, $V \times_{\mathrm{Spec}(k)} \mathrm{Spec}(L)$. The group of k-rational divisors on V is denoted $\mathrm{Div}(V)$. The function field of V is denoted $\kappa(V)$. A divisor is called *principal* if it is the divisor of a function $f \in \kappa(V)$; the group of all such divisors is denoted $\mathrm{Princ}(V)$. The quotient of $\mathrm{Div}(V)$ by $\mathrm{Princ}(V)$ is denoted $\mathrm{Pic}(V)$. When V is a curve $\mathrm{Div}(V)$ is the free abelian group on the set of closed points of V, and there is a well defined notion of degree in $\mathrm{Div}(V)$. For a point $P \in V(\bar{k})$ we use $[P]$ to denote the corresponding element in $\mathrm{Div}(V_{\bar{k}})$. The degree of a principal divisor is 0, so there is also a well defined notion of degree for classes in $\mathrm{Pic}(V)$. We denote the subset consisting of classes of degree i by $\mathrm{Pic}^i(V)$.

Let A be an abelian variety defined over k. A *k-torsor under A* is a variety T over k, together with an algebraic group action of A on T defined over k such that the induced map $A \times T \ni (a, t) \mapsto (a + t, t) \in T \times T$ is an isomorphism. This means that geometrically A acts simply transitively on T. The k-isomorphism classes of k-torsors under A are parametrized by the torsion abelian group $H^1(k, A)$. The trivial class is represented by A acting on itself by translations, and a k-torsor under A is trivial if and only if it possesses a k-rational point. Thus when k is a global field with completions k_v, the Shafarevich-Tate group

$$Ш(A/k) := \ker\Big(H^1(k, A) \to \bigoplus H^1(k_v, A)\Big)$$

parametrizes isomorphism classes of everywhere locally solvable torsors.

We often refer to a variety as a k-torsor under A, taking the group action to be implicit. If T is a k-torsor under A, then any point $t_0 \in T$ gives rise to an isomorphism $T \simeq A$ defined over $k(t_0)$ sending a point $t \in T$ to the unique $a \in A$ such that $a + t_0 = t$. We say an isomorphism $\psi : T \simeq A$ is *compatible with the torsor structure on* T if it is of this type. The action of A on T can be recovered from such an isomorphism by the rule $a + t = \psi^{-1}(\psi(t) + a)$.

2. Coverings and divisibility in Ш

Definition 2.1. Let $\varphi : A' \to A$ be a separable isogeny of abelian varieties. Let T be a k-torsor under A and fix a $\bar{k}$-isomorphism $\psi_T : T \to A$ compatible with the torsor structure. A *φ-covering of T* is a k-variety S together with a morphism $S \xrightarrow{\pi} T$ defined over k such that there exists a $\bar{k}$-isomorphism $\psi_S : S \to A'$ such that $\varphi \circ \psi_S = \psi_T \circ \pi$. Two φ-coverings of T are k-isomorphic if they are k-isomorphic as T-schemes. We use $\mathrm{Cov}^{\varphi}(T/k)$ to denote the set of isomorphism classes of φ-coverings of T. If k is a global field we define the *φ-Selmer set of T* to be the subset $\mathrm{Sel}^{\varphi}(T/k) \subset \mathrm{Cov}^{\varphi}(T/k)$ consisting of those φ-coverings which are everywhere locally solvable.

We will see below that this definition generalizes the usual definition of the φ-Selmer group of an abelian variety. The definition does not depend on the choice for ψ_T, and the isomorphism ψ_S endows S with the structure of a k-torsor under A'.

Lemma 2.2. *Let (S, π) be a φ-covering of T. Then its group of $\bar{k}$-automorphisms is isomorphic to $A'[\varphi]$ as a Galois module.*

Proof. Suppose $\psi : S \to S$ is an isomorphism such that $\pi = \pi \circ \psi$ and consider the endomorphism $\tau = \psi_S \circ \psi \circ \psi_S^{-1} - 1 \in \mathrm{End}(A')$. Since $\pi = \varphi \circ \psi_S = \varphi \circ \psi_S \circ \psi$ we have that $\varphi \circ \tau$ is identically 0. Then τ is a continuous map from $A'(\bar{k})$, which is irreducible, to $A'[\varphi]$, which is discrete. Hence τ is constant. It follows that ψ is translation by a φ-torsion point. Conversely it is clear that translation by any φ-torsion point gives an automorphism of (S, π). □

By definition all φ-coverings of T are twists of one another. So by the twisting principle $\mathrm{Cov}^{\varphi}(T/k)$ is a principal homogeneous space for the group $H^1(k, A'[\varphi])$. In the special case $T = A$ (acting on itself by translations), the morphism $\varphi : A' \to A$ gives A' a canonical structure as a φ-covering of A. This gives a canonical identification of $H^1(k, A'[\varphi])$ and $\mathrm{Cov}^{\varphi}(A/k)$ and consequently endows $\mathrm{Cov}^{\varphi}(A/k)$ with a group structure in which $\varphi : A' \to A$ represents the identity. Under this identification the isomorphism classes of φ-coverings of A which possess k-rational points correspond to the kernel in the Kummer sequence

$$0 \longrightarrow A(k)/\varphi(A'(k)) \longrightarrow H^1(k, A'[\varphi]) \longrightarrow H^1(k, A')[\varphi] \longrightarrow 0. \quad (2\text{-}1)$$

When k is a global field one can deduce from this that $\mathrm{Sel}^{\varphi}(A/k)$ is identified with the kernel of the natural map $H^1(k, A'[\varphi]) \to \bigoplus_v H^1(k_v, A)$. In particular it is a subgroup and it sits in an exact sequence

$$0 \longrightarrow A(k)/\varphi(A'(k)) \longrightarrow \mathrm{Sel}^{\varphi}(A/k) \longrightarrow Ш(A'/k)[\varphi] \longrightarrow 0. \quad (2\text{-}2)$$

Remark. The reader is cautioned that our notation is nonstandard. Our $\mathrm{Sel}^{\varphi}(A/k)$ would typically be referred to as the φ-Selmer group of A' (with A' present in the notation).

More generally φ-Selmer sets are related to divisibility in the Shafarevich-Tate group as follows.

Proposition 2.3. *Suppose $\varphi : A' \to A$ is a separable isogeny of abelian varieties over k and that T is a k-torsor under A. Then $\mathrm{Cov}^{\varphi}(T/k) \neq \varnothing$ if and only if $T \in \varphi H^1(k, A')$. If k is a global field, then $\mathrm{Sel}^{\varphi}(T/k) \neq \varnothing$ if and only if $T \in \varphi Ш(A'/k)$.*

Proof. We will prove the second statement. The first can be proved using the same argument. We may assume $T \in Ш(A/k)$, otherwise the statement is trivial. Suppose T is killed by m and consider the following commutative and exact diagram:

$$\begin{array}{ccccc} \mathrm{Sel}^{m\circ\varphi}(A/k) & \longrightarrow & Ш(A'/k)[m\circ\varphi] & \longrightarrow & 0 \\ \downarrow{\scriptstyle \varphi_*} & & \downarrow{\scriptstyle \varphi} & & \\ \mathrm{Sel}^{m}(A/k) & \longrightarrow & Ш(A/k)[m] & \longrightarrow & 0. \end{array}$$

The torsor T admits a lift to an m-covering $T \xrightarrow{\pi} A$ in the m-Selmer group of A. Each choice of lift gives a map

$$\begin{aligned} \mathrm{Sel}^{\varphi}(T/k) &\longrightarrow \mathrm{Sel}^{(m\circ\varphi)}(A/k) \\ (S, \rho) &\longmapsto (S, \pi \circ \rho). \end{aligned}$$

The image of this map is exactly the fiber above (T, π) under the map denoted φ_* in the diagram above. From this one deduces the result from commutativity and the fact that the horizontal maps are surjective. □

We record here the following well known lemma which relates the condition in Proposition 2.3 to the Cassels-Tate pairing.

Lemma 2.4. *Let $\varphi : A' \to A$ be a separable isogeny of abelian varieties over a global field k with dual isogeny $\varphi^{\vee} : A^{\vee} \to A'^{\vee}$. An element of $Ш(A/k)$ is divisible by φ if and only if it pairs trivially with every element of $Ш(A^{\vee}/k)[\varphi^{\vee}]$ under the Cassels-Tate pairing.*

Proof. The compatibility of the Cassels-Tate pairing with isogenies (see [11, Remark I.6.10(a)]) shows that it induces a complex

$$\varphi Ш(A') \longrightarrow Ш(A) \longrightarrow \mathrm{Hom}\big(Ш(A^{\vee})[\varphi^{\vee}], \mathbb{Q}/\mathbb{Z}\big).$$

The statement is equivalent to claiming that this is exact. When φ is multiplication by an integer this result appears in the paragraph following the proof of [11, Lemma I.6.17]. The general statement can be deduced in exactly the same manner. □

3. Cyclic covers of $\mathbb{P}^1$

Let $\pi : X \to \mathbb{P}^1$ be a cyclic cover of degree p defined over k. By the Riemann-Hurwitz formula, X has genus $g = (d-2)(p-1)/2$, where d is the number of branch points of π. Provided $\mathbb{P}^1(k)$ has sufficiently many points we can make a change of variables to ensure that π is not ramified above $\infty \in \mathbb{P}^1$. As our present interest lies in infinite fields, there is no harm in assuming this to be the case. The pullback $\mathfrak{m} = \pi^*\infty$ is an effective k-rational divisor of degree p on X. Let $\Omega \subset X$ denote the set of ramification points of π. Then for any $\omega \in \Omega$ the divisor $p[\omega]$ is linearly equivalent to $\mathfrak{m}$, and $(2g-2)[\omega]$ is a canonical divisor.

3A. ***The isogeny ϕ.*** Since k contains the p-th roots of unity, the group of deck transformations of π may be identified with $\mu_p(\bar{k})$. The action of $\mu_p(\bar{k})$ on X extends linearly to give a Galois-equivariant action of the group ring $\mathbb{Z}[\mu_p]$ on $\mathrm{Div}(X_{\bar{k}})$. For any divisor D, the element $t = \sum_{\zeta \in \mu_p} \zeta \in \mathbb{Z}[\mu_p]$ sends D to a divisor linearly equivalent to $(\deg D)\mathfrak{m}$. Hence t sends $\mathrm{Div}^0(X_{\bar{k}})$ to $\mathrm{Princ}(X_{\bar{k}})$, so the induced actions of $\mathbb{Z}[\mu_p]$ on J and $\mathrm{Pic}^0(X)$ factor through $\mathbb{Z}[\mu_p]/t$, which is isomorphic to the cyclotomic subring of k generated by μ_p. Fix a generator $\zeta \in \mu_p$ and set $\phi = 1 - \zeta$. Then $\phi : J \to J$ is an isogeny of degree p^{d-2}. We note that the ratio of ϕ^{p-1} and p is a unit in $\mathrm{End}(J)$.

3B. ***The model $y^p = ch(x)$.*** By Kummer theory, X has a (possibly singular) affine model of the form $y^p = ch(x)$, where $c \in k^\times$ and $h(x) \in k[x]$ is a p-th-power-free polynomial with leading coefficient 1. In this model π is given by the x-coordinate and $\zeta \in \mu_p(\bar{k})$ acts via $(x, y) \mapsto (x, \zeta y)$. Our assumption that ∞ is not a branch point implies that the branch points are the roots of $h(x)$ and so we may assume p divides the degree of $h(x)$.

3C. ***The torsor $\mathscr{X}$.*** In what follows we consider the reduced scheme $\mathscr{X} = \mathbf{Pic}^1(X)$ classifying linear equivalence classes of divisors of degree 1 on X. This scheme is defined over k and its set of $\bar{k}$-points is $\mathscr{X}(\bar{k}) = \mathrm{Pic}^1(X_{\bar{k}})$. The obvious injection $\mathrm{Pic}^1(X) \to \mathrm{Pic}^1(X_{\bar{k}})^{\mathfrak{g}_k} = \mathscr{X}(k)$ is not always surjective. The obstruction to a k-rational divisor class being represented by a k-rational divisor can be interpreted as an element of the Brauer group; one has a well known exact sequence (see, for example, [2, Section 9.1])

$$0 \longrightarrow \mathrm{Pic}^1(X) \longrightarrow \mathscr{X}(k) \xrightarrow{\theta_X} \mathrm{Br}(k). \tag{3-1}$$

The obstruction θ_X vanishes identically when $\mathrm{Pic}^1(X)$ is nonempty. When k is a global field, the local-global principle for $\mathrm{Br}(k)$ can be used to show that $\mathrm{Pic}^1(X) = \mathscr{X}(k)$ if $\mathrm{Pic}^1(X_{k_v}) = \varnothing$ for at most one prime. Similarly if $\mathscr{X}(k_v) = \varnothing$ for at most one prime v, then $\mathrm{Pic}^0(X) = \mathrm{Pic}^0(X_{\bar{k}})^{\mathfrak{g}_k}$, which is equal to $J(k)$.

There is a $\bar{k}$-isomorphism $\mathscr{X} \simeq J$, sending a point $P \in \mathscr{X}$ corresponding to the divisor class of D to the class of the divisor $D - [\omega_0]$ in $\mathrm{Pic}^0(X_{\bar{k}}) = J(\bar{k})$. This endows $\mathscr{X}$ with the structure of a k-torsor under J (which does not depend on the choice for ω_0). The class of $\mathscr{X}$ in $H^1(k, J)$ is given by the class of the 1-cocycle sending $\sigma \in \mathfrak{g}_k$ to the class of $[\omega_0] - [\omega_0^\sigma]$ in $\mathrm{Pic}^0(X_{\bar{k}}) = J(\bar{k})$. As the difference of any two ramification points gives a ϕ-torsion point on J, we see that the class of $\mathscr{X}$ in $H^1(k, J)$ is killed by ϕ. In particular, this class has order p if and only if $\mathscr{X}(k) = \varnothing$. This is the case if and only if every k-rational divisor class on X has degree divisible by p.

4. The algebraic Selmer set

4A. ***The* $(x - T, y)$ *map.*** Let $H(x, z)$ be the binary form of degree $n = \deg(h(x))$ such that $H(x, 1) = h(x)$. Then X is birational to the curve $y^p = cH(x, z)$ in the weighted projective plane $\mathbb{P}^2(x : y : z)$ with weights $1, n/p, 1$. Writing $H(x, z)$ as $H(x, z) = H_1(x, z)^{n_1} \cdots H_e(x, z)^{n_e}$ with distinct irreducible factors $H_i(x, z)$, the radical of $H(x, z)$ is $H_{\mathrm{rad}}(x, z) = H_1(x, z) \cdots H_e(x, z)$. Let $L = \mathrm{Map}_k(\Omega, \bar{k}) \simeq k[x]/H_{\mathrm{rad}}(x, 1)$. This is the étale k-algebra associated to the finite $\mathfrak{g}_k$-set Ω. It splits as a product $L \simeq K_1 \times \cdots \times K_e$ of finite extensions of k corresponding to the irreducible factors $h_i(x)$ of $h(x)$. We have a *weighted norm map*

$$N : L \simeq K_1 \times \cdots \times K_e \to k, \quad (\alpha_1, \ldots, \alpha_e) \mapsto \prod_{i=1}^{e} N_{K_i/k}(\alpha_i)^{n_i}. \tag{4-1}$$

Let

$$\Omega' = \big\{ p[\omega] : \omega \in \Omega \big\} \cup \Big\{ \sum_{\omega \in \Omega} n_\omega [\omega] \Big\} \subset \mathrm{Div}(X_{\bar{k}}).$$

The first set appearing in the union above is isomorphic to Ω as a $\mathfrak{g}_k$-set. The divisor $\sum_{\omega \in \Omega} n_\omega [\omega]$ is the zero divisor of the function $y/z^{n/p} \in \kappa(X)^\times$. In particular it is invariant under the action of $\mathfrak{g}_k$. Thus Ω' is a disjoint union of $\mathfrak{g}_k$-sets, and the étale k-algebra corresponding to Ω' splits as $\mathrm{Map}_k(\Omega', \bar{k}) = M \simeq L \times k$. Since the action of $\mathfrak{g}_k$ on Ω' is induced from the action on Ω, we have an induced norm map

$$\partial : L = \mathrm{Map}_k(\Omega, \bar{k}) \to \mathrm{Map}_k(\Omega', \bar{k}) = M, \quad \alpha \mapsto \big(\omega' = \textstyle\sum c_\omega [\omega] \mapsto \prod \alpha(\omega)^{c_\omega}\big).$$

Concretely, this is the map

$$\alpha \mapsto (\alpha^p, N(\alpha)) \in L \times k, \tag{4-2}$$

where N is the weighted norm map defined in (4-1). We can embed k in M via the map $\iota : k \to M \simeq L \times k$ sending a to $(a, a^{n/p})$. The choice is such that $\partial(a) = \iota(a^p)$.

Let $f \in \mathrm{Map}_k(\Omega', \kappa(X_{\bar{k}})^\times)$ be the map

$$\omega' \longmapsto \begin{cases} (x - x(\omega)z)/z & \text{if } \omega' = p[\omega], \\ y/z^{n/p} & \text{if } \omega' = \sum_{\omega \in \Omega} n_\omega [\omega]. \end{cases}$$

Then f is a Galois-equivariant family of functions f_ω parametrized by Ω', whose divisors are supported on the union of Ω and the support of $\mathfrak{m}$. Moreover, if $[\boldsymbol{w}] \in \mathrm{Map}_k(\Omega, \mathrm{Div}(X_{\bar{k}}))$ denotes the map $(\omega \mapsto [\omega])$ and we interpret $\iota(\mathfrak{m})$ as the element

$$\omega' \longmapsto \begin{cases} \mathfrak{m} & \text{if } \omega' = p[\omega], \\ (n/p)\mathfrak{m} & \text{if } \omega' = \sum_{\omega \in \Omega} n_\omega [\omega] \end{cases}$$

of $\mathrm{Map}_k(\Omega', \mathrm{Div}(X_{\bar{k}}))$, then the family of divisors corresponding to f is

$$\mathrm{div}(f) = \partial[\boldsymbol{w}] - \iota(\mathfrak{m}) \in \mathrm{Map}_k\big(\Omega', \mathrm{Div}(X_{\bar{k}})\big).$$

Following the terminology in [13] we will say a divisor is *good* if its support is disjoint from Ω and $\mathfrak{m}$. For any good divisor, $D = \sum_P n_P [P] \in \mathrm{Div}(X_{\bar{k}})$, we may define

$$f(D) = \prod_P f(P)^{n_P} \in \overline{M}^\times.$$

Note that if $D \in \mathrm{Div}(X)$, then $f(D) \in M^\times$. Every k-rational divisor is linearly equivalent to a good k-rational divisor. Using this and applying Weil reciprocity one can prove the following proposition. For details we refer the reader to [7, Proposition 3.1], [13, Section 5] or [21, Section 4].

Proposition 4.1. *The function f induces a unique homomorphism*

$$f : \mathrm{Pic}(X) \to M^\times / \iota(k^\times)\partial(L^\times)$$

with the property that the image of the class of any good divisor $D \in \mathrm{Div}(X)$ is given by $f(D)$ as defined above.

Remark. The $(x - T, y)$ map of Stoll and van Luijk defined in [21] differs from ours slightly. The second factor of their map is defined using the function $\gamma y / z^{n/p}$ where γ is some p-th root of c. Hence their map and ours agree in degree 0 only. The projection $\mathrm{pr}_1 : M \simeq L \times k \to L$ induces a map $M^\times / \iota(k^\times)\partial(L^\times) \to L^\times / k^\times L^{\times p}$. Composing this with either f or the $(x - T, y)$ map defined in [21] one recovers the $(x - T)$ map defined in [13]. The main advantage of our definition over the others is that it defines a homomorphism on all of $\mathrm{Pic}(X)$ and not just the degree divisible by p part. The map used in [5; 12] to do a descent on X is the restriction of $\mathrm{pr}_1 \circ f$ to $X(k) \subset \mathrm{Pic}^1(X)$.

Recall that $c \in k^\times$ is the leading coefficient of the polynomial defining X. For $r \in \mathbb{Z}$ define

$$\boldsymbol{H}_k^r = \frac{\{(\alpha, s) \in L^\times \times k^\times : c^r \cdot N(\alpha) = s^p\}}{\{(\gamma\alpha^p, \gamma^{n/p} N(\alpha)) \in L^\times \times k^\times : \alpha \in L^\times, \gamma \in k^\times\}} \subset M^\times/\iota(k^\times)\partial(L^\times).$$

Lemma 4.2. *For $r \in \mathbb{Z}$,*

$$\boldsymbol{H}_k^r = \big(f(D)\partial(\bar{L}^\times)\big)^{\mathfrak{g}_k}\big/\iota(k^\times)\partial(L^\times),$$

where $D \in \mathrm{Pic}^r(X_{\bar{k}})$ is any divisor class of degree r. In particular,

$$\boldsymbol{H}_k^0 = \big(\partial(\bar{L}^\times)\big)^{\mathfrak{g}_k}\big/\iota(k^\times)\partial(L^\times).$$

Proof. First we claim that

$$\partial(\bar{L}^\times) = \big\{(\alpha, s) \in \bar{L}^\times \times \bar{k}^\times : N(\alpha) = s^p\big\}.$$

By definition $\partial(\bar{L}^\times) = \{(\alpha^p, N(\alpha)) : \alpha \in \bar{L}^\times\}$, so clearly

$$\partial(\bar{L}^\times) \subset \big\{(\alpha, s) \in \bar{L}^\times \times \bar{k}^\times : N(\alpha) = s^p\big\}.$$

For the other inclusion, suppose $(a, s) \in \bar{L}^\times \times \bar{k}^\times$ is such that $N(\alpha) = s^p$. Then for any p-th root $\beta \in \bar{L}^\times$ of α we have $N(\beta)^p = s^p$. Hence $N(\beta) = \nu s$ for some $\nu \in \mu_p(\bar{k})$. Since $h(x)$ is p-th-power-free, the weighted norm map $N : \mu_p(\bar{L}) \to \mu_p(\bar{k})$ is surjective. Hence there must exist $\nu' \in \mu_p(\bar{L})$ such that $\nu'\beta \in \bar{L}^\times$ satisfies $\partial\beta = ((\nu'\beta)^p, N(\nu'\beta)) = (\alpha, s)$. This establishes the claim.

For $i = 1, 2$, let pr_i denote the projection of $M \simeq L \times k$ onto the i-th factor. For every point $P = (x_0, y_0) \in X$ we have

$$cN(\mathrm{pr}_1 \circ f(P)) = c \prod_{\omega \in \Omega} (x_0 - x(\omega))^{n_\omega} = ch(x_0) = y_0^p = \mathrm{pr}_2(f(P))^p,$$

where n_ω denotes the multiplicity of ω as a root of $h(x)$. So for any good divisor D of degree r we have $c^r N(\mathrm{pr}_1 \circ f(D)) = \mathrm{pr}_2(f(D))^p$, and, in light of the claim above, we have

$$f(D)\partial(\bar{L}^\times) = \big\{(\alpha, s) \in \bar{L}^\times \times \bar{k}^\times : c^r \cdot N(\alpha) = s^p\big\}.$$

In particular, the coset $f(D)\partial(\bar{L}^\times)$ depends only on the degree of D. The same is then true of its Galois-invariant subset. The lemma now follows easily. □

Corollary 4.3. *If $\boldsymbol{H}_k^1 = \varnothing$, then $\mathrm{Pic}^1(X) = \varnothing$.*

Proof. The image of $f : \mathrm{Pic}^r(X) \to M^\times/\iota(k^\times)\partial(L^\times)$ is contained in $\boldsymbol{H}_k^r$. □

4B. ***The algebraic Selmer set.*** Over a global field one can combine the information from the various local versions of the map f to obtain a finite subset of $\boldsymbol{H}_k^r$ which contains the image of $\mathrm{Pic}^r(X)$.

Definition 4.4. For a global field k with completions k_v, define *algebraic ϕ-Selmer sets*:

$$\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k) = \left\{\delta \in \boldsymbol{H}_k^0 : \text{ for all primes } v,\ \mathrm{res}_v(\delta) \in f(\mathrm{Pic}^0(X_{k_v}))\right\},$$
$$\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k) = \left\{\delta \in \boldsymbol{H}_k^1 : \text{ for all primes } v,\ \mathrm{res}_v(\delta) \in f(X(k_v))\right\},$$
$$\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k) = \left\{\delta \in \boldsymbol{H}_k^1 : \text{ for all primes } v,\ \mathrm{res}_v(\delta) \in f(\mathrm{Pic}^1(X_{k_v}))\right\}.$$

Recall that the projection $\mathrm{pr}_1 : M \simeq L \times k \to L$ induces a map

$$\mathrm{pr}_1 : M^\times/\iota(k^\times)\partial(L^\times) \to L^\times/k^\times L^{\times p}.$$

The *fake ϕ-Selmer group* considered in [13] is equal to $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k))$. The *unfaked ϕ-Selmer group* considered in [21] is equal to $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k)$. From [21, Theorem 5.1] we see that if X has divisors of degree 1 everywhere locally, then we can identify $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k)$ with the ϕ-Selmer group of J. In particular, $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k)$ is finite. If the set $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ is nonempty, it is a coset of $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k)$ inside $M^\times/\iota(k^\times)\partial(L^\times)$. This implies that $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ is also finite. If, in addition, $\delta \in \mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k) \neq \varnothing$, then $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k) = \delta \cdot \mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k)$, and, in particular, $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k) \subset \mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$. The set $\mathrm{pr}_1\left(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k)\right)$ is equal to the *fake ϕ-Selmer set* considered in [5; 12] where it is shown to be a quotient of the ϕ-Selmer set of X (see Definition 5.1). As we shall see in Corollary 5.5, $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k)$ is in one-to-one correspondence with ϕ-Selmer set of X.

One motivation for considering this set is that it can explain the failure of the Hasse principle for X. Similarly, one can easily deduce the implication

$$\left(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k) = \varnothing\right) \implies \left(\mathrm{Pic}^1(X) = \varnothing\right).$$

When X has points everywhere locally we can say even more.

Theorem 4.5. *Suppose k is a global field and X is everywhere locally solvable. Then $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ is nonempty if and only if the torsor $\mathscr{X}$ is divisible by ϕ in $Ш(J/k)$.*

In light of Proposition 2.3, to prove the theorem it will suffice to show that when X is everywhere locally solvable $\mathrm{Sel}^{\phi}(\mathscr{X}/k)$ and $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ are in one-to-one correspondence. This will be accomplished with Proposition 6.2 below.

Corollary 4.6. *Suppose k is a global field and X is everywhere locally solvable. If $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ is empty, then $\dim_{\mathbb{F}_p} Ш(J/k)[\phi] \geq 2$. If in addition $\dim_{\mathbb{F}_p} Ш(J/k)[\phi] \leq 2$, then $Ш(J/k)(p) \simeq \mathbb{Z}/p \times \mathbb{Z}/p$.*

Proof. Under the assumptions the theorem implies that $\mathscr{X}$ represents a nontrivial class in the finite abelian group $G = Ш(J/k)[\phi]\big/\phi Ш(J/k)[\phi^2]$. Under the canonical identification of J with its dual, ϕ is self dual (up to a unit). It then follows from Lemma 2.4 and [14, Corollary 12] that the Cassels-Tate pairing induces a nondegenerate alternating pairing on G. Hence the order of G is a positive even power of p. This establishes the first statement. For the second, note that the assumptions imply that $\phi Ш(J/k)[\phi^2] = 0$, and use that $\phi^{p-1} = p$ up to a unit. □

Remark. To show G has square order it is enough to assume that p is odd or that X has a k_v-rational divisor of degree 1 for each prime v. We use the assumption that X is everywhere locally solvable to ensure that $\mathscr{X}$ represents a nontrivial element of G. Indeed this assumption is used in our proof of Theorem 4.5 when we apply Lemma 6.1 in the proof of Proposition 6.2. While it may be possible to relax this hypothesis, some assumption on the existence of k_v-rational divisors of degree 1 is required. For the curve $X : y^2 = 3x^6 + 3$, one can show that $\mathrm{Pic}^1(X_{\mathbb{Q}_2}) = \varnothing$, while $\mathscr{X}(\mathbb{Q}) \neq \varnothing$. So the algebraic Selmer set is empty, but $\mathscr{X} \in 2Ш(J/\mathbb{Q})$.

Remark. It is not generally true that $Ш(J/k)[\phi]$ has square order. Well known examples with $p = 2$ are given in [13] and are necessarily explained by the fact that X fails to have a k_v-rational divisor of degree 1 at an odd number of primes. An example with $p = 3$ where X has a rational point is given in [8].

4C. ***Computing the algebraic Selmer set.*** Before carrying on with the proof of Theorem 4.5 we briefly discuss how $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ can be computed in practice. For an extension K/k set $\mathfrak{L}(K) = (L \otimes_k K)^\times / K^\times (L \otimes_k K)^{\times p}$, and use res_K to denote the canonical map $\mathfrak{L}(k) \to \mathfrak{L}(K)$. The weighted norm $N : L \to k$ induces a map $N : \mathfrak{L}(k) \to k^\times / k^{\times p}$. If k is a local field, an element of $\mathfrak{L}(k)$ is said to be *unramified* if its image under res_{k^u} is trivial, where k^u denotes the maximal unramified extension of k. If k is a global field, an element $\delta \in \mathfrak{L}(k)$ is said to be *unramified at a prime* v of k if $\mathrm{res}_{k_v}(\delta)$ is unramified.

Now suppose k is a global field and let S denote the set of primes of k consisting of all primes of bad reduction, all nonarchimedean primes dividing cp and all archimedean primes.[1] Let $\mathfrak{L}(k)_S$ denote the subgroup of $\mathfrak{L}(k)$ consisting of elements which are unramified at all primes outside of S. This is a finite group which can be computed from the S-unit group and class group of each of the constituent fields of L (see Propositions 12.5 and 12.6, Corollary 12.7, and Proposition 12.8 of [13]). For an element $a \in k^\times$, let $\mathfrak{L}(k)_{S,a}$ denote the subset of $\mathfrak{L}(k)_S$ consisting of elements α such that $aN(\alpha) \in k^{\times p}$.

[1] Actually, one can get away with using a smaller set of primes. Compare with [20, Corollary 4.7 and Proposition 5.12], [5, Lemma 4.3], and [12, Lemma 2.6].

Computable descriptions of $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k))$ and $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k))$ are given in [13, Theorem 13.2], [5, Section 6] and [12, Corollary 3.12]. They are the subsets of $\text{Ш}(k)_{S,1}$ and $\text{Ш}(k)_{S,c}$ cut out by certain local conditions. The former is the subgroup of elements which restrict into $\mathrm{pr}_1 \circ f(\mathrm{Pic}^0(X_{k_v}))$ for all $v \in S$ while the latter is the subset which restricts into $\mathrm{pr}_1 \circ f(X(k_v))$ for all primes with norm up to some explicit bound. For explicit descriptions of how to compute these local images, see [5; 12; 20].

Proposition 4.7. *Suppose that $D_v \in X(k_v)$ for each $v \in S$. Then*

$$\mathrm{pr}_1\big(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)\big) = \big\{\delta \in \text{Ш}(k)_{S,c} : \mathrm{res}_{k_v}(\delta) \in \mathrm{pr}_1(f(D_v)) \cdot \mathrm{pr}_1\big(f(\mathrm{Pic}^0(X_{k_v}))\big) \text{ for all } v \in S\big\}.$$

Proof. This follows from the descriptions of $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k))$ and $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k))$ above and the fact that $\mathrm{pr}_1 \circ f$ is a homomorphism. □

Remark. This shows that while doing a ϕ-descent on J — that is, computing $\mathrm{pr}_1(\mathrm{Sel}^{\phi}_{\mathrm{alg}}(J/k))$ — one can determine whether $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ is empty or not with virtually no extra effort.

5. ϕ-coverings of X

Our proof of Theorem 4.5 will involve relating $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathscr{X}/k)$ and $\mathrm{Sel}^{\phi}(\mathscr{X}/k)$. To do this we first relate $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k)$ to a certain set of coverings of X which we now define.

Definition 5.1. A *ϕ-covering of X* is a covering $Y \to X$ which arises as the pullback of some ϕ-covering $\mathscr{Y} \to \mathscr{X}$ along the canonical map $X \to \mathscr{X}$ sending a point P to the class of the divisor $[P]$. We use $\mathrm{Cov}^{\phi}(X/k)$ to denote the set of k-isomorphism classes of ϕ-coverings of X. If k is a global field, the *ϕ-Selmer set of X* is defined to be the subset $\mathrm{Sel}^{\phi}(X/k) \subset \mathrm{Cov}^{\phi}(X/k)$ consisting of those coverings that are everywhere locally solvable.

It follows that any ϕ-covering of X is an X-torsor under $J[\phi]$ and that all ϕ-coverings of X are twists of one another. Hence $\mathrm{Cov}^{\phi}(X/k)$ is also a principal homogeneous space for $H^1(k, J[\phi])$. The action of twisting is compatible with base change, so the obvious map $\mathrm{Cov}^{\phi}(\mathscr{X}/k) \to \mathrm{Cov}^{\phi}(X/k)$ is an affine isomorphism.

Our next goal is to relate $\boldsymbol{H}^1_k$ with a certain subset of $\mathrm{Cov}^{\phi}(X/k)$ and use this to show that $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k)$ and $\mathrm{Sel}^{\phi}(X/k)$ are in one-to-one correspondence. While we work with $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(X/k)$ rather than its image under pr_1, this result was essentially established in [5; 12]. The only new ingredient here is to clarify the affine structure of these sets. This interpretation is, however, crucial to our proof of Theorem 4.5.

We have an exact sequence

$$1 \longrightarrow \mu_p \longrightarrow J_{\mathfrak{m}}[\phi] \xrightarrow{q} J[\phi] \longrightarrow 0, \tag{5-1}$$

where $J_{\mathfrak{m}}$ is the generalized Jacobian associated to the modulus $\mathfrak{m} \in \mathrm{Div}(X)$ (see [13, Section 2] or [19, Chapter 5]). Applying Galois cohomology gives an exact sequence

$$H^1(k, \mu_p) \longrightarrow H^1(k, J_{\mathfrak{m}}[\phi]) \longrightarrow H^1(k, J[\phi]) \xrightarrow{\Upsilon} H^2(k, \mu_p). \tag{5-2}$$

The description of $J_{\mathfrak{m}}[\phi]$ in [13, Section 6] identifies $J_{\mathfrak{m}}[\phi]$ with the kernel of $\partial : \bar{L}^\times \to \bar{M}^\times$. This allows us to interpret the cocycle in the following proposition as taking values in $J[\phi]$.

Proposition 5.2. *There is an isomorphism $\boldsymbol{H}_k^0 \simeq \ker \Upsilon$ which sends the class of $\partial(\alpha) \in \partial(\bar{L}^\times)^{\mathfrak{g}_k}$ to the class of the 1-cocycle $\sigma \mapsto q(\sigma(\alpha)/\alpha)$ in $H^1(k, J[\phi])$.*

Proof. This can be found in [21] (see Proposition 3.1 and Remark 4.3). □

Definition 5.3. Define

$$\mathrm{Cov}_0^\phi(X/k) = \left\{ (Y, \pi) \in \mathrm{Cov}^\phi(X/k) : \begin{array}{l} \pi^*[\omega_0] \text{ is linearly equivalent} \\ \text{to a } k\text{-rational divisor} \end{array} \right\}.$$

The pullbacks of the ramification points are all linearly equivalent, so $\pi^*[\omega_0]$ represents a k-rational divisor class. If k is a global field and Y is everywhere locally solvable, then every k-rational divisor class contains a k-rational divisor. Thus we see that $\mathrm{Sel}^\phi(X/k) \subset \mathrm{Cov}_0^\phi(X/k)$.

Proposition 5.4. *The action of $H^1(k, J[\phi])$ on $\mathrm{Cov}^\phi(X/k)$ restricts to a simply transitive action of $\ker(\Upsilon) \simeq \boldsymbol{H}_k^0$ on $\mathrm{Cov}_0^\phi(X/k)$. The function f induces an affine isomorphism*

$$\mathfrak{f} : \mathrm{Cov}_0^\phi(X/k) \to \boldsymbol{H}_k^1$$

with the property that for any $(Y, \pi) \in \mathrm{Cov}_0^\phi(X/k)$ and any extension K/k, we have

$$f(\pi(Q)) = \mathfrak{f}((Y, \pi)) \text{ in } \boldsymbol{H}_K^1$$

for every point $Q \in Y(K)$.

Corollary 5.5. *Suppose k is a global field. Then $\mathfrak{f}$ restricts to give a bijection $\mathfrak{f} : \mathrm{Sel}^\phi(X/k) \to \mathrm{Sel}_{\mathrm{alg}}^\phi(X/k)$.*

Proof of Proposition 5.4. Let $(Y, \pi) \in \mathrm{Cov}_0^\phi(X/k)$. The complete linear system associated to $\pi^*[\omega_0]$ gives an embedding in $\mathbb{P}^N$ (for some N) with the property that for $\omega \in \Omega$, the divisor $\pi^*[\omega]$ is a hyperplane section defined by the vanishing of some linear form l_ω. Recall that $[\boldsymbol{w}]$ is the map $(\omega \mapsto [\omega]) \in \mathrm{Map}_K(\Omega, \mathrm{Div}(X_{\bar{k}}))$.

These linear forms l_ω may be chosen so as to give a linear form l with coefficients in L defining the $\mathfrak{g}_k$-equivariant family of divisors

$$\big(\pi^*[\boldsymbol{w}] : \omega \mapsto \pi^*[\omega]\big) \in \mathrm{Map}_K(\Omega, \mathrm{Div}(Y_{\bar{k}})).$$

Since the divisor of f is $\partial[\boldsymbol{w}] - \iota(\mathfrak{m}) \in \mathrm{Map}_k(\Omega', \mathrm{Div}(X_{\bar{k}}))$, we see that there is some $\Delta \in M^\times$ such that

$$\pi^* f = \Delta \frac{\partial(l)}{\iota(z \circ \pi)} \in \mathrm{Map}_k(\Omega', \kappa(Y_{\bar{k}})^\times).$$

Define $\mathfrak{f}((Y, \pi)) = \Delta$. A different choice of model for Y or a different choice for the linear form l would serve to modify Δ by an element of $\iota(k^\times)\partial(L^\times)$. So the class of Δ in $\boldsymbol{H}^1_k$ is well defined. For any point $Q \in Y(K)$ not lying above a Weierstrass point or some point at above ∞ on X, the defining property stated in the proposition is immediate. For the finitely many remaining points the result follows by application of the moving lemma.

Given $(\delta, s) \in L^\times \times k^\times$ representing an element of $\boldsymbol{H}^1_k$ one can construct a ϕ-covering of X as follows. Let $\mathbb{P}_\Omega$ be the projective space with coordinates parametrized by Ω. Define a curve $Y_{\delta,s} \subset \mathbb{P}_\Omega \times X$ by declaring that

$$\big((u_\omega)_{\omega\in\Omega}, (x : y : z)\big) \in Y_{\delta,s}$$

if and only if there exists some $a \in k^\times$ such that

$$\delta(\omega) u_\omega^p = a(x - x(\omega)z) \text{ for all } \omega \in \Omega, \quad \text{and} \quad s \prod_\omega u_\omega^{n_\omega} = a^d y. \tag{5-3}$$

Recall that $\delta \in L$ can be interpreted as a map $\delta : \Omega \to \bar{k}$ and that n_ω denotes the weight associated to ω in the weighted norm map $N : L \to k$. Projection onto the second factor gives $Y_{\delta,s}$ the structure of an X-torsor under $J[\phi]$. It is easy to see that the isomorphism class of $Y_\delta \to X$ depends only on the class of (δ, s) in $\boldsymbol{H}^1_k$. Suppose $(Y, \pi) \in \mathrm{Cov}^\phi_0(X/k)$ and $\mathfrak{f}(Y, \pi) = (\epsilon, t)$. Then (with notation as above) we can find a projective embedding $Y \to \mathbb{P}^N$ and linear forms l_ω which cut out the divisors $\pi^*[\omega]$. The rational map $\mathbb{P}^N \to \mathbb{P}_\Omega$ given by $(l_\omega)_{\omega\in\Omega}$ gives an isomorphism (of X-schemes) $Y \to Y_{\epsilon,t}$. This shows that the $Y_{\delta,s}$ are ϕ-coverings. It is evident from the construction that the pullback of any ramification point $\omega \in X$ is the hyperplane section of $Y_{\delta,s}$ cut out by $u_\omega = 0$. So this covering represents an element of $\mathrm{Cov}^\phi_0(X/k)$. Moreover, it is clear that the image of $(Y_\delta, \pi_{\delta,s})$ under $\mathfrak{f}$ is represented by (δ, s). This shows that $\mathfrak{f}$ is surjective.

Now we show that the map is affine with respect to the action of

$$\boldsymbol{H}^0_k \simeq (\partial(\bar{L}^\times))^{\mathfrak{g}_k} / \iota(k^\times)\partial(L^\times).$$

For this suppose $\alpha \in \bar{L}^\times$ with $\partial\alpha = (\alpha^p, N(\alpha)) \in (\partial(\bar{L}^\times))^{\mathfrak{g}_k}$. Multiplication by α induces a $\bar{k}$-automorphism of $\mathbb{P}_\Omega$. It is evident from (5-3) that this induces an isomorphism of X-schemes $\alpha : (Y_{\alpha^p\delta, N(\alpha)s}, \pi_{\alpha^p\delta, N(\alpha)s}) \longrightarrow (Y_{\delta,s}, \pi_{\delta,s})$. The cocycle $\xi \in H^1(k, J[\phi])$ corresponding to this twist sends $\sigma \in \mathfrak{g}_k$ to

$$\alpha^\sigma \circ \alpha^{-1} \in \mathrm{Aut}((Y_\delta, \pi_{\delta,s})) \simeq J[\phi].$$

Under the isomorphism $(\partial(\bar{L}^\times))^{\mathfrak{g}_k}/k^\times\partial(L^\times) \simeq \ker(\Upsilon) \subset H^1(k, J_{\mathfrak{m}}[\phi])$ from Proposition 5.2, the class of $\partial\alpha$ corresponds to the class of the cocycle η that sends $\sigma \in \mathfrak{g}_k$ to $q(\alpha^\sigma/\alpha) \in J[\phi]$, where $q : J_{\mathfrak{m}}[\phi] \to J[\phi]$ is the quotient map in the exact sequence (5-1). It is then clear that ξ and η give the same class in $H^1(k, J[\phi])$. This proves that $\mathfrak{f}$ is affine. □

6. A descent map for coverings of $\mathscr{X}$

We consider the subset $\mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k) \subset \mathrm{Cov}^\phi(\mathscr{X}/k)$ consisting of ϕ-coverings of $\mathscr{X}$ such that the corresponding ϕ-covering of X lies in $\mathrm{Cov}^\phi_0(X/k)$, and we define a map

$$\mathfrak{F} : \mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k) \to \mathrm{Cov}^\phi_0(X/k) \xrightarrow{\mathfrak{f}} \boldsymbol{H}^1_k.$$

Proposition 5.4 implies that $\mathfrak{F}$ is an affine isomorphism.

Lemma 6.1. *Suppose $X(k) \neq \varnothing$.*

(1) *If $(\mathscr{Y}, \pi) \in \mathrm{Cov}^\phi(\mathscr{X}/k)$ and $\mathscr{Y}(k) \neq \varnothing$, then $(\mathscr{Y}, \pi) \in \mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k)$.*

(2) *If $(\mathscr{Y}, \pi) \in \mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k)$ and $Q \in \mathscr{Y}(K)$ for some extension K/k, then*

$$f(\pi(Q)) = \mathfrak{F}((\mathscr{Y}, \pi)) \text{ in } \boldsymbol{H}^1_K.$$

Proof. By assumption there is some point $R \in X(k) \neq \varnothing$. Then there exists $(Y, \pi) \in \mathrm{Cov}^\phi_0(X/k)$ and $R' \in Y(k)$ such that $\pi(R') = R$. Let $(\mathscr{Y}, \widetilde{\pi}) \in \mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k)$ be the corresponding covering and $i_Y : Y \to \mathscr{Y}$ the base change of $i_X : X \to \mathscr{X}$. Clearly $i_Y(R') \in \mathscr{Y}(k) \neq \varnothing$.

The set B of isomorphism classes of ϕ-coverings of $\mathscr{X}$ which contain a k-rational point is a principal homogeneous space for the image of $J(k)$ under the connecting homomorphism in the Kummer sequence (2-2). This image is contained in $\ker(\Upsilon)$, so $B \subset (\mathscr{Y}, \pi) \cdot \ker(\Upsilon) = \mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k)$. This proves statement (1).

For statement (2), consider the map $d : \mathrm{Pic}^1(X) \to \mathrm{Cov}^\phi(\mathscr{X}/k)$ sending a point $P \in \mathrm{Pic}^1(X) = \mathscr{X}(k)$ to the unique covering to which P lifts. This map is affine, since $f : \mathrm{Pic}^0(X) \to \boldsymbol{H}^0_k \simeq \ker(\Upsilon) \subset H^1(k, J[\phi])$ can be identified with the connecting homomorphism in the Kummer sequence [21, Theorem 1.1]. Moreover its image lands in $\mathrm{Cov}^\phi_{\mathrm{good}}(\mathscr{X}/k)$ by statement (1).

It suffices to prove the statement for $K = k$, which amounts to showing that $f(D) = \mathfrak{F}(d(D))$ for every $D \in \mathrm{Pic}^1(X)$. The point $i_Y(R') \in \mathcal{Y}(k)$ is a lift of $[R] \in \mathcal{X}(k)$, so $\mathcal{Y} = d([R])$. From the definition of $\mathfrak{F}$ and the defining property of $\mathfrak{f}$ we have

$$\mathfrak{F}((\mathcal{Y}, \widetilde{\pi})) = \mathfrak{f}((Y, \pi)) = f([\pi(R')]) = f([R]) \in \boldsymbol{H}^1_k.$$

Hence $f([R]) = \mathfrak{F}(d([R]))$.

Now suppose $D \in \mathrm{Pic}^1(X)$. Since d is affine, $d(D)$ is the twist of $d([R])$ by the cocycle $f(D - [R]) \in \boldsymbol{H}^0_k \simeq \ker(\Upsilon)$. Since $\mathfrak{F}$ is affine, we have

$$\mathfrak{F}(d(D)) = \mathfrak{F}(d[R]) \cdot f(D - [R]) = f(D)\mathfrak{F}(d[R])/f([R]) = f(D).$$

This completes the proof. □

We have the following analogue of Corollary 5.5, which, with Proposition 2.3, implies Theorem 4.5.

Proposition 6.2. *Suppose k is a global field and X is everywhere locally solvable. Then $\mathfrak{F}$ restricts to an affine isomorphism* $\mathrm{Sel}^{\phi}(\mathcal{X}/k) \to \mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathcal{X}/k)$.

Proof. First off, let us show that $\mathrm{Sel}^{\phi}(\mathcal{X}/k) \subset \mathrm{Cov}^{\phi}_{\mathrm{good}}(\mathcal{X}/k)$. Suppose that $(\mathcal{Y}, \pi) \in \mathrm{Sel}^{\phi}(\mathcal{X}/k)$ and that X is everywhere locally solvable. Consider the covering $\widetilde{\pi} : Y \to X$ obtained by pulling back. We want to show that the pullback to Y of some ramification point on X is linearly equivalent to a k-rational divisor. The obstruction to a k-rational divisor class being represented by a k-rational divisor is an element of the Brauer group of k. Since the Brauer group of a global field satisfies the local-global principle it suffices to show that $(Y, \widetilde{\pi})$ gives a class in $\mathrm{Cov}^{\phi}_0(X/k_v)$ for every prime v. This follows from Lemma 6.1(1) since we have assumed both X and $\mathcal{Y}$ are everywhere locally solvable.

Now let us show that $\mathfrak{F}$ maps the ϕ-Selmer set to the algebraic ϕ-Selmer set. Let $(\mathcal{Y}, \pi) \in \mathrm{Sel}^{\phi}(\mathcal{X}/k)$ and set $\delta = \mathfrak{F}((\mathcal{Y}, \pi))$. For every completion k_v of k, $X(k_v) \neq \varnothing$, so we may apply Lemma 6.1(2) over k_v. This shows that $\mathrm{res}_v(\delta) \in f(\mathrm{Pic}^1(X_{k_v}))$ for every v. Consequently, δ lies in the algebraic ϕ-Selmer set.

It now suffices to show that the map in the statement is surjective, as it is the restriction of an affine isomorphism. For this let δ be an element in the algebraic ϕ-Selmer set. Then $\delta \in \boldsymbol{H}^1_k$, so $\delta = \mathfrak{F}((\mathcal{Y}, \pi))$ for some $(\mathcal{Y}, \pi) \in \mathrm{Cov}^{\phi}_{\mathrm{good}}(\mathcal{X}/k)$. We need to show that $\mathcal{Y}$ is everywhere locally solvable. For each prime v we can find $P_v \in \mathrm{Pic}^1(X_{k_v}) \subset \mathcal{X}(k_v)$ such that $\mathrm{res}_v(\delta) = f(P_v)$. The point P_v lifts to a k_v-point on some ϕ-covering $(\mathcal{Y}_v, \pi_v)$ defined over k_v. Moreover $(\mathcal{Y}_v, \pi_v) \in \mathrm{Cov}^{\phi}_{\mathrm{good}}(\mathcal{X}/k_v)$ by Lemma 6.1(1) and $\mathfrak{F}((\mathcal{Y}_v, \pi_v)) = \mathrm{res}_v(\delta)$ by Lemma 6.1(2). Since $\mathfrak{F}$ is injective we have that $\mathcal{Y} \otimes k_v$ and $\mathcal{Y}_v$ are isomorphic, for each prime v. This implies that $\mathcal{Y}$ is everywhere locally solvable as required. □

7. Examples

We have implemented the algorithm described in Section 4C in the computer algebra system Magma [3] for degree p cyclic covers of $\mathbb{P}^1$ defined over the p-th cyclotomic field. As a test of the algorithm (and the correctness of the implementation) we performed computations for a large sample of hyperelliptic curves. When at all possible we checked our results for consistency with rank bounds obtained by other means (for example, different implementations of descent on elliptic curves and Jacobians of hyperelliptic curves, points of small height on the Jacobian, information obtained assuming standard conjectures, and so on). Some of the resulting data is presented at the end of this section. In addition to this we offer the following examples.

Example 7.1. The two hyperelliptic curves

$$X_1: \quad y^2 + (x^3 + x + 1)y = x^6 + 5x^5 + 12x^4 + 12x^3 + 6x^2 - 3x - 4,$$
$$X_2: \quad y^2 + (x^3 + x + 1)y = -2x^6 + 7x^5 - 2x^4 - 19x^3 + 2x^2 + 18x + 7$$

over $\mathbb{Q}$ have Mordell-Weil rank 0, and the 2-primary parts of their Shafarevich-Tate groups are isomorphic to $\mathbb{Z}/2 \times \mathbb{Z}/2$.

Proof. Let J_i denote the Jacobian of X_i and $\mathscr{X}_i$ denote $\mathbf{Pic}^1(X_i)$. The X_i are everywhere locally solvable double covers of $\mathbb{P}^1$. Using Magma we computed that $\mathrm{Sel}^2(J_i/\mathbb{Q})$ has $\mathbb{F}_2$-dimension 2 and that the 2-Selmer set of $\mathbf{Pic}^1(X_i)$ is empty for $i = 1, 2$. The result then follows from Theorem 4.5 and its corollary. □

Remark. These curves were taken from [9] (where they were labeled $C_{125,B}$ and $C_{133,A}$), where it is shown that the order of the 2-torsion subgroup of Ш is equal to the order of Ш predicted by the Birch and Swinnerton-Dyer conjectural formula for several modular Jacobian surfaces. In particular, it is proved in [9] that the formula holds for those Jacobians considered if and only if $2\text{Ш} = 0$. For the curves considered one can determine the rank (unconditionally) by analytic means, so a 2-descent on the Jacobian determines Ш[2], but it only determines Ш(2) when $\dim_{\mathbb{F}_2} \text{Ш}[2] \leq 1$. Apart from the two curves above, all curves considered in [9] had $\dim_{\mathbb{F}_2} \text{Ш}[2] \leq 1$. So from the example above one can now conclude for the curves considered in [9] that the conjectural formula holds if and only if Ш has no elements of odd order.

Example 7.2. Let $X/\mathbb{Q}$ be the genus 4 cyclic cover of $\mathbb{P}^1$ with affine equation

$$X: y^3 = 3(x^6 + x^4 + 4x^3 + 2x^2 + 4x + 3).$$

Then X is everywhere locally solvable, yet has no $\mathbb{Q}$-rational divisors of any degree prime to 3. Moreover, the Jacobian J of X has Mordell-Weil rank 1 and the 3-primary part of its Shafarevich-Tate group is isomorphic to $\mathbb{Z}/3 \times \mathbb{Z}/3$.

Proof. We first note that X is everywhere locally solvable. In order to apply the results of this paper, we work over the field $k = \mathbb{Q}(\zeta_3)$ obtained by adjoining a primitive cube root of unity ζ_3. To prove the result we do ϕ-descents on J_k and $\mathbf{Pic}^1(X_k)$, for $\phi = 1 - \zeta_3$. Using Magma we computed that the ϕ-Selmer group of J_k has $\mathbb{F}_3$-dimension 3. From the exact sequence (2-2) it follows that

$$\dim_{\mathbb{F}_3} \frac{J(k)}{\phi J(k)} + \dim_{\mathbb{F}_3} Ш(J/k)[\phi] = 3.$$

We then computed $\mathrm{Sel}^{\phi}_{\mathrm{alg}}(\mathbf{Pic}^1(X_k)/k)$ and found it to be empty. Using Corollary 4.6 this lowers the upper bound for the dimension of $J(k)/\phi J(k)$ to 1.

The divisor on $\mathbb{P}^1$ defined by $x^3 - x^2 + 4x + 4 = 0$ lifts to a degree 3 $\mathbb{Q}$-rational divisor D on X. One can check that the image of the class of $D - \mathfrak{m}$ under $f : \mathrm{Pic}^0(X_k) \to \boldsymbol{H}^0_k$ is nontrivial. So we find that $J(k)/\phi J(k)$ has dimension 1. This gives an upper bound of 2 for the dimension of $Ш(J/k)[\phi]$, so by Corollary 4.6, $Ш(J/k)(3) \simeq Ш(J/k)[\phi] \simeq \mathbb{Z}/3 \times \mathbb{Z}/3$. On the other hand, $\mathbf{Pic}^1(X)$ represents an element of $Ш(J/\mathbb{Q})[3]$ which is not divisible by 3 (since it is not divisible by 3 over k). On the other hand the dimension of $Ш(J/\mathbb{Q})[3]$ is even [14], so it is at least 2. Now the map $Ш(J/\mathbb{Q})(3) \to Ш(J/k)(3)$ obtained by extension of scalars is injective since $[k : \mathbb{Q}] = 2$ is prime to 3, so we must have $Ш(J/\mathbb{Q})(3) \simeq \mathbb{Z}/3 \times \mathbb{Z}/3$.

It remains to compute the rank. The Galois group acts on the ramification points as the full symmetric group, from which it follows that there is no nontrivial k-rational ϕ-torsion in $J(k)$. By [17, Corollary 3.7 and Proposition 3.8] it follows that

$$\mathrm{rank}(J(k)) = [k : \mathbb{Q}] \cdot \left(\dim \frac{J(k)}{\phi J(k)} - \dim J(k)[\phi] \right) = 2, \quad \text{and}$$

$$\mathrm{rank}(J(\mathbb{Q})) = \frac{\mathrm{rank}(J(k))}{[k : \mathbb{Q}]} = 1.$$

In fact, $D - \mathfrak{m}$ represents a point of infinite order in $J(\mathbb{Q})$. □

Remark. From a ϕ-descent on J alone, one is only able to conclude that $1 \leq \dim_{\mathbb{F}_3} J(k)/\phi J(k) \leq 3$, giving $1 \leq \mathrm{rank}(J(\mathbb{Q})) \leq 3$.

Example 7.3 (Data for hyperelliptic curves). For $g \in \{2, 3, 4\}$ we tested our algorithm on various samples of hyperelliptic curves of genus g. For varying values of N, we randomly chose 10,000 separable polynomials $h(x) = \sum_{i=1}^{2g+2} h_i x^i$ of degree at least $2g + 1$ and with integer coefficients h_i bounded in absolute value by N. For each of the genus g curves X defined by $y^2 = h(x)$, we computed $\mathrm{Sel}^2_{\mathrm{alg}}(J/\mathbb{Q})$ and $\mathrm{Sel}^2_{\mathrm{alg}}(\mathscr{X}/\mathbb{Q})$, assuming the generalized Riemann hypothesis for reasons of efficiency. If the latter set was empty, we noted whether or not this was because $\mathrm{Pic}^1(X_{\mathbb{Q}_p})$ was empty for some prime $p \leq \infty$. The resulting data

g	N	$\mathrm{Sel}^2(\mathscr{X}/\mathbb{Q}) = \varnothing$	Rank	Rank*	Improvement	Improvement*
2	5	2146	7873	9819	29%	84%
		981	**848**	**977**		
2	10	3088	5315	9346	22%	73%
		1778	**1295**	**1752**		
2	20	3787	3411	8392	17%	59%
		2420	**1350**	**2317**		
2	50	4297	2156	6955	15%	46%
		2916	**1350**	**2637**		
3	5	2101	2540	7573	8%	32%
		1228	**645**	**1164**		
3	10	2801	1477	5840	8%	29%
		1857	**786**	**1619**		
4	5	1991	1717	6031	6%	22%
		1278	**484**	**1127**		
4	10	2687	1296	5145	8%	25%
		1952	**726**	**1644**		

Table 1. Data for hyperelliptic curves. For reasons of efficiency, all computations summarized in this table were made under the assumption of the generalized Riemann hypothesis; furthermore, in the columns marked by asterisks, we also assumed that $Ш_{\mathrm{div}} = 0$. The first two columns indicate the genus g and the coefficient height bound N of the examples considered in a given row. The third column counts the number of curves (out of 10,000 randomly chosen hyperelliptic curves of the given genus and coefficient height bound) for which $\mathrm{Sel}^2(\mathscr{X}/\mathbb{Q}) = \varnothing$; the bold figures give the number of times the explanation was *not* simply that $\mathrm{Pic}^1(X_{\mathbb{Q}_p})$ is empty for some prime $p \leq \infty$. The "Rank" columns give the number of curves for which the rank could be computed (under the assumptions indicated), with the numbers in bold giving the number of curves for which information from our algorithm was needed to complete the computation. The final two columns give the "improvement factor" in the rank computations: of the sample curves whose ranks could not be determined by earlier methods, the fraction whose ranks could be determined using our algorithm (under the assumptions indicated).

is summarized in Table 1. The boldfaced entries correspond to curves where our algorithm provided information that would not otherwise have been obtained.

It is also interesting to consider how often the combined information yields a sharp upper bound for the Mordell-Weil rank. This will be the case if (i) $\mathscr{X}$ is either trivial or not divisible by 2 in $Ш(J/\mathbb{Q})$; (ii) the number of primes where X fails to have divisors of degree 1 locally is at most one (respectively, not even and positive when the genus is even); and (iii) $Ш(J/\mathbb{Q})[2]$ contains at most two elements

linearly independent from $\mathscr{X}$. The assumptions (i) and (ii) imply, respectively, that in order for $\mathscr{X}(\mathbb{Q})$ to be empty it is necessary and sufficient that $\mathrm{Sel}^2_{\mathrm{alg}}(\mathscr{X}/\mathbb{Q})$ be empty, while (iii) guarantees that determining whether $\mathscr{X}(\mathbb{Q})$ is empty is sufficient to deduce a sharp bound.

With this in mind we used a point search to compute a lower bound for the rank for each curve, both with and without assuming that the divisible subgroup of $Ш(J/\mathbb{Q})$ is trivial (the assumption allows us to determine the parity of the rank). When this matched the upper bound it means we computed the rank, and in such cases we counted the number of curves where the additional information provided by $\mathrm{Sel}^2_{\mathrm{alg}}(\mathscr{X}/\mathbb{Q})$ was needed. We then computed the proportion of curves for which the rank could be determined with the additional information provided by our algorithm among those for which the rank could not be determined by descent on the Jacobian alone.

For example, in the sample of genus 2 curves with $N = 10$ the method yielded new information for about 17% of the curves, which (assuming $Ш_{\mathrm{div}} = 0$) increased our success rate from about 76% to about 93%, handling about 73% of the curves left previously undecided by the descent on the Jacobian.

References

[1] Michael Artin and John Tate (eds.), *Arithmetic and geometry*, *vol. I*, Progress in Mathematics, no. 35, Birkhäuser, Boston, 1983. MR 84j:14005a

[2] Siegfried Bosch, Werner Lütkebohmert, and Michel Raynaud, *Néron models*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), no. 21, Springer, Berlin, 1990. MR 91i:14034

[3] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system, I: The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265. MR 1484478

[4] Nils Bruin and Michael Stoll, *Deciding existence of rational points on curves: An experiment*, Experiment. Math. **17** (2008), no. 2, 181–189. MR 2009d:11100

[5] ———, *Two-cover descent on hyperelliptic curves*, Math. Comp. **78** (2009), no. 268, 2347–2370. MR 2010e:11059

[6] J. W. S. Cassels, *The Mordell-Weil group of curves of genus* 2, in Artin and Tate [1], 1983, pp. 27–60. MR 84k:14032

[7] Brendan Matthew Creutz, *Explicit second p-descent on elliptic curves*, Ph.D. thesis, Jacobs University, Bremen, Germany, 2010. http://www.jacobs-university.de/phd/files/1283816493.pdf

[8] Tom Fisher, *A counterexample to a conjecture of Selmer*, in Reid and Skorobogatov [15], 2003, pp. 119–131. MR 2005a:11077

[9] E. Victor Flynn, Franck Leprévost, Edward F. Schaefer, William A. Stein, Michael Stoll, and Joseph L. Wetherell, *Empirical evidence for the Birch and Swinnerton-Dyer conjectures for modular Jacobians of genus 2 curves*, Math. Comp. **70** (2001), 1675–1697. MR 2002d:11072

[10] Daniel M. Gordon and David Grant, *Computing the Mordell-Weil rank of Jacobians of curves of genus two*, Trans. Amer. Math. Soc. **337** (1993), no. 2, 807–824. MR 93h:11057

[11] J. S. Milne, *Arithmetic duality theorems*, 2nd ed., BookSurge, LLC, Charleston, SC, 2006. MR 2007e:14029

[12] Michael Mourao, *Descent on superelliptic curves*, 2011. arXiv 1010.2360v3 [math.NT]

[13] Bjorn Poonen and Edward F. Schaefer, *Explicit descent for Jacobians of cyclic covers of the projective line*, J. Reine Angew. Math. **488** (1997), 141–188. MR 98k:11087

[14] Bjorn Poonen and Michael Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2) **150** (1999), no. 3, 1109–1149. MR 2000m:11048

[15] Miles Reid and Alexei Skorobogatov (eds.), *Number theory and algebraic geometry*, London Mathematical Society Lecture Note Series, no. 303, Cambridge University Press, 2003. MR 2004k:00024

[16] Edward F. Schaefer, 2-*descent on the Jacobians of hyperelliptic curves*, J. Number Theory **51** (1995), no. 2, 219–232. MR 96c:11066

[17] ———, *Computing a Selmer group of a Jacobian using functions on the curve*, Math. Ann. **310** (1998), no. 3, 447–471, erratum: [18]. MR 99h:11063

[18] ———, *Erratum: "Computing a Selmer group of a Jacobian using functions on the curve"* [*Math. Ann.* **310** (*1998*)*, no. 3, 447–471*], Math. Ann. **339** (2007), no. 1, 1. MR 2008f:11063

[19] Jean-Pierre Serre, *Algebraic groups and class fields*, 2nd ed., Graduate Texts in Mathematics, no. 117, Springer, New York, 1998. MR 88i:14041

[20] Michael Stoll, *Implementing 2-descent for Jacobians of hyperelliptic curves*, Acta Arith. **98** (2001), no. 3, 245–277. MR 2002b:11089

[21] Michael Stoll and Ronald van Luijk, *Unfaking the fake Selmer group*, 2011. arXiv 1108.3364 [math.AG]

BRENDAN CREUTZ: brendan.creutz@sydney.edu.au
School of Mathematics and Statistics, University of Sydney, Sydney, NSW 2006, Australia

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.317

msp

Computing equations of curves with many points

Virgile Ducet and Claus Fieker

We explain how to compute the equations of the abelian coverings of any curve defined over a finite field. Then we describe an algorithm which computes curves with many rational points with respect to their genus. The implementation of the algorithm provides seven new records over $\mathbb{F}_2$.

1. Introduction

The motivation for finding curves defined over a finite field $\mathbb{F}_q$ with many rational points compared to their genus comes from the theory of error-correcting codes. Let C be a (n, k, d)-*code*, that is, a subvector space of $\mathbb{F}_q^n$ of dimension k in which every nonzero vector has at least d nonzero coordinates in a fixed basis. For given parameters n and k, one wishes to find codes with the largest possible correction capacity $(d-1)/2$.

In a 1977 paper, Goppa [7] proposed a method for constructing codes which is based on algebraic geometry. Let X be a (nonsingular projective irreducible) curve X defined over $\mathbb{F}_q$. Let $D_1 = P_1 + \cdots + P_n$ and D_2 be two divisors over X with disjoint support such that the points P_i are rational and such that $2g-2 < \deg D_2 < n$. Let $\Omega_X(D_1 - D_2)$ be the space of differentials ω on X such that $\operatorname{div}(\omega) \geq D_2 - D_1$, and for every differential ω let $\operatorname{res}_{P_i}(\omega)$ denote the residue of ω at P_i. The *Goppa code* $C(X, D_1, D_2)$ associated to this data is the image of the $\mathbb{F}_q$-linear map $\Omega_X(D_1 - D_2) \to \mathbb{F}_q^n$ defined by $\omega \mapsto (\operatorname{res}_{P_1}(\omega), \ldots, \operatorname{res}_{P_n}(\omega))$. For these codes, the Riemann-Roch theorem shows that $k = g - 1 + n - \deg D_2$ and that

$$\frac{k}{n} + \frac{d}{n} \geq 1 + \frac{1}{n} - \frac{g}{n}.$$

By construction, n is bounded by the number of rational points $N(X)$ of X, and from the above inequality, for given n and k, the smaller the genus, the more

MSC2010: primary 11R37; secondary 14H45.
Keywords: explicit class field theory, Kummer theory, Witt vectors, curves with many points, equations of abelian coverings.

efficient the code. So one would like to find, for every n, the smallest genus g such that there exists a curve $X/\mathbb{F}_q$ with at least n rational points. The moral of all this is that one must look for curves with many rational points compared to their genus, for every genus.

The idea of using class field theory to construct abelian coverings with many rational points over a finite field comes from Serre (see [22]). His Harvard course notes [23] remain a very useful reference with a lot of material. Niederreiter and Xing continued the search for good curves and devoted many papers to finding and exploiting new techniques; in particular, they make use of the explicit description of ray class fields provided by the theory of Drinfel'd modules. Their book [18] includes all their work on the subject and much more. In a series of paper in the late 90s, Lauter [12; 13; 14] extended Serre's method and obtained new records by studying the degrees of certain abelian extensions of the rational function field ramified at a single rational place and totally split at the others. She also interpreted several known families of curves as particular class field theoretical constructions. Auer (see his Ph.D. thesis [1] or the ANTS paper [2] for a summary of the results) extended Lauter's work and described an algorithm to compute the degree of the maximal abelian extension of any function field ramified at most one place and with prescribed splitting behavior. This allowed him to find many new curves improving the known records. We conclude this historical survey by noting that only in a few cases can one deduce the equation of the curve from its theoretical construction; in particular, the so-called "explicit" description via Drinfel'd modules is very difficult to use.

In the present article, we use explicit class field theory to compute the equations of the abelian coverings of a curve defined over a finite field, and we apply this method to the problem of finding curves with the maximum possible number of rational points compared to their genus. The paper is divided as follows. In the first section we explain the link between ray class groups and abelian coverings. Then we describe how to use explicit class field theory to compute the equation of an abelian covering of a curve from knowledge of the corresponding ray class group. In Section 4 we present an algorithm to find good curves, and we give an overview of our results in Section 5.

2. Ray class groups

We first recall the main aspects of class field theory in the classical language of ray class groups. The reader is referred to [10], [15], or [25] for the proofs.

Let K be a global function field defined over a finite field $\mathbb{F}_q$; K should be thought of as the function field of a curve X defined over $\mathbb{F}_q$. The set of places of K is denoted by Pl_K. Let $\mathfrak{m}$ be a *modulus* on K, that is, an effective divisor

over K. Let $\mathrm{Div}_{\mathfrak{m}}$ be the group of divisors of K whose support is disjoint from that of $\mathfrak{m}$, and let $P_{\mathfrak{m},1}$ be the subgroup of divisors of functions "congruent to 1 modulo $\mathfrak{m}$":

$$P_{\mathfrak{m},1} = \{\mathrm{div}(f) : f \in K^{\times} \text{ and } v_P(f-1) \geq v_P(\mathfrak{m}) \text{ for all } P \in \mathrm{Supp}(\mathfrak{m})\}.$$

A subgroup H of $\mathrm{Div}_{\mathfrak{m}}$ of finite index is called a *congruence subgroup modulo* $\mathfrak{m}$ if H contains $P_{\mathfrak{m},1}$.

By the Artin reciprocity law, for every finite abelian extension L of K there exist a modulus $\mathfrak{m}$ and a congruence subgroup $H_{\mathfrak{m}}(L)$ modulo $\mathfrak{m}$ such that the Artin map provides an isomorphism of groups

$$\mathrm{Gal}(L/K) \cong \mathrm{Div}_{\mathfrak{m}}/H_{\mathfrak{m}}(L).$$

Such a $\mathfrak{m}$ is called an *admissible modulus for* L/K; it is not unique (whereas for a given $\mathfrak{m}$, $H_{\mathfrak{m}}(L)$ is), but there exists an admissible modulus $\mathfrak{f}_{L/K}$ for L/K, called the *conductor of* L/K, which is smaller than the others in the sense that every admissible modulus $\mathfrak{m}$ for L/K satisfies $\mathfrak{f}_{L/K} \leq \mathfrak{m}$ (as divisors). An important property of the conductor of an abelian extension is that its support consists of exactly those places that are ramified.

The existence theorem of class field theory guarantees, for every modulus $\mathfrak{m}$ and every congruence subgroup $H_{\mathfrak{m}}$ modulo $\mathfrak{m}$, the existence of a unique global function field $L_{\mathfrak{m}}(H_{\mathfrak{m}})$, possibly defined over a constant field extension, that is a finite abelian extension of K such that $\mathrm{Gal}(L_{\mathfrak{m}}(H_{\mathfrak{m}})/K) \cong \mathrm{Div}_{\mathfrak{m}}/H_{\mathfrak{m}}$. The field $L_{\mathfrak{m}}(H_{\mathfrak{m}})$ is called the *class field* of $H_{\mathfrak{m}}$. Note that by definition of the conductor, we have $\mathfrak{f}_{L_{\mathfrak{m}}(H_{\mathfrak{m}})/K} \leq \mathfrak{m}$.

Instead of working with congruence subgroups modulo a certain $\mathfrak{m}$, it is sometimes more convenient to consider subgroups of the *ray class group modulo* $\mathfrak{m}$, which is the quotient group $\mathrm{Pic}_{\mathfrak{m}} = \mathrm{Div}_{\mathfrak{m}}/P_{\mathfrak{m},1}$. To each congruence subgroup H modulo $\mathfrak{m}$, one can associate the subgroup $\overline{H} = H/P_{\mathfrak{m},1}$ of $\mathrm{Pic}_{\mathfrak{m}}$ of finite index. This correspondence is one-to-one, and furthermore we have an isomorphism $\mathrm{Pic}_{\mathfrak{m}}/\overline{H} \cong \mathrm{Div}_{\mathfrak{m}}/H$. We can thus restate what has been said above as follows:

Theorem 1 (Main theorem of class field theory). *Let* $\mathfrak{m}$ *be a modulus. There is a one-to-one inclusion reversing correspondence between subgroups* H *of* $\mathrm{Pic}_{\mathfrak{m}}$ *of finite index and finite abelian extensions* L *of* K *with conductor less than* $\mathfrak{m}$. *Furthermore, the Artin map provides an isomorphism* $\mathrm{Pic}_{\mathfrak{m}}/H \cong \mathrm{Gal}(L/K)$.

3. Computing the equation of an abelian covering

Throughout this section, K is a function field defined over a finite field $\mathbb{F}_q$. We fix a modulus $\mathfrak{m}$ and a congruence subgroup H modulo $\mathfrak{m}$, and we explain how to compute the class field L of H. The similar approach for number fields has

been introduced by the second author in [6], where one will find more algorithmic details; the computations of groups of units and ray class groups are explained in [8].

3.1. ***Reduction to the cyclic case.*** First, we show that we can reduce the problem to the case of a cyclic extension of prime power degree. For this, we use the fundamental theorem of abelian groups to decompose $\overline{H} = \mathrm{Div}_{\mathfrak{m}}/H$ as a finite product of cyclic groups $\overline{H} = \prod_{i=1}^{d} \overline{H}_i$, where each $\overline{H}_i$ is of the form $\mathrm{Div}_{\mathfrak{m}}/H_i$ for a subgroup $H \subseteq H_i \subseteq \mathrm{Div}_{\mathfrak{m}}$ such that $\overline{H}_i \cong \mathbb{Z}/p_i^{m_i}\mathbb{Z}$ for some prime number p_i and some positive integer m_i. For every i, let L_i be the class field of H_i, so $\mathrm{Gal}(L_i/K) \cong \overline{H}_i$, and let L' be the composite field $L_1 L_2 \cdots L_d$. By general Galois theory, $\mathrm{Gal}(L'/K)$ is isomorphic to the subgroup of elements of $\prod_{i=1}^{d} \mathrm{Gal}(L_i/K)$ which agree on $L_1 \cap \cdots \cap L_d$. The functoriality of the Artin map implies that the previous condition is always true, so $\mathrm{Gal}(L'/K) \cong \prod_{i=1}^{d} \mathrm{Gal}(L_i/K)$. Thus $\mathrm{Gal}(L/K)$ and $\mathrm{Gal}(L'/K)$ are equal, and by the uniqueness property of the class field, we conclude that $L = \prod_{i=1}^{d} L_i$. Also, note that if we have equations for two abelian extensions L_1/K and L_2/K, then there are algorithms based on the theory of resultants to compute an equation of $L_1 L_2/K$.

3.2. ***Cyclic case: $l \neq p$.*** Now suppose that $\overline{H}$ is cyclic of prime power degree $n = l^m$ for a prime l different from p and an integer $m \geq 1$. As in the proof of the existence theorem (see [10, Chapter XI, §2]), the idea consists of reducing to the case when K contains the n-th roots of unity, and then to use explicit Kummer theory. So let $K' = K(\zeta_n)$ and set $L' = LK'$: We will "translate" the problem to the extension L'/K'. (Note that the extension K'/K is a constant field extension, hence it is unramified.)

We will refer to Figure 1; the solid lines in the figure connect fields that are actually constructed during the execution of the algorithm, while dotted lines connect fields that are only implicitly used.

Since L/K is cyclic of degree n, the field $L' := L(\zeta_n) = K'L$ is a Kummer extension of K', and hence there exists a nonzero element $\alpha \in K'$ such that $L' = K'(\sqrt[n]{\alpha})$. Since L'/K has to be unramified outside places in the modulus $\mathfrak{m}$ of L/K, there exists a set S of places of K', depending only on $\mathfrak{m}$ and K', such that α can be chosen as an element of the S-units U_S, that is, as an element that has no poles outside S; in particular, L'/K' is unramified[1] outside S. Let $\mathfrak{m}'$ be an admissible modulus for L'/K', and assume without loss of generality that $\mathfrak{m}'$ is supported on S. By the Dirichlet unit theorem, we have $U_S = \langle \epsilon_1, \ldots, \epsilon_s \rangle$ for independent elements ϵ_i $(1 \leq i \leq s-1)$ and a torsion unit ϵ_s. Set $M := K'(\sqrt[n]{U_S})$, so that $\mathrm{Gal}(M/K') = (\mathbb{Z}/n\mathbb{Z})^s$. For any place P of K' unramified in M/K', the Frobenius

[1]This is a general property of Kummer extensions, which follows from Hensel's lemma; see for example [17, Lemma V.3.3].

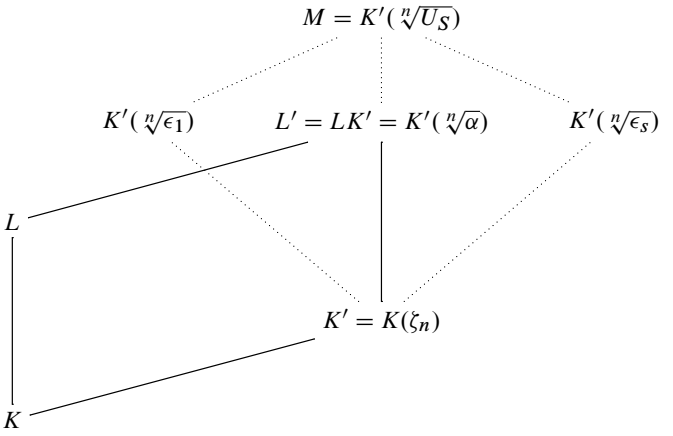


Figure 1. Fields used implicitly in the discussion.

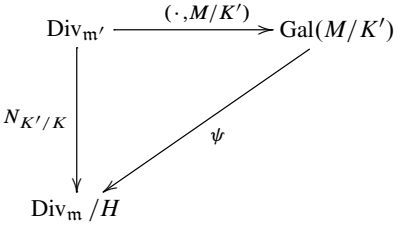


Figure 2. Definition of ψ.

$(P, M/K')$ at P is defined by its operation on the $\sqrt[n]{\epsilon_i}$. Since M/K' is unramified outside S, we see that we get a map $\mathrm{Div}_{\mathfrak{m}'} \to (\mathbb{Z}/n\mathbb{Z})^s$ defined by $P \mapsto (n_i)$, where $\sqrt[n]{\epsilon_i} \mapsto \zeta_n^{n_i}\sqrt[n]{\epsilon_i}$ and $\sqrt[n]{\epsilon_i}^N \equiv \zeta_n^{n_i}\sqrt[n]{\epsilon_i} \bmod P$, where N is the cardinality of the residue field $\mathbb{F}_P$ of K' at P. In particular, $N \equiv 1 \bmod n$ because $\mathbb{F}_P$ contains the n-th roots of unity, and thus n_i is defined by $\epsilon_i^{\lceil N/n \rceil} \equiv \zeta_n^{n_i} \bmod P$. To summarize: The Artin map from $\mathrm{Div}_{\mathfrak{m}'}$ to $(\mathbb{Z}/n\mathbb{Z})^s$ is explicit and can be computed in K' already!

To find L' we need to find divisors $D \in \mathrm{Div}_{\mathfrak{m}'}$ such that $(D, M/K')$ fixes L'. By the existence theorem, this is equivalent to $D \in H'$, where H' is the congruence subgroup modulo $\mathfrak{m}'$ whose class field is L'. By standard properties of the Artin map, this reduces to $\mathrm{N}_{K'/K}(D) \in H$. We use this as summarized in Figure 2 to explicitly construct the map ψ: Computing $(P, M/K')$ on the one side and $\mathrm{N}_{K'/K}(P) + H \in \mathrm{Div}_{\mathfrak{m}}/H$ on the other, we collect (small) places outside S until the full group $\mathrm{Gal}(M/K')$ can be generated. The field L' is then obtained as the field fixed by the kernel of ψ.

In order to find α we apply a similar idea (see [6, §4] for details): L'/K is abelian and the Galois group can be computed explicitly. Once the automorphisms

of L'/K are known, we can easily establish again an explicit Artin map, now from $\mathrm{Div}_{\mathfrak{m}}$ to $\mathrm{Gal}(L'/K)$, and find the subgroup fixing L as above. We note that the conductor of L' can be larger than the conductor of L/K, but since L' is obtained via a constant field extension, the ramified primes remain the same, hence the map is well defined and surjective (but the kernel may not be a congruence subgroup modulo $\mathfrak{m}$).

3.3. ***Cyclic case: l = p.*** Finally we turn to the case when L/K is cyclic of degree $n = p^m$, for an integer $m \geq 1$. To begin with, we recall some aspects of Artin-Schreier-Witt theory.

Let k be an arbitrary field and let $\bar{k}$ be an algebraic closure of k. Let r be an integer and let $\mathrm{W}_r(k)$ and $\mathrm{W}_r(\bar{k})$ be the rings of *Witt vectors of length r* with coefficients in k and $\bar{k}$, respectively. Then any $\vec{\alpha}$ in $\mathrm{W}_r(\bar{k})$ can be used to generate an algebraic extension $k(\vec{\alpha})$ of k in the following way: If $\vec{\alpha} = (\alpha_1, \ldots, \alpha_r)$, then we set $k(\vec{\alpha}) = k(\alpha_1, \ldots, \alpha_r)$. This construction can be visualized as a tower:

$$\begin{array}{l} k_r = k(\vec{\alpha}), \\ \uparrow \\ \vdots \\ \uparrow \\ k_2 = k_1(\alpha_2), \\ \uparrow \\ k_1 = k_0(\alpha_1), \\ \uparrow \\ k_0 = k. \end{array}$$

Suppose now that k has positive characteristic p. Let $\wp$ be the Artin-Schreier-Witt operator acting on $\vec{\alpha} \in \mathrm{W}_r(\bar{k})$ by

$$\wp(\vec{\alpha}) = \vec{\alpha}^p - \vec{\alpha}.$$

Then for $\vec{\beta}$ in $\mathrm{W}_r(k)$ the equation $\wp(\vec{\alpha}) = \vec{\beta}$ is algebraic over k, so as above one can consider the extension $k(\wp^{-1}(\vec{\beta}))$. Actually, by explicit Artin-Schreier-Witt theory (see [11, pp. 330–332]), every abelian extension of exponent p^r of k arises as $k(\wp^{-1}(\Delta_r))$ for some subgroup $\Delta_r \subseteq \mathrm{W}_r(k)$ containing $\wp(\mathrm{W}_r(k))$. In particular, a cyclic extension of degree p^r of k is of the form $k(\vec{\gamma})$ for some $\vec{\gamma}$ in $\wp^{-1}(k) \subset \mathrm{W}_r(\bar{k})$, with Galois group generated by the automorphism $\vec{\gamma} \mapsto \vec{\gamma} + (1, 0, \ldots, 0)$ (see [21]).

So for our purposes we take $r = m$, and we can assume that the cyclic extension of degree p^m of K is of the form $L = K(\vec{y})$ for some $\vec{x} \in \mathrm{W}_m(K)$ and $\vec{y} \in \mathrm{W}_m(\bar{k})$

satisfying $\wp(\vec{y}) = \vec{x}$. Now we explain how to compute $\vec{x}$. It is clear that the Artin-Schreier-Witt extension does not change if one replaces $\vec{x}$ with $\vec{x} + \wp(\vec{z})$ for some $\vec{z}$ in $\mathrm{W}_m(K)$, so one will look for $\vec{x}$ as an element of $\mathrm{W}_m(K)/\wp(\mathrm{W}_m(K))$.

We first look at the case $m = 1$; hence we assume that L/K is a cyclic extension of degree p, and write x for $\vec{x}$.

Lemma 2. *Let $y \in K$ be arbitrary. For every place P of K there exists an element $u_P \in K$ such that either $v_P(y + u_P^p - u_P)$ is negative and coprime to p, or $v_P(y + u_P^p - u_P) \geq 0$.*

Proof. If $v_P(y) \geq 0$ or $v_P(y)$ is coprime to p then $u_P := 0$ works, so we henceforth assume that $v_P(y) < 0$ and $p \mid v_P(y)$. Let $\bar{y} := (y\pi^{-v_P(y)})(P) \in \mathbb{F}_P$, where $\mathbb{F}_P$ is the residue class field of K at P and π is a uniformizing element (that is, $v_P(\pi) = 1$). Since the p-power Frobenius is surjective, we can find a $\bar{u} \in \mathbb{F}_P$ such that $\bar{u}^p = -\bar{y}$. Now let u be a lift of $\bar{u}$ in K: There exists $a \in K$ with $v_P(a) > v_P(y)$ such that $y + u^p \pi^{v_P(y)} = a$. Then, since $v_P(y) < v_P(y)/p < 0$, we have $v_P(y + (u\pi^{v_P(y)/p})^p - u\pi^{v_P(y)/p}) \geq \min\{v_P(a), v_P(y)/p\} > v_P(y)$ (note that $v_P(u) = 0$), and we can recurse. □

We also make use of the fact that the ramified places P in L/K (which appear in the support of $\mathfrak{m}$) are exactly those for which there exists a u_P as above such that $\lambda_P := -v_P(y + u_P^p - u_P)$ is positive and coprime to p; furthermore, the conductor $\mathfrak{f}_{L/K}$ verifies $v_P(\mathfrak{f}_{L/K}) = \lambda_P + 1$ (use [24, Proposition 3.7.8] and Proposition 4 below), so λ_P does not depend on y. Thus, while Lemma 2 helps us understand the ramification in L/K, if we want to explicitly compute L we need to find a Riemann-Roch space containing the generator x. With this in mind, we combine Lemma 2 with the strong approximation theorem to get a global result.

Lemma 3. *Let y be an element of K. For every place P of K, let u_P and λ_P be as above. Let S be the set of places P of K such that $\lambda_P > 0$, and let $S' := \{P \in \mathrm{Pl}_K : v_P(y) < 0\}$, so that $S \subseteq S'$. Fix an arbitrary place $P_0 \notin S'$, and let n_0 be a positive integer such that $D := n_0 P_0 - \sum_{P \in S'} 2P$ is nonspecial. Then there exists some u such that*

- $v_P(y + u^p - u) = -\lambda_P$ *for* $P \in S$,
- $v_P(y + u^p - u) \geq 0$ *for* $P \notin S \cup \{P_0\}$, *and*
- $v_{P_0}(y + u^p - u) \geq -pn_0$.

Proof. By the strong approximation theorem and its proof (see [24, Theorem 1.6.5]), there exists an element u in K such that $v_P(u - u_P) = 1$ for $P \in S'$, $v_P(u) \geq 0$ for $P \notin S' \cup \{P_0\}$, and $v_{P_0}(u) \geq -n_0$. We have

$$\begin{aligned} v := v_P(y + u^p - u) &= v_P(y + u_P^p - u_P + (u - u_P)^p + (u_P - u)) \\ &\geq \min\{v_P(y + u_P^p - u_P), pv_P(u - u_P), v_P(u_P - u)\}, \end{aligned}$$

which shows that $v = -\lambda_P$ if $P \in S$, and $v \geq 0$ if $P \in S' \setminus S$. In the same way,

$$\begin{aligned} v &= v_P(y + u_P^p - u_P + (u^p - u) - (u_P^p - u_P)) \\ &\geq \min\{v_P(y + u_P^p - u_P), v_P(u^p - u), v_P(u_P^p - u_P)\}, \end{aligned}$$

so we also have that $v \geq 0$ if $P \notin S' \cup \{P_0\}$, and $v \geq -pn_0$ if $P = P_0$ (note that $u_P = 0$ when $P \notin S'$). □

Thus we have that $x := y + u^p - u$ is an element of the Riemann-Roch space

$$\mathcal{L}\left(pn_0P_0 + \sum_S \lambda_P P\right) = \left\{f \in K : \operatorname{div}(f) \geq -pn_0P_0 - \sum_S \lambda_P P\right\}.$$

We now return to our hypothesis that L/K is a cyclic extension of degree p^m for some $m \geq 1$, with primitive element $\vec{x}$. Following [21], we study the vector $\lambda_P := -v_P(\vec{x}) := (-v_P(x_1), \ldots, -v_P(x_m))$. By adding elements of the form $\wp(0, \ldots, 0, x, 0, \ldots, 0)$ we can assume that there exist sets $S_i \subset \mathrm{Supp}(\mathfrak{m})$, places $P_{0,i}$ not in S_i, and positive integers $n_{0,i}$ such that x_i is in $\mathcal{L}(pn_{0,i}P_{0,i} + \sum_{S_i} \lambda_{P,i} P)$, where $\lambda_{P,i} := -v_P(x_i) > 0$ and $\gcd(\lambda_{P,i}, p) = 1$ for $P \in S_i$.

Setting $M_P := \max\{p^{m-i}\lambda_{P,i} : 1 \leq i \leq m\}$, we obtain $v_P(\mathfrak{f}_{L/K}) = M_P + 1$ from [21, p. 163]. Given that we already know a modulus $\mathfrak{m}$ such that $\mathfrak{f}_{L/K} \leq \mathfrak{m}$, we immediately get $\lambda_{P,i} \leq (v_P(\mathfrak{m}) - 1)p^{i-m}$. If $\mathfrak{m} = \sum_P n_P P$, then we set

$$D_i := pn_{0,i}P_{0,i} + \sum_{S_i} (n_P - 1)p^{i-m} P.$$

With these notations, we see that x_i is an element of $\mathcal{L}(D_i)$.

By induction, we assume that the x_i have been computed for $1 \leq i \leq m-1$ and explain how to find x_m. Set $M_m := K(\wp^{-1}(x_1, \ldots, x_{m-1}))$ and $D := D_m$; as remarked above, we can identify x_m as an element of the $\mathbb{F}_q$-vector space

$$\overline{\mathcal{L}_K}(D) = \mathcal{L}_K(D)/\wp(\mathcal{L}_K(D)).$$

Let d be the dimension of this space over $\mathbb{F}_p$. We compute an $\mathbb{F}_p$-basis of $\overline{\mathcal{L}_K}(D)$ and lift it to a set of d elements $\{f_1, \ldots, f_d\}$ of $\mathcal{L}_K(D)$. Hence x_m is an element of the subvector space of $\mathcal{L}_K(D)$ generated by the f_i, and we have

$$x_m = \sum_{i=1}^{d} a_i f_i$$

for some unknown elements a_i of $\mathbb{F}_p$. Next, we set

$$M := K(\wp^{-1}((x_1, \ldots, x_{m-1}, \mathcal{L}_K(D)))) = M_m(\wp^{-1}(0, \ldots, 0, \mathcal{L}_K(D))),$$

so that we have a tower $K \subset M_m \subset L \subset M$. Note that as in the Kummer case, neither M nor M_m is actually ever constructed. We will use the explicit action of the Frobenius automorphisms on Witt vectors of length m, so we identify

$(x_1, \ldots, x_{m-1})$ with $(x_1, \ldots, x_{m-1}, 0) \in \mathrm{W}_m(K)$ and f_i with $(0, \ldots, 0, f_i) \in \mathrm{W}_m(K)$. Let P be an unramified place of K; then the Frobenius automorphism $(P, L/K)$ acts on $\vec{y}$ via the formula

$$(P, L/K)(\vec{y}) = \vec{y} + \left\{\frac{\vec{x}}{P}\right\}$$

(see [21]), where the last term is in $\mathrm{W}_m(\mathbb{F}_p) \cong \mathbb{Z}_p \bmod p^m$ and satisfies

$$\left\{\frac{\vec{x}}{P}\right\} = \mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}\left(\vec{x} + \vec{x}^q + \cdots + \vec{x}^{\frac{\mathrm{N}(P)}{q}} \bmod P\right).$$

We now compute $\mathrm{Gal}(M/M_m)$. We have canonical isomorphisms

$$\mathrm{Gal}(M/M_m) \cong \prod_{i=1}^{d} \mathrm{Gal}(M_m(0, \ldots, 0, \wp^{-1}(f_i))/M_m) \cong (\mathbb{Z}/p\mathbb{Z})^d,$$

and this is made explicit via the Frobenius: Every $\mathrm{Gal}(M_m(0, \ldots, 0, \wp^{-1}(f_i))/M_m)$ is generated by the isomorphisms $(Q, M_m(0, \ldots, 0, \wp^{-1}(f_i))/M_m)$, where Q is a place of M_m. Because of the canonical isomorphism

$$\mathrm{Gal}(M_m(0, \ldots, 0, \wp^{-1}(f_i))/M_m) \cong \mathrm{Gal}(K(\wp^{-1}(f_i))/K),$$

the generating isomorphisms are of the form

$$y_i \mapsto y_i + \left\{\frac{f_i}{P}\right\},$$

where y_i is a primitive element of $K(\wp^{-1}(f_i))/K$ and P is the place of K below Q. Since the symbol $\{\cdot\}$ is additive (see [21]), we have

$$\mathrm{Gal}(K(\wp^{-1}(f_i))/K) \cong \left\langle \left\{\frac{f_i}{P}\right\} \right\rangle,$$

and so the isomorphism $\mathrm{Gal}(M/M_m) \cong (\mathbb{Z}/p\mathbb{Z})^d$ is made explicit via the map

$$(Q, M/M_m) \mapsto \left(\left\{\frac{f_1}{P}\right\}, \ldots, \left\{\frac{f_d}{P}\right\}\right).$$

We lift the terms in $\{\cdot\}$ from $\mathrm{W}_m(\mathbb{F}_p)$ to $\mathbb{Z}_p$, and if we can find enough places P_i such that the $\mathbb{Z}_p$-vectors

$$\left(\left\{\frac{f_1}{P_i}\right\}, \ldots, \left\{\frac{f_d}{P_i}\right\}\right)_i$$

form a matrix of rank d over $\mathbb{Z}_p$, then we are done, because by class field theory every element of $\mathrm{Gal}(M/M_m)$ is a Frobenius automorphism for some place Q. The generator is now obtained in exactly the same way as in the previous section for Kummer extensions — for which all that is necessary is an explicit Artin map.

4. An algorithm to find curves with many points

We now turn to the explicit applications of the theory described in the preceding sections, and switch between the language of curves and function fields when necessary. Our aim here is to find curves of low genus ($g \leq 50$) defined over a small finite field ($q \leq 100$) such that the number of rational points is the maximum possible; the current records can be found at www.manypoints.org. So we will only be interested in the abelian extensions L/K defined over the same finite field $\mathbb{F}_q$ such that the number of rational places of the field L is greater than or equal to the corresponding entry in the table[2] (as it was in June 2011). Furthermore, with the aid of the theory of Section 3, we will be able to find the equations of such extensions.

Proposition 4. *Let L/K be a cyclic extension of prime degree l of function fields defined over a finite field $\mathbb{F}_q$. Then the genus of L satisfies*

$$g_L = 1 + l(g_K - 1) + \tfrac{1}{2}(l-1)\deg \mathfrak{f}_{L/K}.$$

Proof. By the Riemann-Hurwitz genus formula, this comes down to showing that the degree of the different $\mathfrak{D}_{L/K}$ of L/K is $(l-1)\deg \mathfrak{f}_{L/K}$. Let Q be a place of L and let P be the place of K below Q. The extension being Galois, the inertia degree of P relatively to Q is independent of Q, so we denote it f_P. Let

$$\mathrm{N} = \mathrm{N}_{L/K} : \mathrm{Div}(L) \to \mathrm{Div}(K)$$

be the norm map defined by linearizing the formula $\mathrm{N}(Q) = f_P P$. From the general relation $\deg Q = f_P \deg P$, we note that $\deg \mathrm{N}(\mathfrak{D}_{L/K}) = \deg \mathfrak{D}_{L/K}$. By the conductor-discriminant formula, $\mathrm{N}(\mathfrak{D}(L_K))$ is equal to $\mathfrak{f}_{L/K}^{l-1}$, so by taking degrees we obtain the proposition. □

From Proposition 4, the genus of a cyclic extension of global function fields L/K of prime degree is exactly determined by its conductor $\mathfrak{f}_{L/K}$, or even simply by its degree. On the other hand, $\mathfrak{f}_{L/K}$ identifies L as the only field such that the Galois group of L/K is a quotient of the ray class group modulo $\mathfrak{f}_{L/K}$ by a certain subgroup of finite index. So, starting from a prime number l and a modulus $\mathfrak{m}$ defined over a global function field K with field of constants $\mathbb{F}_q$, one can enumerate all the cyclic extensions L of K of degree l and of conductor $\mathfrak{f}_{L/K}$ less than $\mathfrak{m}$ by computing all the subgroups of index l of $\mathrm{Pic}_{\mathfrak{m}}$. We also know in advance that the genus of these extensions will be less than

$$1 + l(g_K - 1) + \tfrac{1}{2}(l-1)\deg \mathfrak{m}.$$

[2]Note that L will be defined over $\mathbb{F}_q$ if at least one rational place of K splits totally in L, which will be the case when we are looking for L with many rational places.

Since l is a prime, all places which ramify have the same ramification type: Either they are all wildly ramified, or they are all tamely ramified. The following proposition thus describes what kind of $\mathfrak{m}$ one should test for a given l.

Proposition 5. *Let L/K be an abelian extension of function fields. Let P be a place of K. Then P is wildly ramified in L/K if and only if P appears in the conductor of L/K with multiplicity greater than* 2, *that is*,

$$P \text{ is wildly ramified if and only if } \mathfrak{f}_{L/K} \geq 2P.$$

Proof. From [16, Corollary 7.59], we see that a place P is tamely ramified if and only if the first ramification group in upper numbering is trivial, and from the local-global property of the conductor, this amounts to saying that P has weight one in $\mathfrak{f}_{L_K}$. So a place with weight at least two must be wildly ramified. □

We see that if l is prime to the characteristic p of K, then $\mathfrak{m}$ must be of the form

$$\mathfrak{m} = \sum_{i=1}^{n} P_i,$$

whereas if l equals p, then $\mathfrak{m}$ must be of the form

$$\mathfrak{m} = \sum_{i=1}^{n} m_i P_i,$$

where $m_i \geq 2$.

Because we want the greatest possible number of rational places for the field L, and because of the formula

$$N(L) = l|S| + r$$

(where S is the set of rational places of K which split in L and r is the number of rational places in the support of $\mathfrak{f}_{L/K}$), it seems reasonable to start from a field K which itself has many rational points compared to its genus. In this way, we will find curves with many points and their equations recursively: We start from the projective line or a maximal[3] elliptic curve, compute all of its "best" coverings reaching or improving a lower bound in www.manypoints.org, start the process again on these coverings, and so on. We summarize the process in Algorithm 1. Note that a reasonable restriction, especially when the size of the constant field increases, could be to take only conductors with places of degree 1 in their support.

[3]We call a curve of genus g defined over $\mathbb{F}_q$ *maximal* if no genus g curve defined over $\mathbb{F}_q$ has more rational points. This number of points is denoted $N_q(g)$.

Algorithm 1 (Good abelian coverings).

Input: A function field $K/\mathbb{F}_q$, a prime l, an integer G.
Output: The equations of all cyclic extensions of K of degree l and genus less than G whose number of $\mathbb{F}_q$-rational points improves the best known records.

1: Compute all the moduli of degree less than $B = (2G-2-l(2g(K)-2))/(l-1)$ using Proposition 5.
2: **for** each such modulus $\mathfrak{m}$ **do**
3: Compute the ray class group $\mathrm{Pic}_{\mathfrak{m}}$ modulo $\mathfrak{m}$.
4: Compute the set S of subgroups of $\mathrm{Pic}_{\mathfrak{m}}$ of index l and conductor $\mathfrak{m}$.
5: **for** every s in S **do**
6: Compute the genus g and the number of rational places n of the class field L of s.
7: **if** n is greater or equal to the known record for a genus g curve defined over $\mathbb{F}_q$ **then**
8: Update n as the new lower bound on $N_q(g)$.
9: Compute and output the equation of L.
10: **end if**
11: **end for**
12: **end for**

The complexity of the algorithm is linear in the number of fields (or pairs of divisors and subgroups) we need to consider. The total number of divisors of degree bounded by B is roughly $O(q^B)$ since this is the estimate for the number of irreducible polynomials of degree bounded by B. The number of subgroups to consider depends on the structure of the ray class group. For tamely ramified extensions, the group is the extension of the divisor class group by the product of the multiplicative groups of the divisors (modulo constants), so the number of cyclic factors depends on the number of places such that $l \mid q^{\deg P} - 1$. For wild extensions, the number of ramified places provides the same information. In the wild case, the number is bounded by $B/2$, so the total number of fields to investigate is roughly $O(q^B \cdot q^{B/2})$. For each pair we have to compute the genus and the number of rational places. The computation of the genus can be seen to run in time quartic in the number of (potentially) ramified places: For each place we need to check if it divides the conductor. This test is done by some $\mathbb{Z}$-HNF computation of a matrix whose dimension depends again on the total number of places. The computation of the number of rational places requires the computation of discrete logarithms in the divisor class group for every rational place of the base field. Assuming a small degree, this depends linearly on the number of ramified places.

Name	f
D_1	y^2+y+x^3+x
D_2	$y^2+(x^3+x+1)y+x^5+x^4+x^3+x$
D_3	$y^3+x^2y^2+(x^3+1)y+x^2+x$
D_4	$y^4+(x+1)y^2+(x^3+x)y+x^7+x^3$
D_5	$y^4+(x^2+x+1)y^2+(x^2+x)y+x^7+x^6+x^5+x^4$
D_6	$y^4+(x^6+x^5+x^4+1)y^2+(x^7+x^4+x^3+x^2)y+x^{11}+x^{10}+x^3+x^2$
D_7	$y^4+(x^7+x^6+x^4+x^2+1)y^2+(x^8+x^6+x^5+x^4)y+x^{10}+x^8+x^6+x^4$
D'_1	y^2+xy+x^3+x
D'_2	y^2+y+x^5+x
D'_3	$y^4+(x^2+x+1)y^2+(x^2+x)y+x^6+x^5$
D'_4	$y^4+xy^2+(x+1)y+x^5+x^4+x^3+x^2$
D'_5	$y^4+(x^3+1)y^2+(x^4+x^2)y+x^9+x^5$
D'_6	$y^4+(x^3+x+1)y^2+(x^3+x)y+x^9+x^8+x^5+x^4$
D'_7	$y^4+x^7y^2+(x^7+1)y+x^5+x$

Table 1. Equations $f=0$ for the base curves over $\mathbb{F}_2$ used in our calculations. The curves D_g have genus g and are maximal; the curves D'_g have genus g and satisfy $|D'_g(\mathbb{F}_2)| = N_2(g)-1$.

To summarize: The total complexity is essentially exponential in the genus bound, and is thus limited in scope.

Remark. It is possible to extend the algorithm to coverings of nonprime degrees, to include Artin-Schreier-Witt extensions for example, and this is what we have implemented in Magma. The genus and the conductor can then be computed using techniques from [8]. Note however that the computations then are much longer. This is the reason why we presented the algorithm only for cyclic extensions of prime degree: Since their arithmetic is simpler, the algorithm works best for them and can thus be used more efficiently over finite fields of size greater than 2 or 3.

5. Results

In this section we present the explicit results we obtained by implementing our algorithm. All of our computations were carried out in Magma [4], using a class field theory library implemented by the second author.

We restrict our attention here to the case where the base field is $\mathbb{F}_2$.

In Table 1 we give the equations for the base curves to which we applied our algorithm. The curves D_g have genus g and are maximal; the curves D'_g have genus g and satisfy $|D'_g(\mathbb{F}_2)| = N_q(g)-1$. Note that Rigato [19] has shown that the maximal curves of genus 1, 2, 3, 4, and 5 over $\mathbb{F}_2$ are unique.

g	N	Oesterlé bound	Base curve	Conductor $\mathfrak{f}$	Galois group G	$\lvert S\rvert$	$\lvert T\rvert$	$\lvert R\rvert$
14	16	16	D_4	$2P_7$	$\mathbb{Z}/2\mathbb{Z}$	16	0	0
17	18	18	D_2	$4P_1+6P_1$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$	16	2	0
24	23	23	D_4'	$2P_1+4P_1+2P_2$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$	20	1	2
29	26	27	D_4	$4P_1+8P_1$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/2\mathbb{Z}$	24	2	0
41	34	35	D_3'	$4P_1+4P_1$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/4\mathbb{Z}$	32	2	0
45	34	37	D_2	$4P_1+8P_1$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/4\mathbb{Z}$	32	2	0
46	35	38	D_3	$3P_1+8P_1$	$\mathbb{Z}/2\mathbb{Z}\oplus\mathbb{Z}/4\mathbb{Z}$	32	1	2

Table 2. New results over $\mathbb{F}_2$. For each genus g in the leftmost column, we give the largest number N for which we have constructed a genus-g curve over $\mathbb{F}_2$ having N rational points. The other columns are explained in the text.

Table 2 presents data on the curves we constructed that improved the previous records for the number of points on a genus-g curve over $\mathbb{F}_2$. The first two columns in the table give the genus g and the number of rational points N on the abelian coverings we construct. The third column gives the Oesterlé bound on the number of rational points of a genus-g curve defined over $\mathbb{F}_2$; in the cases we consider this is the best upper bound known. The fourth column gives the name (from Table 1) of the base curve used in the construction. The fifth column gives the conductor of the covering; a summand of the form $n_i P_i$ means that there is a place of degree i occurring in the conductor with weight n_i. The final four columns give the Galois group G of the covering, the number $|S|$ of totally split places, the number $|T|$ of totally ramified places, and the number $|R|$ of partially ramified places. In some cases we obtained the same values of g and N by applying our algorithm to different base curves; in these cases, we only make one entry in our table, corresponding to the construction using the base curve with the smallest genus. Finally, we mention that the average bound on the degree of the possible conductors we have tested was 14.

For each row of Table 2, let C_g denote the covering curve of genus g corresponding to that row. We present explicit equations for each C_g next; these are equations for the C_g as coverings of their base curves, so the equations for the base curves (given in Table 1) are left unstated here. We have attempted to present the equations so that the structure of each cover as a tower of Artin-Schreier covers is clear.

$$C_{14}: \begin{cases} 0 = (x^7+x^3+1)(z^2+z) \\ \qquad + y^3 + (x^4+x)y^2 + (x^4+x^2+1)y + (x^8+x^6+x^5+x^4) \end{cases}$$

$$C_{17}: \begin{cases} 0 = z^2 + x^2 z \\ \qquad + x(x+1)(x^3+x^2+1)y + x^2(x+1)^2(x^4+x^3+x^2+x+1) \\ 0 = w^2 + xw + x(x+1)(x^2+x+1)y + x^2(x+1) \end{cases}$$

$$C_{24}: \begin{cases} 0 = z^2 + x^2(x+1)z \\ \qquad + x(x^3+x^2+1)y^3 + x^3(x+1)^4y^2 + x^2(x^4+x^3+1)y \\ \qquad + x(x+1)(x^7+x^6+x^3+x^2+1) \\ 0 = w^2 + x^2w + x(x+1)y^3 \\ \qquad + x^3(x+1)^2y^2 + x^2(x+1)^2y + x(x+1)(x^2+x+1) \end{cases}$$

$$C_{29}: \begin{cases} 0 = z^2 + x^2(x+1)^4z \\ \qquad + (x+1)(x^6+x^5+x^4+x^3+1)y^3 \\ \qquad + x(x+1)^3(x^5+x^4+x^3+x^2+1)y^2 + (x+1)^2(x^6+x^2+1)y \\ \qquad + x^2(x+1)^3(x^5+x^4+x^3+x^2+1) \\ 0 = w^2 + x^2(x+1)^5w \\ \qquad + (x+1)(x^9+x^8+x^5+x^4+1)y^3 \\ \qquad + x(x+1)^3(x^9+x^8+x^6+x^5+x^4+x^2+1)y^2 \\ \qquad + (x+1)^2(x^9+x^8+x^3+x^2+1)y \\ \qquad + x^2(x+1)^3(x^{11}+x^9+x^8+x^6+x^5+x^4+x^3+x^2+1) \end{cases}$$

$$C_{41}: \begin{cases} 0 = z^2 + \big[x^6(x^2+x+1)y^3 + x^7(x^4+x^3+x^2+x+1)y^2 \\ \qquad + x^6(x+1)(x^2+x+1)y + x^{10}(x+1)^4\big]z \\ \qquad + x(x+1)^7(x^{13}+x^{12}+x^{11}+x^9+x^6+x^4+1)y^3 \\ \qquad + x^2(x+1)^3(x^{17}+x^{15}+x^{12}+x^{11}+x^9+x^3+1)y^2 \\ \qquad + x(x+1)^6(x^{17}+x^{15}+x^{14}+x^{13}+x^4+x^2+1)y \\ \qquad + x^5(x+1)^4(x^{17}+x^{16}+x^{12}+x^{11}+x^6+x^3+1) \\ 0 = v^2 + x^7v + xz \\ 0 = w^2 + x^2w + xy^2 + x^2y \end{cases}$$

$$C_{45}: \begin{cases} 0 = z^2 + (x+1)^2(xy+1)z + x^2(x^{13}+x^{11}+x^9+x+1)y \\ \qquad + x^9(x^8+x^6+x^4+x^3+x^2+x+1) \\ 0 = v^2 + (x+1)^2v + x(x+1)z + x^7(x^4+x+1) \\ 0 = w^2 + (x+1)^2w + (x+1)(x^5+x^2+x)y + (x+1)(x^8+x^5+x^4) \end{cases}$$

$$C_{46}: \begin{cases} 0 = z^2 + \big[(x+1)y^2 + (x^3+x^2+1)y + (x^4+x^3+x^2+x+1)\big]z \\ \qquad + (x+1)^2(x^{11}+x^8+x^6+x+1)y^2 + (x+1)^6(x^9+x^2+1)y \\ \qquad + x^7(x+1)^2(x^7+x^5+x^4+x^3+1) \\ 0 = v^2 + v + x(x+1)z + x^5(x+1) \\ 0 = w^2 + w + xy^2 + x^2(x^3+x^2+1)y \end{cases}$$

Remark. After this article was written, a preprint of Karl Rökaeus appeared in which he undertakes similar computations over the finite fields of size 2, 3, 4, and 5 (see [20]). Over $\mathbb{F}_2$ he recovers our genus-17 record, and he improves our genus-45 bound to 36 points. (He obtains the record-setting genus-45 curve as an abelian

cover of a genus-2 curve D with $|D(\mathbb{F}_2)| = N_2(2) - 2$.) In private communication, Rökaeus indicated that he also found a genus-46 curve over $\mathbb{F}_2$ with 36 points.

Remark. As mentioned above, we have restricted our search to curves over the field $\mathbb{F}_2$. However, our code works over other fields as well, and while we were testing it we found a curve of genus 11 over $\mathbb{F}_3$ with 21 rational points. This curve is a degree-2 cover of the genus-4 maximal curve defined by

$$C : y^4 - y^2 + x^6 + x^4 + x^2 = 0.$$

With notations as above, the conductor of the cover is of the form $P_1 + P_1 + P_1 + P_5$, and we have $|S| = 9$, $|R| = 3$, and $|T| = 0$. The resulting cover C' is given by the equation

$$\begin{aligned} z^2 = -(x^5+x^4+x^3-x^2+x+2)\cdot(y+x^2+x)\cdot(y^2+(-x+1)y+x^3-x^2-x+1) \\ \cdot\big((x^7+x^6+x^5-x^3-1)y^3+(-x^8+x^6+x^5-x^4-x^3-x)y^2 \\ +(-x^{10}-x^9-x^8+x^5+x^4+x^3-x^2+1)y \\ -x^{12}-x^9-x^8+x^6+x^4+x\big). \end{aligned}$$

Acknowledgments

The first author would like to thank his advisor David Kohel and Everett Howe for their support during the preparation of the paper, as well as Jérémie Detrey and Emmanuel Thomé for their help with Magma. Both authors thank the anonymous referees for their useful comments about a first version of the article.

References

[1] Roland Auer, *Ray class fields of global function with many rational places*, Ph.D. thesis, Carl-von-Ossietzky-Universität Oldenburg, 1999. http://oops.uni-oldenburg.de/volltexte/1999/457/

[2] ———, *Curves over finite fields with many rational points obtained by ray class field extensions*, in Bosma [3], 2000, pp. 127–134. MR 2002h:11053

[3] Wieb Bosma (ed.), *Algorithmic number theory*: *Proceedings of the* 4*th International Symposium* (*ANTS-IV*) *held at the Universiteit Leiden*, *Leiden*, *July* 2–7, 2000, Lecture Notes in Computer Science, no. 1838, Berlin, Springer, 2000. MR 2002d:11002

[4] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system*, *I*: *The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265. MR 1484478

[5] Henri Cohen (ed.), *Algorithmic number theory*: *Proceedings of the* 2*nd International Symposium* (*ANTS-II*) *held at the Université Bordeaux I*, *Talence*, *May* 18–23, 1996, Lecture Notes in Computer Science, no. 1122, Berlin, Springer, 1996. MR 97k:11001

[6] Claus Fieker, *Computing class fields via the Artin map*, Math. Comp. **70** (2001), no. 235, 1293–1303. MR 2002e:11153

[7] V. D. Goppa, *Codes that are associated with divisors*, Problemy Peredači Informacii **13** (1977), no. 1, 33–39. MR 58 #15672

[8] Florian Hess, Sebastian Pauli, and Michael E. Pohst, *Computing the multiplicative group of residue class rings*, Math. Comp. **72** (2003), no. 243, 1531–1548. MR 2004f:11126

[9] David Kohel and Robert Rolland (eds.), *Arithmetic*, *geometry*, *cryptography and coding theory* 2009: *Papers from the* 12*th Conference* (AGC^{2}T 12) *held in Marseille*, *March* 30–*April* 3, 2009, *the* 1*st Geocrypt Conference held in Pointe-à-Pitre*, *April* 27–*May* 1, 2009, *and the European Science Foundation Exploratory Workshop on Curves*, *Coding Theory and Cryptography held in Marseille*, *March* 25–29, 2009, Contemporary Mathematics, no. 521, American Mathematical Society, Providence, RI, 2010. MR 2011g:11003

[10] Serge Lang, *Algebraic number theory*, 2nd ed., Graduate Texts in Mathematics, no. 110, Springer, New York, 1994. MR 95f:11085

[11] ———, *Algebra*, 3rd ed., Graduate Texts in Mathematics, no. 211, Springer, New York, 2002. MR 2003e:00003

[12] Kristin Lauter, *Ray class field constructions of curves over finite fields with many rational points*, in Cohen [5], 1996, pp. 187–195. MR 98a:11076

[13] ———, *Deligne-Lusztig curves as ray class fields*, Manuscripta Math. **98** (1999), no. 1, 87–96. MR 2000a:11163

[14] ———, *A formula for constructing curves over finite fields with many rational points*, J. Number Theory **74** (1999), no. 1, 56–72. MR 99k:11088

[15] J. S. Milne, *Class field theory* (*version* 4.01), course notes, 2011. http://www.jmilne.org/math/CourseNotes/cft.html

[16] ———, *Algebraic number theory* (*version* 3.04), course notes, 2012. http://www.jmilne.org/math/CourseNotes/ant.html

[17] Jürgen Neukirch, *Algebraic number theory*, Grundlehren der mathematischen Wissenschaften, no. 322, Springer, Berlin, 1999. MR 2000m:11104

[18] Harald Niederreiter and Chaoping Xing, *Rational points on curves over finite fields*: *Theory and applications*, London Mathematical Society Lecture Note Series, no. 285, Cambridge University Press, 2001. MR 2002h:11055

[19] Alessandra Rigato, *Uniqueness of low genus optimal curves over* $\mathbb{F}_2$, in Kohel and Rolland [9], 2010, pp. 87–105. MR 2011m:11129

[20] Karl Rökaeus, *New curves with many points over small finite fields*, Tech. Report, 2012. arXiv 1204.4355 [math.NT]

[21] Hermann Ludwig Schmid, *Zur Arithmetik der zyklischen p-Körper*, J. Reine Angew. Math. **176** (1936), 161–167 (German). Zbl 0016.05205

[22] Jean-Pierre Serre, *Sur le nombre des points rationnels d'une courbe algébrique sur un corps fini*, C. R. Acad. Sci. Paris Sér. I Math. **296** (1983), no. 9, 397–402. http://gallica.bnf.fr/ark:/12148/bpt6k55351747/f35 MR 85b:14027

[23] ———, *Rational points on curves over finite fields*, unpublished notes by Fernando Q. Gouvéa of lectures at Harvard University, 1985.

[24] Henning Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics, no. 254, Springer, Berlin, 2009. MR 2010d:14034

[25] André Weil, *Basic number theory*, Grundlehren der mathematischen Wissenschaften, no. 144, Springer, Berlin, 1973. MR 96c:11002

VIRGILE DUCET: virgile.ducet@gmail.com
Institut de Mathématiques de Luminy, Campus de Luminy, Case 907, 13288 Marseille Cedex 9, France

CLAUS FIEKER: fieker@mathematik.uni-kl.de
Fachbereich Mathematik, Universität Kaiserslautern, Postfach 3049, 67653 Kaiserslautern, Germany

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.335

msp

Computing the unit group, class group, and compact representations in algebraic function fields

Kirsten Eisenträger and Sean Hallgren

Number fields and global function fields have many similar properties. Both have many applications to cryptography and coding theory, and the main computational problems for number fields, such as computing the ring of integers and computing the class group and the unit group, have analogues over function fields. The complexity of the number field problems has been studied extensively, and quantum computation has provided exponential speedups for some of these problems. In this paper we study the analogous problems in function fields. We show that there are efficient quantum algorithms for computing the unit group, for computing the class group, and for solving the principal ideal problem in function fields of arbitrary degree. We show that compact representations exist, which allows us to show that the principal ideal problem is in NP. We are also able to show that these compact representations can be computed efficiently, in contrast with the number field case.

1. Introduction

Algebraic number theory is concerned with the study of *number fields* — that is, finite extensions L of $\mathbb{Q}$ — and of the rings of algebraic integers $\mathbb{O}_L$ of such L. Similarly, we can consider finite algebraic extensions K of $\mathbb{F}_q(t)$, where $\mathbb{F}_q(t)$ is the quotient field of the polynomial ring $\mathbb{F}_q[t]$. These fields are called *function fields over finite fields* or *global function fields*. It was noticed early on that the integers have many properties in common with $\mathbb{F}_q[t]$, and similarly, that number fields and global function fields have many similar properties. Often, a problem that is posed for number fields admits an analogous problem for global function fields, and the other way around. For example, the Riemann hypothesis for the classical Riemann

MSC2010: primary 11Y16; secondary 11R27, 11R29.
Keywords: function fields, compact representations, infrastructure, unit group, principal ideal problem.

zeta function $\zeta(s)$ is still open, while the function-field analogue of this conjecture was proved by Weil.

The main computational problems for number fields include computing the ring of integers, the class group, and the unit group, and solving the principal ideal problem. These problems have been studied extensively, and there are a large number of classical algorithms for computing with number fields. Applications include the number field sieve, which is the fastest classical algorithm for factoring [29], and the Buchmann-Williams key-exchange system, whose security depends on the hardness of the principal ideal problem [7]. The recent push to make lattice-based cryptography more efficient has relied upon special lattices that come from number fields [33; 30]. Error-correcting codes have also been constructed using such lattices [23]. Quantum algorithms have been the source of exponential speedups for many of these computational problems for number fields. There are efficient quantum algorithms for computing the unit group and class group, and for solving the principal ideal problem in constant degree number fields [25; 38]. Some field extensions have also been computed using quantum algorithms [16]. In this paper we study the analogous computational problems over function fields.

Function fields also have many applications in cryptography and coding theory. There are many cryptographic applications that use elliptic curves or Jacobians of curves of small genus defined over finite fields [13]. Most of these rely on the assumption that the discrete log problem is difficult to solve in the underlying group associated with these curves. Another way to state this is that the discrete log problem is assumed to be hard in the divisor class group of the function field of the curve. Error correcting codes have also been based on function fields [22]. In a recent paper, Guruswami [24] constructed codes where everything was efficient except computing the basis for the Riemann-Roch space of a certain divisor.

For number fields the problems listed above have been studied extensively, and they appear to be computationally hard. For example, computing the ring of integers requires squarefree factorization of integers. The best known classical algorithms for computing the unit group, for computing the class group, and for solving the principal ideal problem are exponentially slower than factoring. On the other hand, computing the class group and unit group is in $\mathrm{NP} \cap \mathrm{coNP}$ for arbitrary degree number fields [42], while the quantum algorithms are only efficient for constant degree number fields. One apparent obstacle is that the only way known to compute with ideals of number fields requires a shortest vector problem in ideal lattices to be solved during computations, in order to keep representation sizes small.

In this paper we examine these computational problems over function fields of arbitrary degree. For function fields, computing the ring of integers is computationally equivalent to factoring polynomials over a finite field, which can be done in (classical) polynomial time, so one might hope that much more can be done.

In fact, even the analogue of the shortest vector problem has an efficient classical algorithm. But problems such as computing the divisor class group should be hard classically since they include as a special case the discrete log problem on an elliptic curve (a curve of genus one whose function field has degree two). For certain special classes of function fields (where the degree is two and the genus is large) there are subexponential algorithms for computing the class group, which make them less secure for cryptographic purposes: In [4] the authors give a subexponential algorithm for computing the class group of a hyperelliptic curve of large genus, and [31] gives a subexponential probabilistic algorithm for computing the class group of a real quadratic congruence function field of large genus. In [37] it is shown that various decision problems for quadratic congruence function fields of large genus are in NP $\cap$ coNP. There are also some exponential algorithms known for more general function fields. Another important computational problem that only exists in the function field case is that of computing Riemann-Roch spaces.

In this paper we show that the principal ideal problem over function fields of arbitrary degree is in NP. To do this we show that compact multiplicative representations exist for elements in function fields. This answers a question of Smart [40] and generalizes [36], which showed the existence of compact representations for real quadratic congruence function fields (which have degree two). Our work adapts work of Thiel, who used compact representations in number fields and showed that the principal ideal problem, the computation of class numbers, and the computation of compact representations of units are in NP $\cap$ coNP for number fields [42]. We also show that, unlike the situation for number fields, compact representations can be computed in (classical) polynomial time for arbitrary degree function fields. The standard representation of an element, for example a unit, may take exponentially many bits to represent. Compact representations give a certain factored form of the element which only requires polynomial representation size.

Given this setup, we also show that there are efficient quantum algorithms for computing the unit group and the class group, and for solving the principal ideal problem in arbitrary degree function fields. This is in contrast to the number field case, where currently only the constant degree case has quantum algorithms. These problems are solved by setting up abelian hidden subgroup problems.

One open question related to our work is whether the function field analogues of the problems treated by Thiel are also in NP $\cap$ coNP. Compact representations played a key role in the number field case. One issue in the function field case is that it is not known how to deterministically compute generators for the class group efficiently.

Another open question is finding an efficient quantum algorithm for computing class field towers of function fields. Certain towers of function fields — namely,

Hilbert class field towers — have applications to coding theory. When the tower is infinite one can construct asymptotically good sequences of codes from the fields in such towers [20, p. 212]. Infinite towers are known to exist [39], but for applications of such codes in practice, an explicit construction of the fields in the tower is required. Class groups of certain subrings of the function fields in the tower appear as the Galois groups of the field extensions in the tower. Therefore, computing the class groups (and compact representations), as we do in this paper, is required to compute such towers, as it is in the number field case [16].

In order to set up our algorithms we need efficient algorithms for doing computations in the infrastructure of a function field. Fontein recently provided these and we prove that his algorithms in [19; 17] are polynomial-time. To compute with the infrastructure it is necessary to efficiently compute the Riemann-Roch space of a divisor D. For this we use Hess's algorithm [26], which is a relatively simple, self-contained algorithm. In the appendix we include a complexity analysis of his algorithm. For other references that analyze Hess's algorithm see [19] (which makes some additional assumptions) and [14]. The algorithms above have been implemented, for example in Magma. The focus of this paper, however, is on the complexity analysis. Analyzing the Riemann-Roch algorithm addresses the missing piece for the codes in [24] to be efficient.

One technical challenge in our work is adapting Thiel's algorithm for computing compact representations [42] from the number field case to the function field case. To do this, and to end up with a polynomial-time algorithm, we must show that we can compute compact representations without searching for minima in a region of exponential size, something that is necessary in the number field case. We also analyze the Riemann-Roch space computation. This involves showing that we can efficiently compute certain prime ideals of the ring $\mathbb{O}_\infty$ (see Section 2 for notation) that we use to compute the Riemann-Roch space $L(D)$ from a given representation of the divisor D. We carry out this computation by factoring and computing radicals of certain ideals; further details, and the complexity analysis, can be found in Appendix B. Our algorithm generalizes the ideal factorization algorithm for number fields [16].

There have been other approaches to the study of some of these problems over function fields. In [27] Huang and Ierardi gave a construction of the Riemann-Roch space that is polynomial-time, assuming that all the singular points of the plane curve defining the function field are ordinary and defined over the base field. For another construction, which uses the Brill-Noether method, see Volcheck [43]. Recently, the authors learned that the (unpublished) Habilitation thesis by Diem [14] also studied Hess's algorithm. Kedlaya [28] showed how to compute zeta functions of curves with a quantum algorithm. His method requires computing the size of the divisor class group $\mathrm{Pic}^0(K)$, and he showed how to compute in the group efficiently.

Our work, by contrast, requires the different representation using the infrastructure of Fontein [17] in order to compute in the unit group and the class group, rather than only in the divisor class group $\mathrm{Pic}^0(K)$. The infrastructure also allows us to show the existence of compact representations.

Infrastructures have also been studied in [35; 18], which give quantum algorithms for computing one-dimensional infrastructures and the period lattice of infrastructures of fixed dimension.

2. Background on algebraic function fields and divisors

Algebraic function fields over finite fields. Let k be a finite field with $q = p^m$ elements for some prime p and integer $m > 0$. An *algebraic function field* K/k is an extension field $K \supseteq k$ such that K is a finite algebraic extension of $k(x)$ for some $x \in K$ which is transcendental over k. After replacing k with a finite extension, if necessary, we may assume that k is the *constant field* of K, that is, that k is algebraically closed in K. By [41, p. 144] such an algebraic function field is separably generated; that is, there exist $x, y \in K$ such that $K = k(x, y)$. The function field K is then specified by the finite field k, the indeterminate x and the minimal polynomial $f \in k(x)[T]$ of y over $k(x)$. Throughout the paper, we assume that K is given to us as $K = k(x, y)$ with x, y as above, and we let $d := [K : k(x)]$.

A *valuation ring* of the function field K/k is a ring $\widetilde{\mathbb{O}} \subseteq K$ such that $k \subsetneq \widetilde{\mathbb{O}} \subsetneq K$ and such that for every $z \in K$ we have $z \in \widetilde{\mathbb{O}}$ or $z^{-1} \in \widetilde{\mathbb{O}}$. A valuation ring is a local ring; that is, it has a unique maximal ideal [41, p. 2]. A *place* of a function field K/k is defined to be the maximal ideal of some valuation ring of K/k. To each place $\mathfrak{p}$ of K, there is an *associated discrete valuation* $v_{\mathfrak{p}} \colon K^* \to \mathbb{Z}$, and there is a one-to-one correspondence between places of K/k and discrete valuations of K/k [41, pp. 5–6]. Denote by $\mathcal{P}_K$ the set of all places of K. If $\mathfrak{p}$ is a place of K with corresponding valuation ring $\mathbb{O}_{\mathfrak{p}}$, we define the *degree of* $\mathfrak{p}$ to be the degree of the field extension of $\mathbb{O}_{\mathfrak{p}}/\mathfrak{p}$ over k; that is, $\deg \mathfrak{p} = [\mathbb{O}_{\mathfrak{p}}/\mathfrak{p} : k]$. If F/K is an extension of algebraic function fields we say that a place $\mathfrak{P} \in \mathcal{P}_F$ *lies above* a place $\mathfrak{p} \in \mathcal{P}_K$ if $\mathfrak{p} \subseteq \mathfrak{P}$.

For the rational function field $k(x)$, the places are completely understood: The places of $k(x)$ correspond to the irreducible polynomials of $k[x]$, together with a "place at infinity", denoted ∞.

Let v_∞ be the discrete valuation corresponding to the infinite place ∞ of the rational function field $k(x)$. Then v_∞ is defined via $v_\infty(f/g) = \deg g - \deg f$, for $f, g \in k[x]$. Let $\mathfrak{o}_\infty := \{a \in k(x) : v_\infty(a) \geq 0\}$. Then $\mathfrak{o}_\infty$ is the valuation ring associated to v_∞ and the unique maximal ideal of $\mathfrak{o}_\infty$ is generated by $1/x$. Let S denote the set of places of K above ∞. Let

$$\mathbb{O}_\infty := \{a \in K : v_{\mathfrak{p}}(a) \geq 0 \text{ for all } \mathfrak{p} \in S\}.$$

Then $\mathbb{O}_\infty$ is the integral closure of $\mathfrak{o}_\infty$ in K, and $\mathbb{O}_\infty$ is a free $\mathfrak{o}_\infty$-module of rank d. The ring $\mathbb{O}_\infty$ is a principal ideal domain whose prime ideals correspond to the elements in S.

Divisors on algebraic function fields. A *divisor* on K is a formal sum

$$D = \sum_{\mathfrak{p} \in \mathcal{P}_K} n_\mathfrak{p}\, \mathfrak{p}$$

such that $n_\mathfrak{p} = 0$ for all but finitely many $\mathfrak{p}$. Let $\mathrm{Div}(K)$ denote the group of divisors on K. For a divisor D which is given as $D = \sum_{\mathfrak{p} \in \mathcal{P}_K} n_\mathfrak{p}\, \mathfrak{p}$, we define the *degree* of D to be $\deg D = \sum_{\mathfrak{p} \in \mathcal{P}_K} n_\mathfrak{p} \deg \mathfrak{p}$. The divisors of degree zero form a subgroup of $\mathrm{Div}(K)$, which we denote by $\mathrm{Div}^0(K)$. For $f \in K^*$, the *divisor* of f is defined to be

$$\mathrm{div}(f) = \sum_{\mathfrak{p} \in \mathcal{P}_K} v_\mathfrak{p}(f)\, \mathfrak{p}.$$

The set of all divisors of the form $\mathrm{div}(f)$ forms the group $\mathrm{Prin}(K)$ of *principal divisors* on K. Note that if D is a principal divisor then $\deg D = 0$. We define the *divisor class group* $\mathrm{Pic}^0(K)$ to be the quotient of the group of divisors of degree zero by the group of principal divisors; that is, $\mathrm{Pic}^0(K) = \mathrm{Div}^0(K)/\mathrm{Prin}(K)$. The divisor class group is a finite group.

A divisor $D = \sum_\mathfrak{p} n_\mathfrak{p}\, \mathfrak{p}$ is *effective* if $n_\mathfrak{p} \geq 0$ for all $\mathfrak{p}$; we write $D_1 \geq D_2$ to mean that $D_1 - D_2$ is effective. Every divisor D can be written uniquely as $D = D_+ - D_-$ with D^+, D^- effective divisors with disjoint support. We define the *height* of a divisor D to be $\mathrm{ht}(D) := \max\{\deg(D_+), \deg(D_-)\}$. For a divisor $D \in \mathrm{Div}(K)$ we define the *Riemann-Roch space* of D to be the set

$$L(D) := \{f \in K : \mathrm{div}(f) + D \geq 0\} \cup \{0\}.$$

The set $L(D)$ is a vector space over k, and we denote its dimension by $\ell(D)$.

Fractional ideals. Let $\mathbb{O}$ be the integral closure of $k[x]$ in K. Then $\mathbb{O}$ is a free $k[x]$-module of rank d. By [11, Theorem 1], a $k[x]$-basis for $\mathbb{O}$ can be computed in time polynomial in d and $\log q$. If $S = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_{n+1}\}$ is the set of places above the infinite place ∞ of $k(x)$, then we also have

$$\mathbb{O} = \{a \in K : v_\mathfrak{p}(a) \geq 0 \text{ for all } \mathfrak{p} \notin S\}.$$

Note that for any nonempty finite set S of places of K one can find an $x \in K$ such that S is the set of infinite places above x. Throughout the paper we assume that $\deg \mathfrak{p}_{n+1} = 1$. This can always be achieved by passing to a finite extension of the constant field k.

A *fractional ideal* of $\mathbb{O}$ is a finitely generated $\mathbb{O}$-submodule of K. Since $\mathbb{O}$ is a Dedekind domain, the nonzero fractional ideals $\mathrm{Id}(\mathbb{O})$ of $\mathbb{O}$ form a (free) abelian

group under multiplication. There is a natural homomorphism $\phi\colon \mathrm{Div}(K) \to \mathrm{Id}(\mathbb{O})$ defined by

$$\sum n_{\mathfrak{p}}\mathfrak{p} \mapsto \prod_{\mathfrak{p}\notin S} (\mathfrak{p}\cap\mathbb{O})^{-n_{\mathfrak{p}}}.$$

This map has a right inverse, namely the map $\mathrm{div}\colon \mathrm{Id}(\mathbb{O}) \to \mathrm{Div}(K)$ that sends a fractional ideal $B = \prod_{\mathfrak{p}\notin S}(\mathfrak{p}\cap\mathbb{O})^{n_{\mathfrak{p}}}$ to $\mathrm{div}(B) := -\sum_{\mathfrak{p}\notin S} n_{\mathfrak{p}}\mathfrak{p}$. Hence each divisor can be represented by a pair $\left(A, \sum t_i\mathfrak{p}_i\right)$, where A is a fractional ideal of $\mathbb{O}$ and $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_{n+1}\}$ are the places in S, that is, the primes above ∞. This is how we will represent divisors throughout the paper.

The *class group* $\mathrm{Cl}(\mathbb{O})$ of $\mathbb{O}$ is defined to be the group of fractional ideals of $\mathbb{O}$ modulo the principal fractional ideals of $\mathbb{O}$. The class group is a finite abelian group, and the map $\phi\colon \mathrm{Div}(K) \to \mathrm{Id}(\mathbb{O})$ extends to a homomorphism

$$\begin{aligned} \phi\colon \mathrm{Pic}^0(K) &\longrightarrow \mathrm{Cl}(\mathbb{O}) \\ \Big[\sum n_{\mathfrak{p}}\mathfrak{p}\Big] &\longmapsto \Big[\prod_{\mathfrak{p}\notin S} (\mathfrak{p}\cap\mathbb{O})^{-n_{\mathfrak{p}}}\Big]. \end{aligned}$$

When $\deg \mathfrak{p}_{n+1} = 1$ this map fits into an exact sequence

$$0 \longrightarrow \mathrm{Ker} \longrightarrow \mathrm{Pic}^0(K) \xrightarrow{\phi} \mathrm{Cl}(\mathbb{O}) \longrightarrow 1.$$

Here Ker is the subgroup of $\mathrm{Pic}^0(K)$ that is generated by all degree-zero divisors with support in S, so the map $\mathrm{Ker} \to \mathrm{Pic}^0(K)$ is just the inclusion map. Since k is a finite field, Ker is finite by [34, Proposition 14.2, p. 243].

3. Computing efficiently in the unit group

In this section we show how to efficiently compute classically in the unit group of $\mathbb{O}$. Recall that $S = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_{n+1}\}$ are the places of K above ∞ and that

$$\mathbb{O} = \{a \in K : v_{\mathfrak{p}}(a) \geq 0 \text{ for all } \mathfrak{p} \notin S\}.$$

Also, we assume that $\mathfrak{p}_{n+1}$ is a place of degree 1.

To compute in the unit group, consider the map $\mathrm{val}_\infty : K^* \to \mathbb{Z}^n$ given by $\mathrm{val}_\infty(a) = (-v_{\mathfrak{p}_1}(a), \ldots, -v_{\mathfrak{p}_n}(a))$. The image of $\mathbb{O}^*$ under val_∞ is a lattice Λ in $\mathbb{Z}^n$. By an analogue of Dirichlet's Unit Theorem for function fields, the *unit rank* — that is, the rank of Λ — is equal to $n = \#S - 1$. Since units can have exponentially many bits in the standard representation, computing the unit group means to compute a basis of that lattice, or to compute compact representations for a fundamental set of units as in Definition 4.3. In Lemma 4.7 we show that the compact representation of an element can be computed from its valuation vector, so it follows that these two problems are polynomial time equivalent in function fields.

Fontein [17] showed that it is possible to compute in a finite abelian group which he denotes $\mathrm{Rep}^{f*}(\mathbb{O})$ and which is isomorphic to $\mathbb{Z}^n/\Lambda$. We discuss his approach in the next section. We then show that these computations are efficient. From the group structure of $\mathbb{Z}^n/\Lambda$ we can obtain the basis for the lattice Λ.

3A. ***Minima and reduced ideals in function fields.*** We now give the definitions of minima and reduced ideals and define $\mathrm{Rep}^{f*}(\mathbb{O})$ (see [17]). In the following, by an ideal of $\mathbb{O}$ we will always mean a *fractional* ideal of $\mathbb{O}$.

For each place $\mathfrak{p}_i \in S$, with its associated discrete valuation $v_{\mathfrak{p}_i}$, there is a corresponding *absolute value* $|\alpha|_i$, defined by

$$|\alpha|_i := q^{-v_{\mathfrak{p}_i}(\alpha)\deg \mathfrak{p}_i}.$$

For an ideal A and integers $t_1, \ldots, t_{n+1} \in \mathbb{Z}$ we define

$$B(A, (t_1, \ldots, t_{n+1})) := \{\alpha \in A : |\alpha|_i \le q^{t_i \deg \mathfrak{p}_i} \text{ for } i = 1, \ldots, n+1\}.$$

This is a Riemann-Roch space; we have

$$B(A, (t_1, \ldots, t_{n+1})) = L\Big(\mathrm{div}(A) + \sum_{i=1}^{n+1} t_i \mathfrak{p}_i\Big).$$

For an ideal A and $\alpha \in K^*$, let $B(A, \alpha) := B(A, (-v_{\mathfrak{p}_1}(\alpha), \ldots, -v_{\mathfrak{p}_{n+1}}(\alpha)))$.

Definition 3.1 (Minima and reduced ideals).

(1) Let A be an ideal of $\mathbb{O}$ and let μ be a nonzero element of A. The element μ is a *minimum* of A if for every nonzero $\alpha \in B(A, \mu)$ we have $|\alpha|_i = |\mu|_i$ for $i = 1, \ldots, n+1$.

(2) An ideal A is *reduced* if 1 is a minimum of A.

Denote by $\mathrm{Red}(A)$ the set of reduced ideals of $\mathbb{O}$ which are in the same ideal class as A in $\mathrm{Cl}(\mathbb{O})$. There is a close connection between the set of minima of an ideal A and the set of reduced ideals equivalent to A. First, if μ is a minimum of A and $\epsilon \in \mathbb{O}^*$, then $\epsilon\mu$ is also a minimum of A. This action of $\mathbb{O}^*$ on the set of minima gives rise to a bijection

$$\begin{aligned} \{\text{minima of } A\}/\mathbb{O}^* &\longrightarrow \mathrm{Red}(A) \\ \mu\mathbb{O}^* &\longmapsto (1/\mu)A. \end{aligned}$$

So every element of $\mathrm{Red}(A)$ is of the form $(1/\mu)A$ with μ a minimum of A. Next define a map from the set of reduced ideals equivalent to A to $\mathbb{Z}^n/\Lambda$ by defining

$$\begin{aligned} d\colon \mathrm{Red}(A) &\longrightarrow \mathbb{Z}^n/\Lambda \\ (1/\mu)A &\longmapsto \mathrm{val}_\infty(\mu) + \Lambda \end{aligned}$$

This map is well-defined since $\deg \mathfrak{p}_{n+1} = 1$ (see [17, Corollary 5.3]), and it is also injective [17, Proposition 5.5]. Now we can define Fontein's group $\mathrm{Rep}^{f*}(\mathbb{O})$, which is isomorphic to $\mathbb{Z}^n/\Lambda$.

Definition 3.2. Let A be an ideal of $\mathbb{O}$. An *$f*$-representation* is a tuple

$$(I, (t_1, \ldots, t_n)) \in \mathrm{Red}(A) \times \mathbb{Z}^n$$

such that $B(I, (t_1, \ldots, t_n, 0)) = k$. Denote the set of all $f*$-representations in $\mathrm{Red}(A) \times \mathbb{Z}^n$ by $\mathrm{Rep}^{f*}(A)$.

When A and B are two ideals that are in the same ideal class in $\mathrm{Cl}(\mathbb{O})$, then clearly $\mathrm{Rep}^{f*}(A) = \mathrm{Rep}^{f*}(B)$. Let

$$\Phi_A \colon \mathrm{Rep}^{f*}(A) \to \mathbb{Z}^n/\Lambda$$

be defined by

$$\Phi_A\left((1/\mu)A, t\right) = \mathrm{val}_\infty(\mu) + t + \Lambda.$$

Here $t = (t_1, \ldots, t_n) \in \mathbb{Z}^n$. In [17, Theorem 6.8] it is proved that this map is a bijection. In particular, $\mathrm{Rep}^{f*}(\mathbb{O})$ is isomorphic to $\mathbb{Z}^n/\Lambda$. So to each element (I, t) of $\mathrm{Rep}^{f*}(A)$, there is an associated point in $\mathbb{Z}^n/\Lambda$, and if $I = (1/\mu)A$, we say that (I, t) *represents* the element $\mathrm{val}_\infty(\mu) + t + \Lambda$ of $\mathbb{Z}^n/\Lambda$. Let $[A]$ be the set of ideals equivalent to A in the class group. It is possible to extend Φ_A to a well-defined (but no longer injective) map $\Phi_A \colon [A] \times \mathbb{Z}^n \to \mathbb{Z}^n/\Lambda$ by letting $\Phi_A\left((1/\alpha)A, f\right) = \mathrm{val}_\infty(\alpha) + f + \Lambda$.

In [17, Proposition 8.1] the following is shown:

Proposition 3.3. *Let $(A, (t_1, \ldots, t_n))$ be an element of $\mathrm{Rep}^{f*}(B)$ for some ideal B. Then $\mathrm{div}(A) \geq 0$ and $t_i \geq 0$ for $1 \leq i \leq n$. Moreover,*

$$0 \leq \deg \mathrm{div}(A) + \sum_{i=1}^{n} t_i \deg \mathfrak{p}_i \leq g.$$

Here g denotes the genus of the function field.

We want to compute a basis for the n-dimensional lattice Λ. Since $\mathbb{Z}^n/\Lambda$ is isomorphic to $\mathrm{Rep}^{f*}(\mathbb{O})$, it is enough to obtain generators and relations for the finite group $\mathrm{Rep}^{f*}(\mathbb{O})$.

3B. *Reduction and obtaining generators for* $\mathbf{Rep}^{f*}(\mathbb{O})$.

Let

$$\Phi := \Phi_{\mathbb{O}} \colon \mathrm{Rep}^{f*}(\mathbb{O}) \to \mathbb{Z}^n/\Lambda$$

and its extension to $[\mathbb{O}] \times \mathbb{Z}^n \to \mathbb{Z}^n/\Lambda$ be the maps defined above. The group $\mathbb{Z}^n/\Lambda$ is generated by the standard basis vectors e_i $(1 \leq i \leq n)$, so in order to find

generators for $\mathrm{Rep}^{f*}(\mathbb{O})$ we need to find elements $((1/\mu_i)\mathbb{O}, f_i)$ such that

$$\Phi((1/\mu_i)\mathbb{O}, f_i) = e_i + \Lambda.$$

To obtain such elements we consider the elements $(\mathbb{O}, e_i)$, for $i = 1, \ldots, n$. These elements are not in $\mathrm{Rep}^{f*}(\mathbb{O})$, but they do have the property that $\Phi(\mathbb{O}, e_i) = e_i + \Lambda$. So to obtain the right elements in $\mathrm{Rep}^{f*}(\mathbb{O})$ we reduce the elements $(\mathbb{O}, e_i)$ to elements $((1/\mu)\mathbb{O}, f_i) \in \mathrm{Rep}^{f*}(\mathbb{O})$ with Algorithm 3.4 below, and use the fact that under Φ, the element $(\mathbb{O}, e_i)$ and its reduction have the same image (see Remark 3.6 below).

The general reduction algorithm that we are describing next works for $\mathrm{Rep}^{f*}(I)$ for any ideal I of $\mathbb{O}$.

Algorithm 3.4 (Reduce).

Input: An ideal A and a vector $t = (t_1, \ldots, t_n) \in \mathbb{Z}^n$.

Output: A minimum μ of A, the reduced ideal $(1/\mu)A$, and a vector $t - \mathrm{val}_\infty(\mu)$ such that $((1/\mu)A, t - \mathrm{val}_\infty(\mu)) \in \mathrm{Rep}^{f*}(A)$.

1. Find the minimum ℓ in the interval
$$\Big[-\deg \mathrm{div}(A) - \sum_{i=1}^{n} t_i \deg \mathfrak{p}_i,\ \ g - \deg \mathrm{div}(A) - \sum_{i=1}^{n} t_i \deg \mathfrak{p}_i\Big]$$
such that $\dim B(A, (t_1, \ldots, t_n, \ell)) > 0$.
2. Set $u_1, \ldots, u_n = 0$. For each $1 \le i \le n$, increase u_i to find the largest value u_i with $\dim B(A, (t_1 - u_1, \ldots, t_n - u_n, \ell)) > 0$.
3. Let μ be a nonzero element of $B(A, (t_1 - u_1, \ldots, t_n - u_n, \ell))$.
Output $(\mu, (1/\mu)A, (u_1, \ldots, u_n))$.

Proposition 3.5. *Algorithm 3.4 is correct and returns* $(\mu, (1/\mu)A, (u_1, \ldots, u_n))$ *in time polynomial in* $d, \log q, \mathrm{ht}(\mathrm{div}(A))$ *and* $\|t\|_\infty$.

Proof. Let ℓ be minimal such that $\dim B(A, (t_1, \ldots, t_n, \ell)) > 0$. By [19, Theorem 4.4.3], we have

$$\ell \in \Big[-\deg \mathrm{div}(A) - \sum_{i=1}^{n} t_i \deg \mathfrak{p}_i,\ \ g - \deg \mathrm{div}(A) - \sum_{i=1}^{n} t_i \deg \mathfrak{p}_i\Big],$$

so the first step of the algorithm requires at most g Riemann-Roch computations. By Theorem B.9, each of these computations

$$B(A, (t_1, \ldots, t_n, \ell)) = L\Big(\mathrm{div}(A) + \sum_{i=1}^{n} t_i \cdot \mathfrak{p}_i + \ell \cdot \mathfrak{p}_{n+1}\Big)$$

can be performed in time polynomial in d, $\log q$, $\mathrm{ht}(\mathrm{div}(A))$, and $\|t\|_\infty$, because ℓ is at most a polynomial in g, $\mathrm{div}(A)$, and $\|t\|_\infty$, and g is a polynomial in d.

The second step computes the valuation that μ has in the third step. For coordinate i, there are at most t_i Riemann-Roch computations, so in total there are at most $n \max |t_i|$, which is polynomial in d and $\|t\|_\infty$ since $n \le d$. The correctness of steps 2 and 3 follows from the correctness proof of Algorithm 5.4.2 in [19]. □

Remark 3.6. Let A be an ideal of $\mathbb{O}$ and let $t=(t_1,\dots,t_n)\in\mathbb{Z}^n$. Then $(A,(t_1,\dots,t_n))$ represents the same point in $\mathbb{Z}^n/\Lambda$ as its reduction

$$((1/\mu)A, t - \mathrm{val}_\infty(\mu)) \in \mathrm{Rep}^{f*}(A),$$

because

$$\begin{aligned}\Phi_A(A,t) &= t+\Lambda \\ &= \mathrm{val}_\infty(\mu) + (t - \mathrm{val}_\infty(\mu)) + \Lambda \\ &= \Phi_A((1/\mu)A, t - \mathrm{val}_\infty(\mu)).\end{aligned}$$

Denote by $\mathrm{Reduce}(A,e)$ the element of $\mathrm{Rep}^{f*}(A)$ computed by Algorithm 3.4. By the above discussion we have $\Phi_A(\mathrm{Reduce}(A,e)) = e+\Lambda$, and if $e' = e+v$ with $v\in\Lambda$, then $\Phi_A(\mathrm{Reduce}(A,e')) = e'+\Lambda = e+\Lambda$. Since $\Phi_A\colon \mathrm{Rep}^{f*}(A)\to\mathbb{Z}^n/\Lambda$ is injective this implies that $\mathrm{Reduce}(A,e) = \mathrm{Reduce}(A,e')$ whenever $e-e'\in\Lambda$.

Definition 3.7. When $\alpha\in K$, the norm of α can be expressed uniquely as $N(\alpha) = f/h$, where f and h are coprime elements of $k[x]$ and h is monic. We define $\mathrm{dg}(N(\alpha))$ to be $\mathrm{dg}(N(\alpha)) = \max\{\deg f, \deg h\}$.

Remark 3.8. When $A=\alpha\mathbb{O}$ then being polynomial in $\mathrm{ht}(\mathrm{div}(A))$ is the same as being polynomial in $\mathrm{dg}\,N(\alpha)$ (see [17, p. 28]).

3C. ***Composition and computing inverses in*** $\mathbf{Rep}^{f*}(\mathbb{O})$ ***and bounding the representation size of elements.*** By [17, Proposition 8.2], elements in $\mathrm{Rep}^{f*}(\mathbb{O})$ can be represented by $O(d^2 g\log q)$ bits. Here g denotes the genus of the function field, which is of size polynomial in d.

Composition of two elements $(A, f), (A', f')$ of $\mathrm{Rep}^{f*}(\mathbb{O})$ is done by multiplying the ideals, adding the two vectors, and then applying Algorithm 3.4 to $(AA', f+f')$. To compute the inverse of $(A, f_1,\dots,f_n)$, compute the inverse A^{-1}, find ℓ minimal such that $\dim B(A^{-1}, (-f_1,\dots,-f_n,\ell)) > 0$ and then reduce using Algorithm 3.4 [19, Proposition 4.3.4]. The ideal arithmetic in $\mathbb{O}$ is polynomial in $\log q$, d, and $\mathrm{ht}(\mathrm{div}(A))$, $\mathrm{ht}(\mathrm{div}(A'))$ [14, Proposition 2.66, and Proposition 2.69(b)] and $\mathrm{ht}(\mathrm{div}(A))$ is of size polynomial in d and $\log q$ when $(A, f)\in\mathrm{Rep}^{f*}(\mathbb{O})$. Hence Proposition 3.5 implies that composition of two elements and computing inverses in $\mathrm{Rep}^{f*}(\mathbb{O})$ are both polynomial in $\log q$ and d.

4. Compact representations in global function fields

In this section we show how to efficiently compute compact representations of elements $\alpha \in K$ classically. This allows us to show that the principal ideal problem is in NP, and to compute compact representations of units. We adapt the definitions and approach for number fields given in [42, p. 82] to the function field case. The sizes are adapted to match the parameters that are appropriate for number fields and that come from our algorithms. In the function field case we show that an exponential search for minima is no longer necessary.

Definition 4.1. For $\alpha \in K$ and $s \in \mathbb{Q}^n$ we say that α *is close to* s if

$$\|\mathrm{val}_\infty(\alpha) - s\|_1 \leq n + g,$$

where g is the genus of K.

Definition 4.2. A *multiplicative representation* of an element $\alpha \in K$ is a pair

$$M = ((\beta_1, \ldots, \beta_\ell), (e_1, \ldots, e_\ell)),$$

where $\beta_i \in K$, $e_i \in \mathbb{Z}$, $\ell \in \mathbb{N}$, and such that $\alpha = \prod_{i=1}^{\ell} \beta_i^{e_i}$.

A *binary multiplicative representation* (*BMR*) of an element $\alpha \in K$ is a multiplicative representation such that for all $i \leq \ell$ we have both that $e_i = 2^{\ell - i}$ and that $((\beta_1, \ldots, \beta_i), (e_1, \ldots, e_i))$ is a minimum of $\mathbb{O}$. Since the exponents e_i are determined, a BMR can be represented as $(\beta_1, \ldots, \beta_k)$.

Definition 4.3. A *compact representation* of $\alpha \in K$ is a pair $B = (\gamma, (\beta_1, \ldots, \beta_\ell))$, where $(\beta_1, \ldots, \beta_\ell)$ is a BMR for a minimum β of $\mathbb{O}$ with $\gamma = \alpha\beta$, and where

$$\begin{aligned} \ell &\leq \log(\|\mathrm{val}_\infty(\alpha)\|_\infty + g), \\ \mathrm{size}(\gamma) &\leq \mathrm{poly}(\log q, d, \deg N(\alpha)), \text{ and} \\ \mathrm{size}(\beta_i) &\leq \mathrm{poly}(\log q, d). \end{aligned}$$

Here size denotes the number of bits required to represent the element.

Remark 4.4. Definition 4.3 depends on certain implied constants hidden in expressions like $\mathrm{poly}(\log q, d)$. What is meant is that there exist specific polynomials that can be used in the definition so that all subsequent statements in this paper hold.

The bound on ℓ is chosen to handle the length of the generator after reducing $\alpha\mathbb{O}$, which is $\mathrm{val}_\infty(\gamma/\alpha)$. The factor γ comes from ideal reduction, so γ has size polynomial in d, $\log q$, and $\deg N(\alpha)$.

Claim 4.5. *Given a BMR* $(\beta_1, \ldots, \beta_\ell)$ *of a minimum* β *of* $\mathbb{O}$, *the ideal* $(1/\beta)\mathbb{O}$ *can be efficiently computed.*

Proof. At the first step, the ideal $(1/\beta_1)\mathbb{O}$, which is reduced by the definition of a BMR, can be efficiently computed. In general, let $\beta_i' = \prod_{j=1}^{i} \beta_j^{2^{i-j}}$. By the definition of a BMR, β_i' is a minimum of $\mathbb{O}$ for all i. Given the reduced ideal $(1/\beta_i')\mathbb{O}$, the reduced ideal $(1/\beta_{i+1}')\mathbb{O} = (1/(\beta_{i+1}\beta_i'^2))\mathbb{O}$ can be efficiently computed by squaring $(1/\beta_i')\mathbb{O}$ and multiplying by $1/\beta_{i+1}$. □

Our next algorithm produces a compact representation of a generator of an ideal. It calls Close (Algorithm 4.8), which calls Double (Algorithm 4.10); we will postpone the description of these algorithms, and the proofs of their correctness, until after the proof that Algorithm 4.6 is correct.

Algorithm 4.6 (Compact representation).

Input: A vector $v \in \mathbb{Z}^n$ and an ideal A such that $v = \mathrm{val}_\infty(\alpha)$ and $A = \alpha\mathbb{O}$ for some $\alpha \in K$.

Output: A compact representation of such an α.

1. Call Reduce$(A, 0)$ to get a reduced ideal $(1/\gamma)A$ and element $\gamma \in K$.
2. Let $(\beta_1, \ldots, \beta_\ell) = \mathrm{Close}(\mathbb{O}, \mathrm{val}_\infty(\gamma/\alpha))$.
3. Output $(\gamma, (\beta_1, \ldots, \beta_\ell))$.

Lemma 4.7. *Algorithm 4.6 returns a compact representation of $\alpha \in K$ in time polynomial in* $\log q$, d, $\mathrm{dg}\, N(\alpha)$, *and* $\log(\|\mathrm{val}_\infty(\alpha)\|_\infty)$.

Proof. Proposition 3.5 and Remark 3.8 show that the element γ in Step 1 can be computed with Algorithm 3.4 in time polynomial in d, $\log q$, and $\mathrm{dg}\, N(\alpha)$. Therefore the size of γ is bounded by the same amount. Also, γ is a minimum of $A = \alpha\mathbb{O}$, so $\beta := \gamma/\alpha$ is a minimum of $\mathbb{O}$. By Corollary 4.13 below, $\mathrm{Close}(\mathbb{O}, \mathrm{val}_\infty(\gamma/\alpha))$ returns the BMR $(\beta_1, \ldots, \beta_\ell)$ of the minimum $\beta = \gamma/\alpha$ of $\mathbb{O}$ (and not just the BMR of a minimum close to γ/α). Hence the algorithm computes the compact representation $(\gamma, (\beta_1, \ldots, \beta_\ell))$ of $\alpha = \gamma/\beta$.

We have already noted that Step 1 takes time polynomial in d, $\log q$, and $\mathrm{dg}\, N(\alpha)$. In Step 2, Algorithm Close is called, which executes $\ell = \log(\|\mathrm{val}_\infty(\gamma/\alpha)\|_\infty) + 1$ calls of Algorithm Double. Therefore it suffices to show that Double takes time polynomial in d, $\log q$, and $\mathrm{dg}\, N(\alpha)$.

Each call of Double calls Reduce once on input of the form $(B, \lfloor u \rceil)$; here B is the square of a reduced ideal, u is a vector in $\mathbb{Q}^k$ for some $k \le \ell$, with ℓ as above, and where $\lfloor u \rceil$ denotes the nearest integer vector to u. The ideal B is the square of a reduced ideal, and so is small. On the other hand, u is obtained from doubling a vector $t - \mathrm{val}_\infty(\mu)$ with $\|t - \mathrm{val}_\infty(\mu)\|_1 \le n + g$, so $\|u\|_1 \le 2n + 2g$. Rounding u to $\lfloor u \rceil$ adds at most $k/2$ to the 1-norm, and $k \le n$, so $\|\lfloor u \rceil\|_1 \le 5n/2 + 2g$. By Proposition 3.5, we find that each call of Reduce takes time polynomial in d, $\log q$, $\mathrm{dg}\, N(\alpha)$, as we were to show. □

Algorithm 4.8 (Close).

Input: A reduced ideal A and a vector $s \in \mathbb{Q}^n$.

Output: A BMR $(\beta_1, \ldots, \beta_\ell)$ of a minimum $\beta \in A$ which is close to s, where $\ell = \log(\|s\|_\infty) + 1$.

1. Let $\beta_0 = 1$, $\ell = \log(\|s\|_\infty) + 1$ and $t = s/2^\ell$.
2. For k from 1 to ℓ
   (a) Let $(\beta_1, \ldots, \beta_k) := \text{Double}(A, t, (\beta_0, \beta_1, \ldots, \beta_{k-1}))$.
   (b) Let $t := 2t$.
3. Return $(\beta_1, \ldots, \beta_\ell)$.

Proposition 4.9. *Algorithm 4.8 is correct.*

Proof. This follows from Proposition 4.11, together with the fact that in Step 1 of the algorithm $\beta_0 = 1$ is a minimum of A which is close to $t = s/2^\ell$. □

Algorithm 4.10 (Double).

Input: A reduced ideal A, a vector $t \in \mathbb{Q}^n$, and a BMR $(\beta_1, \ldots, \beta_{k-1})$ of a minimum β of A which is close to t.

Output: A BMR $(\beta_1, \ldots, \beta_{k-1}, \beta_k)$ of a minimum of A which is close to $2t$, where β_k is a minimum of $(1/\beta^2)A$ that has size polynomial in d, $\log q$, $\text{ht}(\text{div } A)$.

1. Let $B := (1/\beta^2)A$ and $u := 2t - \text{val}_\infty(\beta^2)$.
2. Reduce $(B, \lfloor u \rceil)$ to get a minimum β_k of B that is close to u. (Here $\lfloor u \rceil$ denotes the integer vector closest to u.)
3. Return $(\beta_1, \ldots, \beta_{k-1}, \beta_k)$. (This is a BMR of $\beta^2 \cdot \beta_k$.)

Proposition 4.11. *Algorithm 4.10 is correct.*

Proof. First we show that there exists a minimum β_k of B such that

$$\|\text{val}_\infty(\beta_k) - u\|_1 \leq n/2 + g.$$

When we reduce the pair $(B, \lfloor u \rceil)$ we get a pair

$$((1/\beta_k)B, \lfloor u \rceil - \text{val}_\infty(\beta_k)) \in \text{Rep}^{f*}(B).$$

Let $t = (t_1, \ldots, t_n) = \lfloor u \rceil - \text{val}_\infty(\beta_k)$. By Proposition 3.3, we have

$$\sum_{i=1}^{n} t_i \deg \mathfrak{p}_i \leq g,$$

where g is the genus of the function field K. The difference between the ℓ_1-norms of u and $\lfloor u \rceil$ is at most $n/2$, so $\mathrm{val}_\infty(\beta_k) - u$ has ℓ_1-norm bounded by $n/2 + g$. Thus there exists a minimum β_k of B that is close to $u = 2t - \mathrm{val}_\infty(\beta^2)$, and this minimum is computed in Step 3 of Double. Moreover, by Proposition 3.5, the size of the minimum β_k is polynomial in d, $\log q$, $\mathrm{ht}(\mathrm{div}(B))$, and $\|u\|_\infty$. Then, since β_k is close to $2t - \mathrm{val}_\infty(\beta^2)$, we have that $\beta^2\beta_k$ is close to $2t$, because

$$\|2t - \mathrm{val}_\infty(\beta^2\beta_k)\|_1 = \|(2t - (\mathrm{val}_\infty(\beta^2)) - \mathrm{val}_\infty(\beta_k)\|_1. \qquad \square$$

In the next proposition we show that if there is a minimum of A whose valuation vector equals $2t$, then Double returns a BMR of this minimum.

Proposition 4.12. *Let A be a reduced ideal. Suppose there is a minimum μ of A such that $\mathrm{val}_\infty(\mu) = 2t$. Then* Double$(A, t, \beta = (\beta, \ldots, \beta_{k-1}))$ *returns the BMR $(\beta_1, \ldots, \beta_k)$ of this minimum; that is, $\mu = \beta^2\beta_k$.*

Proof. In Step 3 of Double the algorithm reduces the pair $((1/\beta^2)A, 2t - \mathrm{val}_\infty(\beta^2))$, where β is the given minimum of A which is close to t. Since $2t = \mathrm{val}_\infty(\mu)$, we see that $2t$ has integer coordinates, so it is not necessary to round $u = 2t - \mathrm{val}_\infty(\beta^2)$.

After reducing $((1/\beta^2)A, 2t - v(\beta^2))$ we obtain an element

$$((1/(\beta_k\beta^2))A, 2t - \mathrm{val}_\infty(\beta^2) - \mathrm{val}_\infty(\beta_k))$$

of $\mathrm{Rep}^{f*}(\mathbb{O})$, where β_k is a minimum of $(1/\beta^2)A$. Since reduction produces a unique element in $\mathrm{Rep}^{f*}(\mathbb{O})$ and elements of $\mathrm{Rep}^{f*}(\mathbb{O})$ have unique representatives, this implies that β_k is uniquely determined (up to multiplication by an element of $\mathbb{F}_q^*$). Since μ is a minimum of A, we have $((1/\mu)A, 0) \in \mathrm{Rep}^{f*}(\mathbb{O})$. We also have that $\nu := \mu/(\beta^2)$ is a minimum of $(1/\beta^2)A$. Then

$$\big((1/\nu)(1/\beta^2)A, 2t - \mathrm{val}_\infty(\beta^2) - \mathrm{val}_\infty(\nu)\big) = ((1/\mu)A, 0) \in \mathrm{Rep}^{f*}(\mathbb{O}).$$

Hence we must have $\beta_k = \mu/\beta^2$; that is, Double returns a BMR of $\mu = \beta_k\beta^2$. $\square$

Corollary 4.13. *If the input in* Close *(Algorithm 4.8) is a reduced ideal A and a vector $s \in \mathbb{Q}^n$ such that $s = \mathrm{val}_\infty(\mu)$ for a minimum μ of A, then* Close *outputs a BMR of μ.*

Proof. At the last step of the for loop in Step 2 of Close, we have a BMR of a minimum β of A that is close to $s/2$, and the last call of Double produces a BMR of a minimum β' of A that is close to s. By Proposition 4.12, Double outputs a BMR of μ, so Algorithm Close returns a BMR of μ with $s = \mathrm{val}_\infty(\mu)$. $\square$

Corollary 4.14. *The principal ideal problem is in NP.*

Proof. Given a function field and an ideal I of $\mathbb{O}$ represented in Hermite normal form (HNF), if the ideal is principal, then the proof is a compact representation

$B = (\gamma, (\beta_1, \ldots, \beta_\ell))$ for α, where $I = \alpha\mathbb{O}$. By Definition 4.3, the compact representation B has size bounded by $\log(\|\mathrm{val}_\infty(\alpha)\|_\infty + g)$ and $\mathrm{poly}(\log q, d, \deg N(\alpha))$. The field parameters are $\log q$, d, and g. By Remark 3.8, being polynomial in $\deg(N(\alpha))$ is the same as being polynomial in $\mathrm{ht}(\mathrm{div}(A))$, which is the size of the ideal $A = \alpha\mathbb{O}$. Propositions 3.3 and 3.5 tell us that $\|\log \mathrm{val}_\infty(\alpha)\|_\infty$ is bounded.

The verifier must efficiently test whether $A = (\gamma/\beta)\mathbb{O}$, where $\beta = \prod \beta_i^{2^{n-i}}$. The verifier can efficiently compute the ideal as follows. By Claim 4.5, $(1/\beta)\mathbb{O}$ can be efficiently computed. Multiplication by γ is efficient. Finally, comparing the HNF of A and $(\gamma/\beta)\mathbb{O}$ is efficient since the representation of an ideal is unique. □

5. Quantum algorithms for the unit group, the principal ideal problem, and the class group

In this section we give efficient quantum algorithms for computing the unit group, solving the principal ideal problem and computing the class group. Recall from Section 3 that for the unit group and the principal ideal problem this means the objects are computed in the val_∞ embedding, and that compact representations can then be computed.

The basic approach is to show that each of these problems reduces to an instance of the abelian hidden subgroup problem (HSP), which is known to have an efficient quantum algorithm [10]. The class group case is slightly more general since the HSP instances will take values that are quantum states.

Theorem 5.1. *There is a polynomial-time quantum algorithm for computing the unit group of a function field.*

Proof. A hidden subgroup problem for the unit group can be defined by the function $g\colon \mathbb{Z}^n \to \mathrm{Rep}^{f*}(\mathbb{O})$ defined as $g(e) = \mathrm{Reduce}(\mathbb{O}, e)$. Here $\mathrm{Reduce}(\mathbb{O}, e)$ is the element of $\mathrm{Rep}^{f*}(\mathbb{O})$ that is computed by Algorithm 3.4; it is polynomial-time computable by Proposition 3.5. By Remark 3.6,

$$\mathrm{Reduce}(\mathbb{O}, e) = \mathrm{Reduce}(\mathbb{O}, e + v)$$

for every $v \in \Lambda$, so the function g is constant on cosets. Therefore g is also defined on $\mathbb{Z}^n/\Lambda$, and it gives a bijection between $\mathbb{Z}^n/\Lambda$ and $\mathrm{Rep}^{f*}(\mathbb{O})$, so it is also distinct on different cosets. Using the HSP instance g, a quantum algorithm can compute a basis for Λ efficiently. Compact representations can then be computed if desired. □

In the *decision* version of the principal ideal problem, an ideal I of $\mathbb{O}$ is given in HNF and the problem is to decide if it is principal. If it is principal, then the *search* version of the problem is to compute a generator; that is, to compute an α such that $I = \alpha\mathbb{O}$. Since generators may take an exponential number of bits to

represent in general, we only require computing $\mathrm{val}_\infty(\alpha)$. Knowing $\mathrm{val}_\infty(\alpha)$ and $\alpha\mathbb{O}$ determines α up to multiplication by an element of k^*. So given an arbitrary ideal I that is given to us in HNF, the strategy is to solve the search problem and compute a candidate value for $\mathrm{val}_\infty(\alpha)$, and then to test whether $I = \alpha\mathbb{O}$ to see if the ideal is principal or not. A compact representation of α can then be computed from $\mathrm{val}_\infty(\alpha)$ and I using Algorithm 4.6.

Theorem 5.2. *There is a polynomial-time quantum algorithm for the principal ideal problem in a function field.*

Proof. Recall that for a vector $v \in \mathbb{Z}^n$, calling Algorithm 3.4 on $(\mathbb{O}, v)$ results in a pair $(I_v, f_v) \in \mathrm{Rep}^{f*}(\mathbb{O})$. Here I_v is a reduced ideal and f_v is a vector such that f_v has positive coordinates. If $(1/\mu)\mathbb{O} = I_v$ then $\mathrm{val}_\infty(1/\mu) + f_v = v$ by Remark 3.6.

To solve the principal ideal problem we do the following: Given any ideal I we first call Algorithm 3.4 on $(I, 0)$ to get a reduced ideal I_v. The reduction algorithm also computes γ such that $(1/\gamma)I = I_v$. Hence it suffices to solve the principal ideal problem for reduced ideals. If $I_v = (1/\mu)\mathbb{O}$ is reduced, then I_v represents the point $v + \Lambda \in \mathbb{Z}^n/\Lambda$ with $v = \mathrm{val}_\infty(\mu)$. By the above discussion, solving the principal ideal problem means computing v. First, by Theorem 5.1, a basis B of the unit group (under the embedding val_∞) can be computed efficiently with a quantum algorithm. A hidden subgroup problem can be set up as follows. By abuse of notation we denote by $\mathbb{Z}^n/B$ the quotient of $\mathbb{Z}^n$ by the lattice generated by the elements in B. Let $G = \mathbb{Z}_M \times \mathbb{Z}^n/B$, where $M = |\mathbb{Z}^n/B|$. Define $h\colon G \to \mathrm{Rep}^{f*}(K) = \bigcup_A \mathrm{Rep}^{f*}(A)$ by the following algorithm: On input (a, b), use the composition operation in Section 3C and repeated doubling to compute a times the group element (this does reductions along the way, giving an element in $\mathrm{Rep}^{f*}(K)$); then compose the result with $(\mathbb{O}, -b)$ and reduce. When the ideal I is principal, we have $h(a, b) = (I_{av-b}, f_{av-b})$. The hidden subgroup in this case is $H = \langle(1, v)\rangle$, and $h(H) = (\mathbb{O}, 0)$. A set of coset representatives for H is $\{(0, w) : w \in \mathbb{Z}^n/B\}$. Then $h((0, w) + n(1, v)) = h(n, w + nv) = (I_{-w}, f_{-w})$, and so the different values of w correspond to the set of elements in $\mathrm{Rep}^{f*}(\mathbb{O})$. So h is constant on cosets and distinct on different cosets. The function h can be computed efficiently using a small modification of Close (Algorithm 4.8). Therefore there is an efficient quantum algorithm for finding generators for H. Given an element $(n, nv) \in \mathbb{Z}_M \times \mathbb{Z}^n/B$ of H, it is easy to compute v. □

Theorem 5.3. *There is a polynomial-time quantum algorithm for computing the ideal class group of a function field.*

Proof. To compute the class group we also reduce to an abelian hidden subgroup problem where the function takes quantum states as values. Since it is not known how to compute unique representatives in the class group we instead create quantum states to represent each element, as a superposition over all elements of

$\mathrm{Rep}^{f*}(J)$ for the ideal class of J. Let $g_1, \ldots, g_m$ be a set of generators for $\mathrm{Cl}(\mathbb{O})$; Appendix A shows how to compute such a set. For an ideal J, let

$$|\phi_J\rangle = \sum_{(I,v)\in \mathrm{Rep}^{f*}(J)} |I, v\rangle.$$

Define

$$f\colon \mathbb{Z}^m \to \mathbb{C}^{|\mathrm{Pic}^0(K)|}$$

by $f(e_1, \ldots, e_m) = |\phi_J\rangle$, where J is the ideal resulting from $\mathrm{Reduce}(g_1^{e_1} \cdots g_m^{e_m}, 0)$. The function f can be efficiently evaluated using the algorithm for the principal ideal problem as follows. Given $|e_1, \ldots, e_m\rangle$, compute $|e_1, \ldots, e_m, J\rangle$, where J is the ideal resulting from $\mathrm{Reduce}(g_1^{e_1} \cdots g_m^{e_m}, 0)$. The ideal in the last register, call it J, is now used to create the superposition over reduced ideals. Create $\sum_{v\in\mathbb{Z}^n/B} |J, v\rangle$, then $\sum_{v\in\mathbb{Z}^n/B} |J, v, (J_v, f_v)\rangle$ where (J_v, f_v) is the result of calling $\mathrm{Reduce}(J, v)$. Next use the principal ideal algorithm on $J \cdot J_v^{-1}$, which outputs v, to create $\sum_{v\in\mathbb{Z}^n/B} |J, v, (J_v, f_v), v\rangle$. Next uncompute v in the second register using the fourth, then uncompute the fourth register by running the principal ideal algorithm backwards. Finally, uncompute J using $e_1, \ldots, e_m$. □

Acknowledgments

This work was supported in part by the National Security Agency (NSA) under Army Research Office (ARO) contract number W911NF-08-1-0298. The first author was partially supported by National Science Foundation grant DMS-1056703 and a Sloan Research Fellowship. The second author was partially supported by National Science Foundation grant CCF-0747274.

Appendix A. Computing generators for $\mathrm{Cl}(\mathbb{O})$

As usual, let K be an algebraic function field over a finite field of constants $k = \mathbb{F}_q$. As discussed in Section 2, when $S = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_{n+1}\}$ is the set of places at infinity and $\deg \mathfrak{p}_{n+1} = 1$, we have a short exact sequence

$$0 \longrightarrow \mathrm{Ker} \longrightarrow \mathrm{Pic}^0(K) \longrightarrow \mathrm{Cl}(\mathbb{O}) \to 1$$

where the map from $\mathrm{Pic}^0(K) \to \mathrm{Cl}(\mathbb{O})$ is given as

$$\sum_{\mathfrak{p}\in\mathcal{P}_K} n_{\mathfrak{p}}\mathfrak{p} \longmapsto \prod_{\mathfrak{p}\in\mathcal{P}_K - S} (\mathfrak{p} \cap \mathbb{O})^{-n_{\mathfrak{p}}}.$$

Given a function field K as above, there is a smooth projective geometrically irreducible curve C whose function field is K. Let g denote the genus of this curve.

In [28] Kedlaya proved that for q with $q^{1/2} > 16g$ there exists a randomized algorithm that produces a generating set for $\mathrm{Pic}^0(K)$ in time polynomial in g and

$\log q$. The genus of the curve C does not change if we increase the size of the base field k. Hence by enlarging the constant field, if necessary, we may assume that $q^{1/2} > 16g$. From the exact sequence above it follows that the image of the generating set for $\mathrm{Pic}^0(K)$ under the map described above gives a generating set of $\mathrm{Cl}(\mathbb{O})$.

Appendix B. Computing Riemann-Roch spaces

In this section we analyze the complexity of computing the Riemann-Roch space $L(D) := \{f \in K : \mathrm{div}(f) + D \geq 0\} \cup \{0\}$. The input to the problem is a function field K and a divisor

$$D = \left(A, \sum_{i=1}^{n+1} t_i \mathfrak{p}_i \right)$$

of K. The fractional ideal A of $\mathbb{O}$ is given in HNF relative to an $\mathbb{O}$-basis. The second part of D is given by a set of integers $\{t_i : 1 \leq i \leq n+1\}$ that determine the multiplicity of the infinite places, that is, the places of $S = \{\mathfrak{p}_1, \ldots, \mathfrak{p}_{n+1}\}$, in D.

We follow Hess's [26] algorithm to compute the Riemann-Roch space. In [26] Hess does not include any proofs for the complexity of his algorithm, so in this section we show that the Riemann-Roch space $L(D)$ can be computed in time polynomial in $d, \log q$ and $\mathrm{ht}(D)$. (For the definition of $\mathrm{ht}(D)$ see Section 2.) Hess's algorithm is a relatively simple, self-contained algorithm. We also investigate more closely the complexity of computing $\mathfrak{o}_\infty$-bases of the ideals we are working with.

The main idea in [26] is that the Riemann-Roch space can be computed as the intersection of two ideals that come from the divisor D, where the two ideals are in the rings $\mathbb{O}$ and $\mathbb{O}_\infty$.

First we show that we can compute an $\mathfrak{o}_\infty$-basis for $\mathbb{O}_\infty$ in polynomial time.

Proposition B.1 ([15], Proposition 4.13). *Let $R \subset S$ be commutative rings with identity and let U be a multiplicatively closed subset of R. If S' is the integral closure of R in S, then $S'[U^{-1}]$ is the integral closure of $R[U^{-1}]$ in $S[U^{-1}]$.*

Lemma B.2. *An $\mathfrak{o}_\infty$-basis for $\mathbb{O}_\infty$ is computable in time polynomial in d and $\log q$.*

Proof. By [11, Theorem 1] applied to $k[1/x]$, we can compute a basis $\beta_1, \ldots, \beta_d$ of the integral closure of $k[1/x]$ in K. By Proposition B.1, taking integral closures commutes with localization, so when we apply the proposition to the rings $R = k[1/x]$ and $S = K$, with U being the complement of the prime ideal $(1/x)$ of R, we find that $\mathfrak{o}_\infty = k[1/x][U^{-1}]$. Let S' be the integral closure of $k[1/x]$ in K. Then $\mathbb{O}_\infty = S'[U^{-1}]$, which implies that $\beta_1, \ldots, \beta_d$ is an $\mathfrak{o}_\infty$-basis for $\mathbb{O}_\infty$. □

Lemma B.3. *Let A be a fractional ideal of $\mathbb{O}$ given by a $k[x]$-basis, and let B be a fractional ideal of $\mathbb{O}_\infty$ given by an $\mathfrak{o}_\infty$-basis. There exist bases $a_1, \ldots, a_d$ of A and $b_1, \ldots, b_d$ of B and uniquely determined integers λ_i such that $x^{-\lambda_i} b_i = a_i$.*

Proof. Let $a'_1, \dots, a'_d \in K$ be a $k[x]$-basis of A and $b'_1, \dots, b'_d \in K$ a $\mathfrak{o}_\infty$-basis of B. Both of these are bases for K as a $k(x)$-vector space. Let $M \in k(x)^{d\times d}$ be such that

$$(a'_1, \dots, a'_d) = (b'_1, \dots, b'_d)M.$$

By [26, Corollary 4.3] there exists a unimodular $T_1 \in \mathfrak{o}_\infty^{d\times d} \subset k[[x^{-1}]]^{d\times d}$ and a unimodular $T_2 \in k[x]^{d\times d}$ such that $T_1 M T_2 = (x^{-\lambda_j}\delta_{ij})_{ij}$.

Let $(a_1, \dots, a_d) = (a'_1, \dots, a'_d)T_2$ and $(b_1, \dots, b_d) = (b'_1, \dots, b'_d)T_1^{-1}$. Then

$$(b_1, \dots, b_n)T_1 M T_2 = (b'_1, \dots, b'_d)M T_2 = (a'_1, \dots, a'_d)T_2 = (a_1, \dots, a_d). \quad \square$$

Lemma B.4. *If $a_1, \dots, a_d$ and $b_1, \dots, b_d$ are bases as in Lemma* B.3, *then $A \cap B$ has k-basis $\{x^j a_i : 1 \le i \le d, 0 \le j \le \lambda_j\}$.*

Proof. Assume $\lambda \ge 0$. Because $x \in \mathbb{C}$, the elements $x^j a_i$ lie in A for $j \ge 0$, so all we have to show is that $x^j a_i \in B$ if and only if $0 \le j \le \lambda$. We have $a_i = x^{-\lambda_i} b_i \in B$ since $1/x \in \mathfrak{o}_\infty$, B is an $\mathfrak{o}_\infty$-module and $\lambda_i \ge 0$. Similarly, $x^j a_i = x^{j-\lambda_i} b_i \in B$ if and only if $j - \lambda_i \le 0$, that is, if and only if $j \le \lambda_i$. But $x^j a_i \in A$ if and only if $j \ge 0$, so $x^j a_i \in A \cap B$ for $0 \le j \le \lambda_i$.

To see that this set forms a k-basis note that $A \cap B = \bigcup_{i=1}^d (A \cap B \cap k(x)a_i)$, and a k-basis for $A \cap B$ is the union of the k-bases for $A \cap B \cap k(x)a_i$.

But for i with $\lambda_i \ge 0$ we have $A \cap B \cap k(x)a_i = A \cap B \cap a_i k[x]$, so it suffices to determine which monomials $(x^j)a_i$ are in this intersection. By the above argument the only monomials in this intersection are $a_i, xa_i, \dots, x^{\lambda_i} a_i$, and these elements are clearly linearly independent over k, so they form a k-basis for $A \cap B \cap k(x)a_i$ (for i with $\lambda_i \ge 0$). $\square$

Lemma B.5. *The elements $a_1, \dots, a_d$ and the integers $\lambda_1, \dots, \lambda_d$ above can be computed in polynomial time.*

Proof. We will first show that the matrices M and T_2 from the proof of Lemma B.3 can be computed in polynomial time. The lemma then follows from the fact that $(a_1, \dots, a_d) = (a'_1, \dots, a'_d)T_2$, and that the maximum degree of elements of the j-th column of MT_2 is equal to $-\lambda_j$([19, p. 15], [26, Corollary 4.3]). When elements in K are specified as polynomials in y, that is, as $\sum_{i=0}^n a_i y^i$ for coefficients $a_i \in k(x)$, then writing a element $\alpha \in K$ in terms of a basis $\omega_1, \dots, \omega_n$ is a vector space transformation, with vector space generators $1, y, y^2, \dots, y^{n-1}$. The columns of the matrix $A \in k(x)^{n\times n}$ contain the coefficients of the polynomials $\omega_1, \dots, \omega_n$. Then solving the equation $Az = b$ over $k(x)$ for z gives the coefficients of b in terms of the basis. For M, this can be done for each column.

The matrix T_2 is computed using Paulus's polynomial-time algorithm [32] by keeping track of the operations during the basis reduction. $\square$

Algorithm B.6 (Ideal intersection for ideals in two different rings).

Input: A function field K; an element $x \in K$; a $k[x]$-basis $\omega_1, \ldots, \omega_d$ of $\mathbb{O}$; a $k[x]$-basis $a'_1, \ldots, a'_d$ of the fractional ideal A of $\mathbb{O}$; an $\mathfrak{o}_\infty$-basis $v_1, \ldots, v_d$ of $\mathbb{O}_\infty$; and an $\mathfrak{o}_\infty$-basis $b'_1, \ldots, b'_d$ of the fractional ideal B of $\mathbb{O}_\infty$.

Output: Elements $a_1, \ldots, a_d$ of K and integers $\lambda_1, \ldots, \lambda_d$ such that $\{x^j a_i : 1 \leq i \leq d, 0 \leq j \leq \lambda_i\}$ is k-basis of the k-vector space $A \cap B$.

1. Compute a matrix M such that $(b'_1, \ldots, b'_d)M = (a'_1, \ldots, a'_d)$.
2. Do a basis reduction on M. Keep track of the operations and let $T_2 \in Gl_d(k[x])$ be the transformation. Let $-\lambda_i$ be the maximum degree in the i-th column of the reduced matrix MT_2.
3. Let $(a_1, \ldots, a_d) = (a'_1, \ldots, a'_d)T_2$.
4. Return $(a_1, \ldots, a_d; \lambda_1, \ldots, \lambda_d)$.

Proposition B.7. *Algorithm B.6 is correct, and runs in polynomial time.*

Proof. The matrix M computed in Step 1 of the algorithm is exactly the matrix from Lemma B.3 that leads to the special basis for A; that is, $(a_1, \ldots, a_d) = (a'_1, \ldots, a'_d)T_2$. By Lemma B.4 and its proof, if $-\lambda_j$ is the maximum column degree in the j-th column of MT_2, then $\{x^j a_i : 1 \leq i \leq d, 0 \leq j \leq \lambda_i\}$ is a k-basis for the intersection $A \cap B$. By Lemma B.5, the a_i and the λ_i can be computed in polynomial time. □

Algorithm B.8 (Riemann-Roch space).

Input: A function field K; a $k[x]$-basis $\omega_1, \ldots, \omega_d$ of $\mathbb{O}$; and a divisor $D = (A, \sum_{i=1}^{n+1} t_i \mathfrak{p}_i)$, where A is a fractional ideal of $\mathbb{O}$ given in a $k[x]$-basis.

Output: Elements $a_1, \ldots, a_d$ of K and integers $\lambda_1, \ldots, \lambda_d$ such that $\{x^j a_i : 1 \leq i \leq d, 0 \leq j \leq \lambda_i\}$ is a basis of the Riemann-Roch space $L(D)$.

1. Compute a $k[x]$-basis of A^{-1}.
2. Compute an $\mathfrak{o}_\infty$-basis of $B := \prod_{i=1}^{n+1} (\mathfrak{p}_i \cap \mathbb{O}_\infty)^{t_i} \subseteq \mathbb{O}_\infty$.
3. Compute an $\mathfrak{o}_\infty$-basis of B^{-1}.
4. Use Algorithm B.6 to compute $A^{-1} \cap B^{-1}$.
5. Return the $(a_1, \ldots, a_d; \lambda_1, \ldots, \lambda_d)$ computed by Algorithm B.6.

Theorem B.9. *Algorithm B.8 computes the Riemann-Roch space $L(D)$ in time polynomial in d, $\log q$, and* $\mathrm{ht}(D)$.

Proof. Computing the inverse of a fractional ideal A of $\mathbb{O}$ can be done in time polynomial in $\log q$, d, and $\mathrm{ht}(\mathrm{div}(A))$ [14, Proposition 2.69(b)]. The ideals $\mathfrak{p}_i \cap \mathbb{O}_\infty$ in Step 2 are the prime ideals of $\mathbb{O}_\infty$ corresponding to the places in S. These can be computed in polynomial time with an algorithm similar to the one given for number fields in [16]. Hence we can compute an $\mathfrak{o}_\infty$-basis for the ideal B in Step 2 in polynomial time. The inverse of an ideal B in this ring can be computed efficiently as well. Finally, by Proposition B.7 above, a basis for the k-vector space $A^{-1} \cap B^{-1}$ can be computed in polynomial time. □

References

[1] ACM (ed.), *STOC'*05: *Proceedings of the* 37*th Annual ACM Symposium on Theory of Computing*, *held in Baltimore*, *MD*, *May* 22–24, 2005, New York, Association for Computing Machinery, 2005. MR 2006f:68006

[2] ACM (ed.), *STOC'*07*—Proceedings of the* 39*th Annual ACM Symposium on Theory of Computing*, *held in San Diego*, *CA*, *June* 11–13, 2007, New York, Association for Computing Machinery, 2007. MR 2009a:68002

[3] ACM (ed.), *STOC'*09*—Proceedings of the* 2009 *ACM International Symposium on Theory of Computing*, *held in Bethesda*, *MD*, *May* 31–*June* 2, 2009, New York, Association for Computing Machinery, 2009. MR 2012d:68003

[4] Leonard M. Adleman, Jonathan DeMarrais, and Ming-Deh Huang, *A subexponential algorithm for discrete logarithms over the rational subgroup of the Jacobians of large genus hyperelliptic curves over finite fields*, in Adleman and Huang [5], 1994, pp. 28–40. MR 96b:11078

[5] Leonard M. Adleman and Ming-Deh Huang (eds.), *Algorithmic number theory*: *Proceedings of the First International Symposium* (*ANTS-I*) *held at Cornell University*, *Ithaca*, *New York*, *May* 6–9, 1994, Lecture Notes in Computer Science, no. 877, Berlin, Springer, 1994. MR 95j:11119

[6] G. Brassard (ed.), *Advances in cryptology—CRYPTO '*89: *Proceedings of the Conference on the Theory and Applications of Cryptology held at the University of California*, *Santa Barbara*, *California*, *August* 20–24, 1989, Lecture Notes in Computer Science, no. 435, New York, Springer, 1990. MR 91b:94002

[7] Johannes A. Buchmann and Hugh C. Williams, *A key exchange system based on real quadratic fields* (*extended abstract*), in Brassard [6], 1990, pp. 335–343. MR 91f:94014

[8] J. P. Buhler (ed.), *Algorithmic number theory*: *Proceedings of the* 3*rd International Symposium* (*ANTS-III*) *held at Reed College*, *Portland*, *OR*, *June* 21–25, 1998, Lecture Notes in Computer Science, no. 1423, Berlin, Springer, 1998. MR 2000g:11002

[9] Moses Charikar (ed.), *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms held in Austin*, *TX*, *January* 17–19, 2010, Philadelphia, PA, Society for Industrial and Applied Mathematics, 2010. MR 2012f:68008

[10] Kevin K. H. Cheung and Michele Mosca, *Decomposing finite abelian groups*, Quantum Inf. Comput. **1** (2001), no. 3, 26–32. MR 2003e:81030

[11] A. L. Chistov, *The complexity of the construction of the ring of integers of a global field*, Dokl. Akad. Nauk SSSR **306** (1989), no. 5, 1063–1067. MR 90g:11170

[12] Henri Cohen (ed.), *Algorithmic number theory*: *Proceedings of the* 2*nd International Symposium* (*ANTS-II*) *held at the Université Bordeaux I*, *Talence*, *May* 18–23, 1996, Lecture Notes in Computer Science, no. 1122, Berlin, Springer, 1996. MR 97k:11001

[13] Henri Cohen, Gerhard Frey, Roberto Avanzi, Christophe Doche, Tanja Lange, Kim Nguyen, and Frederik Vercauteren (eds.), *Handbook of elliptic and hyperelliptic curve cryptography*, Chapman & Hall/CRC, Boca Raton, FL, 2006. MR 2007f:14020

[14] Claus Diem, *On arithmetic and the discrete logarithm problem in class groups of curves*, Habilitationsschrift, Universität Leipzig, 2008. http://www.math.uni-leipzig.de/∼diem/preprints/habil.pdf

[15] David Eisenbud, *Commutative algebra*, *with a view toward algebraic geometry*, Graduate Texts in Mathematics, no. 150, Springer, New York, 1995. MR 97a:13001

[16] Kirsten Eisenträger and Sean Hallgren, *Algorithms for ray class groups and Hilbert class fields*, in Charikar [9], 2010, pp. 471–483. MR 2012i:11103

[17] Felix Fontein, *The infrastructure of a global field of arbitrary unit rank*, Math. Comp. **80** (2011), no. 276, 2325–2357. MR 2012f:11243

[18] Felix Fontein and Pawel Wocjan, *Quantum algorithm for computing the period lattice of an infrastructure*, 2011. arXiv 1111.1348 [quant-ph]

[19] Felix Wolfgang Fontein, *The infrastructure of a global field and baby step-giant step algorithms*, Ph.D. thesis, University of Zurich, 2009. http://dx.doi.org/10.5167/uzh-24993

[20] Arnaldo García and Henning Stichtenoth, *A tower of Artin-Schreier extensions of function fields attaining the Drinfel′d-Vlăduţ bound*, Invent. Math. **121** (1995), no. 1, 211–222. MR 96d:11074

[21] Henri Gilbert (ed.), *Advances in cryptology—EUROCRYPT* 2010: *Proceedings of the* 29*th Annual International Conference on the Theory and Applications of Cryptographic Techniques held on the French Riviera*, *May* 30–*June* 3, 2010, Lecture Notes in Computer Science, no. 6110, Berlin, Springer, 2010. MR 2011g:94001

[22] V. D. Goppa, *Geometry and codes*, Mathematics and its Applications (Soviet Series), no. 24, Kluwer Academic Publishers Group, Dordrecht, 1988. MR 91a:14013

[23] Venkatesan Guruswami, *Constructions of codes from number fields*, IEEE Trans. Inform. Theory **49** (2003), no. 3, 594–603. MR 2004g:94093

[24] ———, *Artin automorphisms*, *cyclotomic function fields*, *and folded list-decodable codes* (*extended abstract*), in ACM [3], 2009, pp. 23–32. MR 2780046

[25] Sean Hallgren, *Fast quantum algorithms for computing the unit group and class group of a number field*, in ACM [1], 2005, pp. 468–474. MR 2006g:81032

[26] F. Hess, *Computing Riemann-Roch spaces in algebraic function fields and related topics*, J. Symbolic Comput. **33** (2002), no. 4, 425–445. MR 2003j:14032

[27] Ming-Deh Huang and Doug Ierardi, *Efficient algorithms for the Riemann-Roch problem and for addition in the Jacobian of a curve*, J. Symbolic Comput. **18** (1994), no. 6, 519–539. MR 96h:14077

[28] Kiran S. Kedlaya, *Quantum computation of zeta functions of curves*, Comput. Complexity **15** (2006), no. 1, 1–19. MR 2007b:14042

[29] A. K. Lenstra and H. W. Lenstra, Jr. (eds.), *The development of the number field sieve*, Lecture Notes in Mathematics, no. 1554, Springer, Berlin, 1993. MR 96m:11116

[30] Vadim Lyubashevsky, Chris Peikert, and Oded Regev, *On ideal lattices and learning with errors over rings*, in Gilbert [21], 2010, pp. 1–23. MR 2660480

[31] Volker Müller, Andreas Stein, and Christoph Thiel, *Computing discrete logarithms in real quadratic congruence function fields of large genus*, Math. Comp. **68** (1999), no. 226, 807–822. MR 99i:11119

[32] Sachar Paulus, *Lattice basis reduction in function fields*, in Buhler [8], 1998, pp. 567–575. MR 2000i:11193

[33] Chris Peikert and Alon Rosen, *Lattices that admit logarithmic worst-case to average-case connection factors*, in ACM [2], 2007, pp. 478–487. MR 2010e:68099

[34] Michael Rosen, *Number theory in function fields*, Graduate Texts in Mathematics, no. 210, Springer, New York, 2002. MR 2003d:11171

[35] Pradeep Sarvepalli and Pawel Wocjan, *Quantum algorithms for one-dimensional infrastructures*, 2011. arXiv 1106.6347 [quant-ph]

[36] R. Scheidler, *Compact representation in real quadratic congruence function fields*, in Cohen [12], 1996, pp. 323–336. MR 98c:11126

[37] ———, *Decision problems in quadratic function fields of high genus*, J. Complexity **16** (2000), no. 2, 411–423. MR 2001e:11112

[38] Arthur Schmidt and Ulrich Vollmer, *Polynomial time quantum algorithm for the computation of the unit group of a number field* (*extended abstract*), in ACM [1], 2005, pp. 475–480. MR 2006g:81038

[39] René Schoof, *Algebraic curves over* $\mathbf{F}_2$ *with many rational points*, J. Number Theory **41** (1992), no. 1, 6–14. MR 93h:11062

[40] Nigel P. Smart, *Reduced ideals in function fields*, Tech. Report, HP Laboratories Bristol, 1998. http://www.hpl.hp.com/techreports/98/HPL-98-201.html

[41] Henning Stichtenoth, *Algebraic function fields and codes*, 2nd ed., Graduate Texts in Mathematics, no. 254, Springer, Berlin, 2009. MR 2010d:14034

[42] Christoph Thiel, *On the complexity of some problems in algorithmic algebraic number theory*, Ph.D. thesis, Universität des Saarlandes, 1995. http://tinyurl.com/thiel-phd

[43] Emil J. Volcheck, *Computing in the Jacobian of a plane algebraic curve*, in Adleman and Huang [5], 1994, pp. 221–233. MR 96a:14033

KIRSTEN EISENTRÄGER: eisentra@math.psu.edu
Department of Mathematics, Penn State University, University Park, PA 16802, United States

SEAN HALLGREN: hallgren@cse.psu.edu
Department of Computer Science and Engineering, Penn State University, University Park, PA 16802, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.359

msp

The complex polynomials $P(x)$ with $\mathrm{Gal}(P(x)-t) \cong M_{23}$

Noam D. Elkies

We find the polynomials $P \in \mathbb{C}[X]$ of degree 23 such that the Galois group of $P(x)-t$ over $\mathbb{C}(t)$ is the Mathieu group M_{23}. This completes the computation of polyomials P for which the Galois group of $P(x)-t$ is among the exceptional groups listed by Müller.

1. Introduction

For $P \in \mathbb{C}[x]$ of degree $n > 0$, define G_P to be the Galois group of $P(x) - t$ over $\mathbb{C}(t)$. Since $P(x) - t$ is irreducible, G_P is a transitive subgroup of the symmetric group S_n. Generically[1] G_P is all of S_n, but it can be as small as the cyclic or dihedral group for special choices such as $P = x^n$ or $P = T_n(x)$ (Chebyshev polynomial) respectively. If P decomposes as $P(x) = P_1(P_2(x))$ with each $\deg(P_i) > 1$, then G_P permutes the proper subsets $\{x : P_2(x) = u\}$ of the roots with $P_1(u) = t$, and is therefore imprimitive. The converse implication is shown in [8, Proposition 3.4]. Müller [12] determined all G_P that can arise for indecomposable polynomials: they are the symmetric and alternating groups, the cyclic groups of prime order, the dihedral groups of order twice an odd prime, and twelve exceptional permutation groups with $n = 6, 7, \ldots, 23, 31$, the last two for the sporadic Mathieu group M_{23} and the linear group $\mathrm{GL}_5(\mathbb{Z}/2\mathbb{Z})$.

The proof uses covering-space methods and Riemann's existence theorem, and thus does not yield explicit polynomials. But it is still a natural question to exhibit all P that realize each possible group G_P, except for the cases of A_n and S_n,

MSC2010: primary 12F12; secondary 20D08.
Keywords: Mathieu group M_{23}, Galois groups, Chebotarev density theorem.

[1] In particular, $G_P = S_n$ if dP/dx has $n-1$ distinct roots at which P takes distinct values; equivalently, if $\mathrm{disc}_t(\mathrm{disc}_x(P(x)-t)) \neq 0$. This sufficient (but far from necessary) condition was already noted by Hilbert ([10], see also [15, §4.4]); the formulation in terms of the discriminant of the discriminant is attributed to Davenport in [3, p. 422].

which occur in "many, not reasonably classifiable types" [12]. Say $P, Q \in \mathbb{C}[x]$ are *equivalent* if $Q(x) = L_1(P(L_2(x)))$ for some polynomials L_1, L_2 both of degree 1; then $G_P = G_Q$. Up to this equivalence, the cyclic and dihedral groups occur only for powers and Chebyshev polynomials respectively. Some of the exceptional groups were realized in [12], or earlier by Matzat [11]; most of the others were realized by Cassou-Noguès and Couveignes [4],[2] leaving only M_{23}. Here we find the polynomials P with $G_P \cong M_{23}$.

The main novelty here is not in the computation of P but in the proof that $G_P \cong M_{23}$. The coefficients of P were computed using a known p-adic method for finding polynomial identities by solving the equivalent system of nonlinear equations in the coefficients, though here the search for the initial approximation took several CPU-days. The difficulty was that these equations cannot distinguish between polynomials with Galois group M_{23} and A_{23}, and there are four M_{23}-covers but numerous A_{23}-covers with the same cycle structure (with all the A_{23}-covers probably defined only over number fields of rather high degree). Once we found P with coefficients in a quartic number field F, we quickly convinced ourselves that G_P must be M_{23} by factoring $P(x)-t_0 \bmod \lambda$ for many primes λ of F and choices of $t_0 \bmod \lambda$ at which $P(x)-t_0$ has distinct roots: in each case the degrees of the factors matched one of the 12 cycle structures of elements of M_{23}, out of the 632 that arise in A_{23}. Moreover, the fraction of t_0 values that yield each cycle structure was quite near to the fraction of elements of M_{23} with that cycle structure, as promised by the Chebotarev density theorem for Galois extensions of function fields. (I later learned from Mark Watkins that Samir Siksek had independently used much the same technique to find P and gather overwhelming evidence that $G_P \cong M_{23}$.)

Still this did not amount to a proof that $G_P \cong M_{23}$. However, if G_P were actually A_{23} then we would observe a very different distribution of cycle structures, which would contradict the Chebotarev theorem once the residue field of λ got large enough. In our function-field setting such a calculation turns out to be feasible thanks to Weil's proof of the Riemann hypothesis for curves over finite fields. We did this for a λ whose residue field is prime of characteristic $l = 10^8 + 7$ (the smallest 9-digit prime, which happens to lie under a degree-1 prime of F). We showed that the resulting distribution of cycle structures implies that G_P is not 5-transitive, which soon yields $G_P \cong M_{23}$ as desired.

The factorization of 10^8 polynomials mod λ was a somewhat extravagant computation (two days of CPU time in gp [13]). This is not the only way to prove that $G_P \cong M_{23}$; for example, one could do it also by numerically lifting monodromy generators to permutations of 23 preimages, as Granboulan did for the 24

[2]Michael Zieve had already obtained but not published polynomials for a few of these cases, with groups $\mathrm{PGL}_2(\mathbb{Z}/7\mathbb{Z})$ ($n = 8$), $\mathrm{P\Gamma L}_2(\mathbb{F}_8)$ ($n = 9$, both classes), and M_{11} ($n = 11$); he also calculated that there are four M_{23} polynomials up to equivalence, but was not able to exhibit such a polynomial.

preimages of an M_{24}-cover [9]. Still our technique using Chebotarev plus Weil has some advantages over the monodromy computation: while our computation took rather long to run, it was very easy to code, whereas the monodromy calculation would require some careful estimates to guarantee that the precision was sufficient to obtain the correct permutations; and our technique works also for Galois groups of extensions in positive characteristic. This approach also raises the theoretical question of how large a residue field is necessary: perhaps it can be shown that the counts over a field of size much smaller than 10^8 would have sufficed.

In the next section we exhibit F and $P \in F[x]$ and give some details on its calculation. In the following section we report on the results of our computation mod λ, use them to prove that $G_P \ncong A_{23}$, and deduce that a polynomial P_1 satisfies $G_{P_1} \cong M_{23}$ if and only if P_1 is equivalent to the image of our P under one of the four embeddings of F into $\mathbb{C}$.

2. Computation of P

Suppose $G_P \cong M_{23}$. By [12], the map $P : \mathbb{P}^1 \to \mathbb{P}^1$ is branched above only three points, with orders 23 (at $t = \infty$), 2, and 4. The group M_{23} contains only one conjugacy class of order 2 and one of order 4. The corresponding monodromy generators γ_2 and γ_4 must have $\gamma_2\gamma_4$ of order 23. Up to conjugation in M_{23}, there are four such pairs (γ_2, γ_4), two for each of the two conjugacy classes of elements of order 23 in M_{23}, and in each case γ_2 and γ_4 generate M_{23}. Since M_{23} is its own normalizer in S_{23}, we conclude that there are four equivalence classes of M_{23} polynomials, each defined over a number field F containing $\mathbb{Q}(\sqrt{-23})$ with degree 1 or 2. We eventually found that F is the dihedral quartic field of discriminant $3 \cdot 23^3$ generated by a root of $g^4 + g^3 + 9g^2 - 10g + 8$, which indeed contains the square roots $\pm(2g^3 + 4g^2 + 16g - 7)/3$ of -23.

The permutations γ_2 and γ_4 of 23 objects have cycle structures $1^7 2^8$ and $1^3 2^2 4^4$. Thus P is equivalent to a monic polynomial with two double and four quadruple roots. Then, if τ is the value of P at its finite critical points other than zeros, we can write

$$P = P_2^2 P_3 P_4^4 = P_7 P_8^2 + \tau, \tag{1}$$

where the P_i $(i = 2, 3, 4, 7, 8)$ are pairwise coprime monic polynomials of degree i, and τ is a nonzero constant. It may seem that we have 10 coefficients to determine: the $2+3+4$ non-leading coefficients of P_2, P_3, P_4, together with τ. We can reduce this to 8 variables using the remaining equivalences (translate x, and multiply x by some nonzero μ and divide each P_i by μ^i). One further variable is eliminated using a familiar[3] differentiation trick: dP/dx has leading term $23x^{22}$ and is a

[3]The earliest published references I know of are [6; 2], but the trick must have been known and used long before that.

multiple of $P_2P_4^3P_8$, so must equal $23P_2P_4^3P_8$; hence

$$P_8 = \frac{1}{23}\frac{dP/dx}{P_2P_4^3} = \frac{1}{23}\big(2P_2'P_3P_4 + P_2P_3'P_4 + 4P_2P_3P_4'\big). \tag{2}$$

Still the remaining nonlinear equations are too complicated to solve directly by techniques such as Gröbner bases, especially since they do not distinguish between M_{23}- and A_{23}-covers.

Instead we use the following strategy. Suppose the solution is defined over a number field F with a prime π of small residue field at which the cover $P : \mathbb{P}^1 \to \mathbb{P}^1$ has good reduction. We can then find our cover mod π by exhaustive search. An arbitrary lift to the π-adic numbers is then an approximate solution, which can be improved by a multivariate Newton iteration. Once we have the solution to high enough π-adic precision, we can recognize it as an F-rational point by lattice reduction, and verify that it satisfies the equations exactly.

For a general system of nonlinear equations we could not know in advance which π satisfy the condition of good reduction. In our setting, we are seeking a "Belyi map" (a cover of $\mathbb{P}^1$ ramified only above three points), so Beckmann's theorem [1] gives a sufficient condition: if the characteristic of the residue field of π does not divide the order of the Galois group then the cover has good reduction at π. But we do not know F in advance, and thus do not know which residue fields arise. We therefore tried small prime fields $\mathbb{Z}/p\mathbb{Z}$ in the hope that one would work. But searches over $(\mathbb{Z}/p\mathbb{Z})^7$ became ever longer without finding the desired cover. For example, a search mod 13 (the smallest prime not dividing $|M_{23}|$) found only

$$P_2 = x^2 - 3x - 6, \quad P_3 = x^4 - 4x - 4, \quad P_4 = x^4 + 5x^2 - 5x - 1$$

with $\tau = 5$; but the resulting $P = P_2^2P_3P_4^4$ cannot have Galois group M_{23} because there are $t_0 \neq 0, 5$ for which the factorization of $P - t_0$ mod 13 has degrees not seen in any of the M_{23} cycle structures — for instance, $P - 1$ has an irreducible factor of degree 19. In retrospect we know there is no M_{23} polynomial over $\mathbb{Z}/13\mathbb{Z}$, because F has no prime of degree 1 above 13 (even though 13 does split in the quadratic subfield $\mathbb{Q}(\sqrt{-23})$).

To bring larger p within reach, we applied the following refinement. For $j \geq 0$ and any $Q \in \mathbb{C}[x]$, denote by $c_j(Q)$ the x^j coefficient of Q; for example $c_i(P_i) = 1$ for each $i = 2, 3, 4, 7, 8$. For any monic P_2, P_3, P_4, let R be the remainder when P_{23} is divided by P_8^2, where P_{23} and P_8 are defined by (1) and (2). Then R has degree $\deg(P_8^2) - 1 = 15$ generically, but must vanish at the desired solution. We noticed that if we hold all but $c_0(P_4)$ and $c_1(P_4)$ fixed then $c_{15}(R)$ and $c_{14}(R)$ are polynomials of degree only 2 in $c_0(P_4)$ and of degree 3 in $c_1(P_4)$; in fact, $c_{15}(R)$ and $c_{14}(R)$ have degrees 2 and 3 respectively in $(c_0(P_4), c_1(P_4))$ together. We could have solved the simultaneous equations $c_{15}(R) = c_{14}(R) = 0$ in

$(c_0(P_4), c_1(P_4))$, reducing the search from $O(p^7)$ to $O(p^5)$ but with quite a large O-constant. Instead we opted for the following strategy, which is still $O(p^7)$ but with a much smaller constant. Having fixed all but $c_0(P_4)$ and $c_1(P_4)$, compute R at the 12 sample points with $c_0(P_4) = 0, 1, 2$ and $c_1(P_4) = 0, 1, 2, 3$, and then use the fact that both $c_{15}(R)$ and $c_{14}(R)$ are quadratic in $c_0(P_4)$ and cubic in $c_1(P_4)$ to recursively evaluate them at all other choices of $c_0(P_4)$ and $c_1(P_4)$. If both vanish, test whether $\deg(R) = 0$. This way, instead of computing p^2 polynomial remainders we need on average only 13: twelve sample points, and one more for the expected number of solutions of $c_{15}(R) = c_{14}(R) = 0$.

We implemented this search in gp (which we used also for the earlier $O(p^7)$ method), and finally succeeded at $p = 29$. We assumed that $c_2(P_3) = 0$, and that $c_0(P_3) = c_1(P_3)$ if both $c_0(P_3)$ and $c_0(P_1)$ are nonzero; every choice of P_2, P_3, P_4 with $c_0(P_3)c_1(P_3) \neq 0$ is equivalent to exactly one satisfying these conditions. (One can also make a unique choice if $c_0(P_3) = 0$ or $c_1(P_3) = 0$, but here this was not necessary.) The search took 46 CPU-hours, compressed to less than five hours by running on 10 heads in parallel, which is an order of magnitude smaller than the time to compute some 29^7 polynomial remainders. The resulting list of solutions contained two for which every $P(x) - t_0$ has a factorization consistent with $G_P \cong M_{23}$. One of these was

$$P_2 = x^2 - x - 3, \quad P_3 = x^3 - 3x - 3, \quad P_4 = x^4 - 3x^3 - 11x^2 + 13x + 7$$

with $\tau = 5$. Lifting to $\mathbb{Z}/p^{128}\mathbb{Z}$ (while retaining the conditions $c_2(P_3) = 0$ and $c_0(P_3) = c_1(P_3)$) gave more than enough precision to identify all the coefficients as elements of the quartic field $F = \mathbb{Q}[g]/(g^4 + g^3 + 9g^2 - 10g + 8)$.

These elements of F are quite complicated because of the normalization $c_0(P_3) = c_1(P_3)$. Once we have found one choice of $P_2, P_3, P_4 \in F[x]$ that works, we can find equivalent but simpler ones by removing this normalization and the spurious bad reduction that it entails. One reasonably simple choice we found (dropping also the condition that the P_i be monic) is as follows:

$$\begin{aligned}
P_2 &= (8g^3 + 16g^2 - 20g + 20)x^2 - (7g^3 + 17g^2 - 7g + 76)x \\
&\quad - 13g^3 + 25g^2 - 107g + 596; \\
P_3 &= 8(31g^3 + 405g^2 - 459g + 333)x^3 + (941g^3 + 1303g^2 - 1853g + 1772)x \\
&\quad + 85g^3 - 385g^2 + 395g - 220; \\
P_4 &= 32(4g^3 - 69g^2 + 74g - 49)x^4 + 32(21g^3 + 53g^2 - 68g + 58)x^3 \\
&\quad - 8(97g^3 + 95g^2 - 145g + 148)x^2 + 8(41g^3 - 89g^2 - g + 140)x \\
&\quad - 123g^3 + 391g^2 - 93g + 3228.
\end{aligned}$$

With this choice,

$$\tau = \frac{2^{38}3^{17}}{23^3}(47323g^3 - 1084897g^2 + 7751g - 711002),$$

the last factor having norm $2^{27}3^{23}5^{10}$.

3. Proof of Gal$(P(x)-t) \cong M_{23}$

We chose the degree-1 prime λ of F above the rational prime $l = 10^8 + 7$ at which $g \equiv 36436770 \bmod l$. We reduced $P \bmod \lambda$ to obtain a polynomial $\overline{P}$ with coefficients in $F_\lambda = \mathbb{Z}/l\mathbb{Z}$, and factored $\overline{P} - t_0$ for each of the $l-2$ values of $t_0 \bmod l$ for which $\overline{P} - t_0$ has no repeated roots. In each case the degrees of the irreducible factors, and thus the cycle structure of the action of Frobenius at $t = t_0$, agreed with the cycle structure of one or two of the conjugacy classes of M_{23}. Table 1 lists the key information for each class or pair of classes $c \subset M_{23}$, including the difference between the expected and the actual number of occurrences of c's cycle structure. The agreement is quite close: the discrepancy never exceeds twice the square root of the expected value.

In particular, because each of the M_{23} cycle structures occurs (and $G_P \subseteq A_{23}$ because $\mathrm{disc}_x(P(x)-t)$ is a square) we know that G_P is a transitive subgroup of A_{23} containing elements of order p for each of the prime factors $p = 2, 3, 5, 7, 11, 23$

			Occurrences		
ATLAS label	Cycle structure	$\lvert c\rvert/\lvert M_{23}\rvert$	Expected	Actual	Δ
1A	1^{23}	$1/\lvert M_{23}\rvert$	10	9	−1
2A	$1^7 2^8$	1/2688	37202	37235	33
3A	$1^5 3^6$	1/180	555556	556547	991
4A	$1^3 2^2 4^4$	1/32	3125000	3123317	−1683
5A	$1^3 5^4$	1/15	6666667	6665816	−851
6A	$1\,2^2 3^2 6^2$	1/12	8333334	8329354	−3980
7A, 7B	$1^2 7^3$	2/14	14285715	14290600	4885
8A	$1\,2\,4\,8^2$	1/8	12500001	12493007	−6994
11A, 11B	$1\,11^2$	2/11	18181819	18185450	3631
14A, 14B	2 7 14	2/14	14285715	14289505	3790
15A, 15B	3 5 15	2/15	13333334	13331689	−1645
23A, 23B	23	2/23	8695653	8697476	1823

Table 1. Data on conjugacy classes. For each class or pair of classes $c \subset M_{23}$, we list the ATLAS label [5, p. 71], the cycle structure, the fraction $|c|/|M_{23}|$, the integer nearest to $(|c|/|M_{23}|)(l-2)$ (which is the expected number of occurrences of this cycle structure), the actual number of times it appeared, and the difference between the actual and expected counts.

of $|M_{23}| = 2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23 = 10200960$. This shows that G_P is either M_{23} or A_{23}.

One could try various strategies for deducing $G_P \ncong A_{23}$ from the counts in Table 1. The following approach was the one that worked most easily. We shall take C_0 and C_1 to be the projective t- and x-lines in the following general setup.

Suppose C_1/C_0 is a degree-n covering of curves over some finite field F_λ. Let $\widetilde{C}$ be the Galois closure, with Galois group $G \subseteq S_n$. Assume that G is k-transitive. Let G_k be the stabilizer of a k-element *set*, so the action of G_k on that set gives a surjective homomorphism $G_k \to S_k$ whose kernel is the k-point stabilizer; write $C_k = \widetilde{C}/G_k$, so C_k/C_0 is a cover of degree $\binom{n}{k}$. If the cover C_1/C_0 is given by a polynomial Q of degree n, then with finitely many exceptions a point of C_k corresponds to a degree-k factor of a specialization of Q.

Let N_k be the number of F_λ-rational points of C_k. For an unramified F_λ-rational point t_0 on C_0, let $N_k(t_0)$ be the number of F_λ-rational points of C_k lying over t_0. We next express $N_k(t_0)$ in terms of the Galois structure of the preimage of t_0 in C_1. Let ϕ be the Frobenius permutation of the preimage of t_0 in C_1.

Lemma. *Let $c_1, c_2, \ldots, c_m$ (with $\sum_{i=1}^m c_m = n$) be the cycle lengths of ϕ. Then $N_k(t_0)$ is the X^k coefficient of the polynomial $\prod_{i=1}^m (1 + X^{c_i})$.*

Proof. A k-element subset of the preimage of t_0 yields a rational point of C_k if and only if it is taken to itself by ϕ; equivalently, if and only if it is the union of orbits of ϕ. Since these orbits have sizes c_i, the expansion of $\prod_{i=1}^m (1 + X^{c_i})$ yields a sum of 2^m monomials, with each monomial X^k corresponding to a k-element subset. □

We now take C_0 and C_1 to be the t- and x-lines. Then $G = G_P$ by Beckmann's criterion [1] (since l is too large to be a factor of $|G|$ even if $G = A_{23}$). Using the entries in Table 1, we find for each $k = 1, 2, \ldots, 22$ the sum of $\prod_{i=1}^m (1 + X^{c_i})$ over the $l - 2$ unramified points t_0. The sum is invariant under $k \leftrightarrow n - k$, so we need only tabulate up to $k = 11$. In each case we write $\sum_{t_0} N_k(t_0) = Al - B$ with $A \in \mathbb{Z}$ minimizing $|B|$; the results are given in Table 2.

In each case $Al - B$ is a lower bound for N_k, with the difference coming from the counts above the three ramified points. If G acts k-transitively then C_k is an

k	A	B	k	A	B	k	A	B
1	1	10	5	2	10892	9	5	487620
2	1	6592	6	3	60120	10	5	742744
3	1	19784	7	4	109978	11	7	883854
4	1	2326	8	5	243430			

Table 2. Integers A and B such that $\sum_{t_0} N_k(t_0) = Al - B$, with $|B|$ minimal.

irreducible curve, and then the Weil bound gives $|N_k - (l+1)| \le 2l^{1/2} g(C_k)$. Table 2 suggests that this might happen for $k \le 4$ but not for $k = 5$ (and indeed C_5 has two components, one for each of the orbits of the action of M_{23} on 5-element subsets). We next prove that G is not 5-transitive by bounding $g(C_5)$. If $G_{\overline{P}} = A_{23}$ then C_k has genus at most

$$1 + \frac{1}{2}\left(1 - \frac{1}{2} - \frac{1}{4} - \frac{1}{23}\right)[C_k : C_0] = 1 + \frac{1}{2}\,\frac{19}{92}\binom{23}{k}$$

by the Riemann-Hurwitz formula. For $k = 5$ this gives $27805/8$, so $g(C_5) < 3476$. Therefore

$$|N_5 - (l+1)| < 2l^{1/2} \cdot 3476 < 7 \cdot 10^7. \tag{3}$$

But the $k = 5$ row of Table 2 gives

$$N_5 - (l+1) \ge l - 10893 > 9 \cdot 10^7, \tag{4}$$

even without including the preimages of the ramified points. The conflict between inequalities (3) and (4) refutes the hypothesis that $G_P = A_{23}$ and completes the proof that $G_P \cong M_{23}$. □

Acknowledgments

I thank Michael Zieve for telling me of the remaining M_{23} case in Müller's list, and of the earlier work of himself and of Cassou-Noguès and Couveignes on the other groups. I also thank: the referee, for a careful reading resulting in several local corrections and improvements; John Voight, for the reference [2]; Mark Watkins, for apprising me of Siksek's independent work on this problem; and Watkins, Voight, and Zieve also for other helpful correspondence on this and similar problems.

This research was supported in part by NSF grants DMS-0501029 and DMS-1100511.

References

[1] Sybilla Beckmann, *Ramified primes in the field of moduli of branched coverings of curves*, J. Algebra **125** (1989), no. 1, 236–255. MR 90i:11063

[2] Bryan Birch, *Noncongruence subgroups, covers and drawings*, in Schneps [14], 1994, pp. 25–46. MR 95k:11055

[3] B. J. Birch and H. P. F. Swinnerton-Dyer, *Note on a problem of Chowla*, Acta Arith. **5** (1959), 417–423. MR 22 #4675

[4] Pierrette Cassou-Noguès and Jean-Marc Couveignes, *Factorisations explicites de $g(y) - h(z)$*, Acta Arith. **87** (1999), no. 4, 291–317. MR 99m:11023

[5] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *ATLAS of finite groups: maximal subgroups and ordinary characters for simple groups*, Oxford University Press, Eynsham, 1985. MR 88g:20025

[6] Noam D. Elkies, *ABC implies Mordell*, Internat. Math. Res. Notices (1991), no. 7, 99–109. MR 93d:11064

[7] Michael D. Fried, Shreeram S. Abhyankar, Walter Feit, Yasutaka Ihara, and Helmut Voelklein (eds.), *Recent developments in the inverse Galois problem*: *Papers from the Joint Summer Research Conference held at the University of Washington*, *Seattle*, *Washington*, *July* 17–23, 1993, Contemporary Mathematics, no. 186, American Mathematical Society, Providence, RI, 1995. MR 96c:00033

[8] Michael D. Fried and R. E. MacRae, *On the invariance of chains of fields*, Illinois J. Math. **13** (1969), no. 1, 165–171. MR 39 #179

[9] Louis Granboulan, *Construction d'une extension régulière de $\mathbb{Q}(T)$ de groupe de Galois M_{24}*, Experiment. Math. **5** (1996), no. 1, 3–14. MR 98c:12006

[10] David Hilbert, *Ueber die Irreducibilität ganzen rationaler Functionen mit ganzzahligen Coefficienten*, J. Reine Angew. Math. **110** (1892), 104–129.

[11] B. Heinrich Matzat, *Konstruktion von Zahl- und Funktionenkörpern mit vorgegebener Galoisgruppe*, J. Reine Angew. Math. **349** (1984), 179–220. MR 85j:11164

[12] Peter Müller, *Primitive monodromy groups of polynomials*, in Fried et al. [7], 1995, pp. 385–401. MR 96m:20004

[13] PARI Group, *PARI/GP* (*version* 2.4.3), 2011. http://pari.math.u-bordeaux.fr/

[14] Leila Schneps (ed.), *The Grothendieck theory of dessins d'enfants*: *Papers from the Conference on Dessins d'Enfant held in Luminy*, *April* 19–24, 1993, London Mathematical Society Lecture Note Series, no. 200, Cambridge University Press, 1994. MR 95f:11001

[15] Jean-Pierre Serre, *Topics in Galois theory*, Research Notes in Mathematics, no. 1, Jones and Bartlett Publishers, Boston, 1992. MR 94d:12006

NOAM D. ELKIES: elkies@math.harvard.edu
Department of Mathematics, Harvard University, Cambridge, MA 02138, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.369

msp

Experiments with the transcendental Brauer-Manin obstruction

Andreas-Stephan Elsenhans and Jörg Jahnel

We report on some experiments and theoretical investigations concerning weak approximation and the transcendental Brauer-Manin obstruction for Kummer surfaces of certain products of elliptic curves.

1. Introduction

Weak approximation. Consider a geometrically integral, projective variety S over the field $\mathbb{Q}$ of rational numbers. We say that S *fulfills weak approximation* when the following is true: For every finite set $\{p_1, \ldots, p_l\}$ of prime numbers and every vector

$$(x_0, x_1, \ldots, x_l) \in S(\mathbb{R}) \times S(\mathbb{Q}_{p_1}) \times \cdots \times S(\mathbb{Q}_{p_l}),$$

there exists a sequence of $\mathbb{Q}$-rational points that simultaneously converges to x_i in the p_i-adic topology for $i = 1, \ldots, l$ and to x_0 with respect to the real topology. In a more formal language, this means that the set $S(\mathbb{Q})$ of the rational points on S is dense in the set $S(\mathbb{A}_{\mathbb{Q}})$ of all adelic points.

Even for Fano varieties, which are generally expected to have many rational points, weak approximation is not always fulfilled. Well-known counterexamples are due to Sir Peter Swinnerton-Dyer [26], L. J. Mordell [20], J. W. S. Cassels and M. J. T. Guy [4], and many others.

For varieties of intermediate type — K3 surfaces, for example — the situation is yet more obscure. In fact, proving the much weaker statement that $\#S(\mathbb{Q}) = \infty$ is usually a formidable task in its own [18; 17]. It seems therefore that proving weak approximation, even for a single K3 surface, is presently out of reach and that experiments are asked for.

MSC2010: primary 11D41; secondary 11Y50, 11G35, 14J28.
Keywords: Kummer surface, weak approximation, transcendental Brauer-Manin obstruction, multivariate paging.

Obstructions and colorings. To test weak approximation experimentally is, however, an ill-posed problem, at least from the strictly formal point of view. The reason is that weak approximation is not a finite phenomenon. It is strongly infinite in nature.

An interesting situation occurs when a certain "obstruction" is responsible for the failure of weak approximation. This means that $S(\mathbb{Q}_p)$ breaks somehow regularly into open-closed subsets, each of which behaves uniformly as far as approximation by $\mathbb{Q}$-rational points is concerned. As $S(\mathbb{Q}_p)$ is compact, it is clear that finitely many subsets $U_1, \ldots, U_k \subset S(\mathbb{Q}_p)$ will suffice. When such a behavior appears, we speak of a *coloring* and call the subsets the *colors* of $S(\mathbb{Q}_p)$.

The Brauer-Manin obstructions. It is well-known that a class $\alpha \in \mathrm{Br}(S)$ in the Grothendieck-Brauer group of S induces such a coloring. For a point $x \in S(\mathbb{Q}_p)$, its color is obtained as $\mathrm{inv}_{\mathbb{Q}_p}(\alpha|x) \in \mathbb{Q}/\mathbb{Z}$. If α is of order N then not more than N colors may occur.

As a result, a failure of weak approximation may appear. Indeed, for a $\mathbb{Q}$-rational point x one must have $\sum_p \mathrm{inv}_{\mathbb{Q}_p}(\alpha|x) = 0$, but the same need not be true for an adelic point. This phenomenon is called the Brauer-Manin obstruction [19].

There is a canonical filtration on $\mathrm{Br}(S)$, which gives rise to a distinction between *algebraic* and *transcendental* Brauer classes. Correspondingly, there are the algebraic and the transcendental Brauer-Manin obstructions.

The algebraic Brauer-Manin obstruction is rather well understood. At least on $S(\mathbb{Q}_p)_{\mathrm{good}} \subseteq S(\mathbb{Q}_p)$, the p-adic points with good reduction, it yields extremely regular colorings [5; 6; 11]. For example, a coloring by two colors is possible only when there is an unramified two-sheeted covering $\pi : X \to S(\mathbb{Q}_p)_{\mathrm{good}}$. The two colors are then given by the subsets

$$\{x \in S(\mathbb{Q}_p) \mid \pi^{-1}(x) = \varnothing\} \quad \text{and} \quad \{x \in S(\mathbb{Q}_p) \mid \#\pi^{-1}(x) = 2\}.$$

Explicit computations of the algebraic Brauer-Manin obstruction have been done for many classes of varieties. Most of the examples were Fano. For instance, we gave a systematic treatment of the (algebraic) Brauer-Manin obstruction for cubic surfaces in [9; 10]. Concerning K3 surfaces, computations for diagonal quartic surfaces are provided by M. Bright [3]. Furthermore, it is known that there is no algebraic Brauer-Manin obstruction on a generic Kummer surface, or in the generic case of a Kummer surface associated to the product of two elliptic curves [25, Proposition 1.4(ii)].

The transcendental Brauer-Manin obstruction. The transcendental Brauer-Manin obstruction is much less understood and seems to be by far more difficult, at least at present. Historically, the first example of a variety where weak approximation is violated due to a transcendental Brauer class was constructed by D. Harari [12].

Concerning K3 surfaces, the available literature is still rather small. The interested reader is encouraged to consult the articles [13; 14; 15; 16; 22; 24; 27], at least in order to recognize the enormous efforts made by the authors. For example, the entire Ph.D. thesis of Th. Preu is devoted to the computation of the transcendental Brauer-Manin obstruction for single diagonal quartic surface.

An exceptional case, which seems to be a bit more accessible, is provided by the Kummer surfaces $S := \mathrm{Kum}(E \times E')$ for two elliptic curves E and E'. Here the Brauer group, which is typically purely transcendental, was described in detail by A. N. Skorobogatov and Yu. G. Zarhin in [25].

The present article. For this reason, in the present article we will deal with Kummer surfaces, defined over $\mathbb{Q}$, of the type described in the preceding paragraph. To keep the theory simple, we will restrict ourselves to the case that both curves have their full 2-torsion defined over the base field. We may start with equations for the elliptic curves of the form

$$E : y^2 = x(x-a)(x-b) \quad \text{and} \quad E' : y^2 = x(x-a')(x-b'),$$

for $a, b, a', b' \in \mathbb{Q}$. Then $S := \mathrm{Kum}(E \times E')$ is a double cover of $\mathbf{P}^1 \times \mathbf{P}^1$, an affine chart of which is given by the equation

$$z^2 = x(x-a)(x-b)u(u-a')(u-b'). \tag{1}$$

The goal of the article is to report on our experiments and theoretical investigations concerning weak approximation and the transcendental Brauer-Manin obstruction for Kummer surfaces of this particular type.

Remark 1.1. To be precise, Equation (1) defines a model of the Kummer surface with 16 singular points of type A_1. In the minimal regular model, the singularities are replaced by projective lines. As $\mathrm{Br}(\mathbf{P}^1_k) = \mathrm{Br}(k)$, the evaluation of a Brauer class on a projective line is automatically constant. Thus, we may work as well with the singular model.

The results. Among the Kummer surfaces of type (1) for integers a, b, a', b' of absolute value at most 200, we determined all those for which there is a transcendental Brauer-Manin obstruction arising from a 2-torsion Brauer class.

We found that there are exactly 3418 such surfaces having a nontrivial 2-torsion Brauer class. In three cases, this class was algebraic. Moreover, we identified the adelic subsets of the surfaces where the Brauer class gives no obstruction. On only six of the surfaces, it happened that no adelic point was excluded.

On the other hand, we developed a memory-friendly point searching algorithm for Kummer surfaces of the form above. The sets of $\mathbb{Q}$-rational points found turned out to be compatible with the idea that the Brauer-Manin obstruction might be the only obstruction to weak approximation.

2. The transcendental Brauer group

Generalities. The cohomological Grothendieck-Brauer group of an algebraic variety S over a field k is equipped with a canonical three-step filtration, defined by the Hochschild-Serre spectral sequence.

(i) $\mathrm{Br}_0(S) \subseteq \mathrm{Br}(S)$ is the image of $\mathrm{Br}(k)$ under the natural map. When S has a k-rational point, we have $\mathrm{Br}_0(S) \cong \mathrm{Br}(k)$; when k is a number field, the existence of an adelic point suffices. The group $\mathrm{Br}_0(S)$ does not contribute to the Brauer-Manin obstruction.

(ii) The quotient $\mathrm{Br}_1(S)/\mathrm{Br}_0(S)$ is isomorphic to $H^1(\mathrm{Gal}(k^{\mathrm{sep}}/k), \mathrm{Pic}(S_{k^{\mathrm{sep}}}))$. This subquotient is called the *algebraic* part of the Brauer group. For k a number field, it is responsible for the algebraic Brauer-Manin obstruction.

(iii) Finally, $\mathrm{Br}(S)/\mathrm{Br}_1(S)$ injects into $\mathrm{Br}(S_{k^{\mathrm{sep}}})$. This quotient is called the *transcendental* part of the Brauer group. Nevertheless, every Brauer class that is not algebraic is usually said to be transcendental. For k a number field, the corresponding obstruction is a transcendental Brauer-Manin obstruction.

When S is the Kummer surface corresponding to the product of two elliptic curves, the Brauer group of S is well understood due to the work of A. N. Skorobogatov and Yu. G. Zarhin [25]. For us, the proposition below will be sufficient.

Notation. We will denote the 2-torsion part of an abelian group A by A_2.

Proposition 2.1 (Skorobogatov and Zarhin). *Let*

$$E : y^2 = x(x-a)(x-b) \quad \textit{and} \quad E' : v^2 = u(u-a')(u-b')$$

be elliptic curves over a field k of characteristic zero, and let $S := \mathrm{Kum}(E \times E')$ be the corresponding Kummer surface. Suppose that the 2-torsion points of E and E' are defined over k and that $E_{\bar{k}}$ and $E'_{\bar{k}}$ are not isogenous to one another. Then

$$\mathrm{Br}(S)_2/\mathrm{Br}(k)_2 = \mathrm{im}\big(\mathrm{Br}(S)_2 \to \mathrm{Br}(S_{\bar{k}})_2\big) \cong \ker(\mu : \mathbb{F}_2^4 \to (k^*/k^{*2})^4),$$

where μ is given by the matrix

$$M_{aba'b'} := \begin{pmatrix} 1 & ab & a'b' & -aa' \\ ab & 1 & aa' & a'(a'-b') \\ a'b' & aa' & 1 & a(a-b) \\ -aa' & a'(a'-b') & a(a-b) & 1 \end{pmatrix}. \tag{2}$$

Remark 2.2. The reader should keep in mind that the matrix in Equation (2) is supposed to be giving a linear map from $\mathbb{F}_2^4$ to $(k^*/k^{*2})^4$. Thus, the entries of the matrix (although written as elements of k) represent classes of k^*/k^{*2}, and the null space of the matrix is a subspace of $\mathbb{F}_2^4$.

Proof of Proposition 2.1. The equality on the left hand side expresses the absence of algebraic Brauer classes, which is shown in [25, Proposition 3.5.i]. The isomorphism on the right is established by combining [25, Propositions 3.5.ii and 3.5.iii] with [25, Lemma 3.6]. The reader might want to compare [25, Proposition 3.7]. □

Consider the case where k is algebraically closed. Then the Kummer sequence induces a short exact sequence

$$0 \to \mathrm{Pic}(S)/2\,\mathrm{Pic}(S) \to H^2_{\text{ét}}(S, \mu_2) \to \mathrm{Br}(S)_2 \to 0.$$

We have $\dim_{\mathbb{F}_2} \mathrm{Pic}(S)/2\,\mathrm{Pic}(S) = 16 + \dim_{\mathbb{F}_2} \mathrm{NS}(E \times E')/2\,\mathrm{NS}(E \times E') = 18$ and $\dim_{\mathbb{F}_2} H^2_{\text{ét}}(S, \mu_2) = 22$. This explains why $\mathrm{Br}(S)_2 \cong \mathbb{F}_2^4$. More canonically, there are isomorphisms

$$\mathrm{Br}(S)_2 \cong H^2_{\text{ét}}(E \times E', \mu_2)/(H^2_{\text{ét}}(E, \mu_2) \oplus H^2_{\text{ét}}(E', \mu_2)) \cong \mathrm{Hom}(E[2], E'[2]).$$

Remark 2.3. If k is a field of characteristic zero, the assumption that the 2-torsion points are defined over k implies that $\mathrm{Gal}(\bar{k}/k)$ operates trivially on $\mathrm{Br}(S_{\bar{k}})_2$. We see explicitly that

$$\mathrm{Br}(S)_2/\,\mathrm{Br}(k)_2 \subsetneq \mathrm{Br}(S_{\bar{k}})_2^{\mathrm{Gal}(\bar{k}/k)} \cong \mathbb{F}_2^4,$$

in general.

Assume that k is algebraically closed. For two rational functions $f, g \in k(S)$, we denote by (f, g) the quaternion algebra

$$k(S)\{I, J\}/(I^2 - f, J^2 - g, IJ + JI)$$

over $k(S)$. Cohomologically, f and g define classes in $H^1(\mathrm{Gal}(\overline{k(S)}/k(S)), \mu_2)$ via the Kummer sequence. The Brauer class of (f, g) is the cup product of these two classes in

$$\begin{aligned} H^2(\mathrm{Gal}(\overline{k(S)}/k(S)), \mu_2^{\otimes 2}) &= H^2(\mathrm{Gal}(\overline{k(S)}/k(S)), \mu_2) \\ &\subseteq H^2(\mathrm{Gal}(\overline{k(S)}/k(S)), \overline{k(S)}^*). \end{aligned}$$

The symbol $(\cdot, \cdot)$ is thus bilinear and symmetric.

Fact 2.4. *Let k be an algebraically closed field of characteristic 0, let a, b, a', b' be elements of k, and let S be as in Proposition 2.1. Then, in terms of the canonical injection $\mathrm{Br}(S) \hookrightarrow \mathrm{Br}(k(S))$, a basis of $\mathrm{Br}(S)_2$ is given by the four quaternion algebras*

$$A_{\mu,\nu} := ((x - \mu)(x - b), (u - \nu)(u - b')),$$

for $\mu \in \{0, a\}$ and $\nu \in \{0, a'\}$. Here the standard vectors in $\mathbb{F}_2^4$ correspond to these four algebras. More precisely, e_1 corresponds to $A_{a,a'}$, e_2 to $A_{a,0}$, e_3 to $A_{0,a'}$, and e_4 to $A_{0,0}$.

Proof. This is [25, Lemma 3.6] together with [25, formula (20)]. □

Remark 2.5. Using bilinearity, we find for nine of the 15 nontrivial classes a description as a single quaternion algebra, similar to the type above. For the six classes corresponding to the vectors $(1,0,0,1)$, $(0,1,1,0)$, $(1,1,1,0)$, $(1,1,0,1)$, $(1,0,1,1)$, and $(0,1,1,1)$, we need at least two such algebras.

Observations 2.6 (Isomorphy, twisting).

(i) We may replace (a,b) by $(-a,b-a)$ or $(-b,a-b)$ without changing S, and similarly for (a',b'). Indeed, these substitutions simply come from applying the translations $\mathbf{A}^1_k \to \mathbf{A}^1_k$ given by $x \mapsto x-\mu$, for $\mu = a,b$.

(ii) It is also possible to replace (a,b,a',b') with the vector $(\lambda^2 a, \lambda^2 b, a', b')$ or the vector $(\lambda a, \lambda b, \lambda a', \lambda b')$, for $\lambda \in k$. The reason is that the twist

$$E^{(\lambda)} : \lambda y^2 = x(x-a)(x-b)$$

is isomorphic to the elliptic curve given by $Y^2 = X(X-\lambda a)(X-\lambda b)$.

One hypothesis of Proposition 2.1 is that $E_{\bar{k}}$ and $E'_{\bar{k}}$ are not isogenous. Only minor modifications to the proposition are necessary to deal with the case when these curves *are* isogenous. The isogeny causes $\mathrm{NS}(E_{\bar{k}} \times E'_{\bar{k}})/2\,\mathrm{NS}(E_{\bar{k}} \times E'_{\bar{k}})$ to have dimension higher than two, so the homomorphism $\mathbb{F}_2^4 \cong \mathrm{Hom}(E[2], E'[2]) \to \mathrm{Br}(S_{\bar{k}})_2$ is only a surjection, not a bijection.

Over a non-algebraically closed field, the situation is as follows. If E and E' are isogenous over k then $\dim_{\mathbb{F}_2} \mathrm{Pic}(S)/2\,\mathrm{Pic}(S) > 16+2 = 18$. As the additional generator evaluates trivially, it will be found in $\ker M_{aba'b'}$ [25, Lemma 3.6]. Thus, the homomorphism $\ker M_{aba'b'} \twoheadrightarrow \mathrm{Br}(S)_2/\mathrm{Br}(k)_2$ has a nontrivial kernel.

An isogeny defined over a proper field extension l/k causes the same effect over l, but not over k. As $\mathrm{Pic}(S)/2\,\mathrm{Pic}(S) \subsetneq \mathrm{Pic}(S_l)/2\,\mathrm{Pic}(S_l)$, it may, however, happen that a Brauer class is annihilated by the extension l/k; that is, that a vector in $\ker M_{aba'b'}$ describes an *algebraic* Brauer class. By the Hochschild-Serre spectral sequence, we have $H^2_{\text{ét}}(S,\mu_2)/\mathrm{Br}(k)_2 \subseteq H^2_{\text{ét}}(S_{\bar{k}},\mu_2)$. Hence, there are no other algebraic 2-torsion Brauer classes than these.

The transcendental Brauer-Manin obstruction.

Lemma 2.7. *Let k be a local field of characteristic zero, let $E : y^2 = x(x-a)(x-b)$ and $E' : v^2 = u(u-a')(u-b')$ be elliptic curves over k with all 2-torsion points defined over k, and let $S := \mathrm{Kum}(E \times E')$, given explicitly by*

$$z^2 = x(x-a)(x-b)u(u-a')(u-b'). \tag{3}$$

Let $\alpha \in \mathrm{Br}(S)_2$ be a Brauer class, represented by an Azumaya algebra over $k(S)$ of the type $\bigotimes_i A_{\mu_i,\nu_i}$. Then the local evaluation map $\mathrm{ev}_\alpha : S(k) \to \frac{1}{2}\mathbb{Z}/\mathbb{Z}$ is given

by

$$(x,u;z)\mapsto \mathrm{ev}_\alpha((x,u;z))=\sum_i\big((x-\mu_i)(x-b),(u-\nu_i)(u-b')\big)_k.$$

Here $(\cdot,\cdot)_k$ *denotes the* k*-Hilbert symbol* [2, Chapter 1, §6].

Proof. By definition, $\mathrm{ev}_\alpha((x,u;z))=\mathrm{inv}(\alpha|_{(x,u;z)})$. Further, $\alpha|_{(x,u;z)}$ is the Azumaya algebra $\bigotimes_i((x-\mu_i)(x-b),(u-\nu_i)(u-b'))$ over k. Now observe that the quaternion algebra (s,t) splits if and only if t is a norm from $k(\sqrt{s})$. This is tested by the norm residue symbol $(t,k(\sqrt{s})/k)$, which agrees with the classical Hilbert symbol $(s,t)_k$. □

Remarks 2.8.

(i) For us, the Hilbert symbol takes values in $(\frac{1}{2}\mathbb{Z}/\mathbb{Z},+)$. This differs from the classical setting, where the values are taken in $(\{\pm1\},\cdot)$.

(ii) According to Proposition 2.1, $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2\subseteq\mathbb{F}_2^4$. Further, by Fact 2.4, we have an explicit basis, which is given by Azumaya algebras; that is, for each class in $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2$, we chose a lift to $\mathrm{Br}(S)_2$. For k a local field, this lift is normalized such that $\mathrm{ev}_\alpha((\infty,\infty;\cdot))=0$. Indeed, for x close to ∞ in k, $(x-\mu)(x-b)$ is automatically a square.

The next lemma shows that the evaluation map is constant near the singular points.

Lemma 2.9. *Let* $p>2$ *be a prime number, and let* a,b,a',b' *be elements of* $\mathbb{Z}_p$ *such that* $E:y^2=x(x-a)(x-b)$ *and* $E':v^2=u(u-a')(u-b')$ *are elliptic curves that are not isogenous to each other. Suppose that*

$$\min(v_p(a),v_p(b))=\min(v_p(a'),v_p(b'))=0$$

and put

$$l:=\max(v_p(a),v_p(b),v_p(a-b),v_p(a'),v_p(b'),v_p(a'-b')).$$

Consider the surface S *over* $\mathbb{Q}_p$ *defined by* $z^2=x(x-a)(x-b)u(u-a')(u-b')$. *Then for every* $\alpha\in\mathrm{Br}(S)_2$, *the evaluation map* $S(\mathbb{Q}_p)\to\mathbb{Q}/\mathbb{Z}$ *is constant on the subset*

$$T:=\big\{(x,u;z)\in S(\mathbb{Q}_p)\;\big|\;v_p(x)<0\text{ or }v_p(u)<0\big\}\cup\bigcup_{\substack{\mu\in\{0,a,b\}\\ \nu\in\{0,a',b'\}}}\big\{(x,u;z)\in S(\mathbb{Q}_p)\;\big|\;x\equiv\mu,u\equiv\nu\ (\mathrm{mod}\ p^{l+1})\big\}$$

depicted in Figure 1.

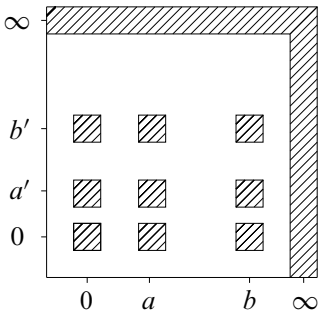


Figure 1. The set T.

Proof. It suffices to prove the lemma for α ranging over a lift to $\mathrm{Br}(S)_2$ of a basis for $\mathrm{Br}(S)_2/\mathrm{Br}(\mathbb{Q}_p)_2$; we will use the basis given in Fact 2.4. We first consider the basis element e_1, corresponding to the Hilbert symbol $((x-a)(x-b),(u-a')(u-b'))_p$. We will show that if ab and $a'b'$ and $-aa'$ are all squares, then the Hilbert symbol will be 0 on the set T.

Using the equation of the surface, we see that

$$\begin{aligned}((x-a)(x-b),(u-a')(u-b'))_p &= ((x-a)(x-b),-xu)_p \\ &= (-xu,(u-a')(u-b'))_p.\end{aligned} \tag{4}$$

Let us distinguish three cases. In all cases, we observe that a Hilbert symbol is zero when at least one of its arguments is a square.

First case. Suppose that either $v_p(x)<0$ or $v_p(u)<0$. If the first condition holds, then $(x-a)(x-b)$ is a square, while if the second condition holds, then $(u-a')(u-b')$ is a square. Thus the Hilbert symbol is 0 in this case.

Second case. Suppose that $x\equiv 0$ or $u\equiv 0 \pmod{p^{l+1}}$.

If $x\equiv 0 \pmod{p^{l+1}}$ then $(x-a)(x-b)\equiv ab \pmod{p^{l+1}}$. Since

$$v_p(ab)=v_p(a)+v_p(b)=\max(v_p(a),v_p(b))\le l,$$

the numbers $(x-a)(x-b)$ and ab belong to the same square class. Thus, if ab is a square, the Hilbert symbol will be 0.

Analogously, if $u\equiv 0 \pmod{p^{l+1}}$ then $(u-a')(u-b')\equiv a'b' \pmod{p^{l+1}}$, so that $(u-a')(u-b')$ is in the square class of $a'b'$. It follows that if $a'b'$ is a square, the Hilbert symbol will be 0.

Third case. Suppose that $x\equiv\mu$ and $u\equiv\nu \pmod{p^{l+1}}$ where $\mu\in\{a,b\}$ and $\nu\in\{a',b'\}$.

Suppose, for example that $x\equiv a \pmod{p^{l+1}}$ and $u\equiv a' \pmod{p^{l+1}}$. Then, in particular, $x\equiv a \pmod{p^{v(a)+1}}$ and $u\equiv a' \pmod{p^{v(a')+1}}$. This implies that $-xu\equiv -aa' \pmod{p^{v(a)+v(a')+1}}$ so that $-xu$ is in the square class of $-aa'$. In

particular, if $-aa'$ is a square then the Hilbert symbol is 0. The other possibilities for the residue classes of x and u yield the square classes of $-ab'$, $-ba'$, and $-bb'$, which are all trivial when ab, $a'b'$, and $-aa'$ are squares.

We see that the evaluation map is constant on the set T if and only if the vector

$$(1, ab, a'b', -aa')^t \in (\mathbb{Q}_p^*/\mathbb{Q}_p^{*2})^4$$

is zero. This is exactly the first column of the matrix $M_{aba'b'}$ given in Equation (2).

For the Hilbert symbols $((x-a)(x-b), u(u-b'))_p$, $(x(x-b), (u-a')(u-b'))_p$, and $(x(x-b), u(u-b'))_p$, the calculations are completely analogous. They lead to the second, third, and fourth columns of $M_{aba'b'}$.

Hence we see that, for a combination of Hilbert symbols, the evaluation map is constant on the set T if and only if it represents a Brauer class. □

Remarks 2.10.

(i) In Lemma 2.9, the assumption that

$$\min(v_p(a), v_p(b)) = \min(v_p(a'), v_p(b')) = 0$$

is not a restriction in view of Remark 2.16(i), below.

(ii) A result similar to Lemma 2.9 holds for $p = 2$ as well; however, the condition in the first set in the definition of T must be strengthened to $v_2(x) < -2$ or $v_2(u) < -2$, and the congruences in the other sets in the definition of T must be taken modulo 2^{l+3}. The proof is essentially the same.

Proposition 2.11. *Let k be either $\mathbb{R}$ or the field $\mathbb{Q}_p$ for a prime p. Let $E : y^2 = x(x-a)(x-b)$ and $E' : v^2 = u(u-a')(u-b')$ be elliptic curves over k with all 2-torsion points defined over k, and let $S := \mathrm{Kum}(E \times E')$ be the corresponding Kummer surface. Suppose that E and E' are not isogenous to one another, and that both E and E' have good reduction if $k = \mathbb{Q}_p$. Then for every $\alpha \in \mathrm{Br}(S)_2$, the evaluation map $\mathrm{ev}_\alpha : S(k) \to \mathbb{Q}/\mathbb{Z}$ is constant.*

Proof. First suppose that $k = \mathbb{Q}_p$. Then the assertion of the lemma is a particular case of a very general result [6, Proposition 2.4] due to J.-L. Colliot-Thélène and A. N. Skorobogatov. (Using Lemma 2.9 and elementary properties of the Hilbert symbol, one could also provide an elementary argument that is specific for the present situation.)

Next, suppose that $k = \mathbb{R}$. Without loss of generality, we may assume that $a > b > 0$ and $a' > b' > 0$. Then it will suffice to prove the assertion for representatives of e_2 and e_3, that is, for $((x-a)(x-b), u(u-b'))_{\mathbb{R}}$ and $(x(x-b), (u-a')(u-b'))_{\mathbb{R}}$.

Consider e_2. Suppose $(x, u; z)$ is an $\mathbb{R}$-rational point on the model of S given by Equation (3). If $((x-a)(x-b), u(u-b'))_{\mathbb{R}} = \frac{1}{2}$ then $(x-a)(x-b) < 0$ and

$u(u-b') < 0$. Hence, $b < x < a$ and $0 < u < b'$. But then

$$z^2 = x(x-a)(x-b)u(u-a')(u-b') < 0,$$

a contradiction. Thus, the evaluation map is constant.

For e_3, the argument is analogous. □

Algorithm 2.12.

Input: Integers a, b, a', and b'; a prime number p; and a Brauer class $\alpha \in \mathrm{Br}(S)_2$ for the surface

$$S : z^2 = x(x-a)(x-b)u(u-a')(u-b'),$$

given as a combination of Hilbert symbols.

Output: The coloring of $S(\mathbb{Q}_p)$ defined by $\mathrm{ev}_\alpha : S(\mathbb{Q}_p) \to \frac{1}{2}\mathbb{Z}/\mathbb{Z}$.

1. Calculate $l := \max(v_p(a), v_p(b), v_p(a-b), v_p(a'), v_p(b'), v_p(a'-b'))$, the bound established in Lemma 2.9.
2. Initialize three lists S_0, S_1, and S_2, the first two being empty, the third containing all triples (x_0, u_0, p) for $x_0, u_0 \in \{0, \ldots, p-1\}$. A triple (x_0, u_0, p^e) shall represent the subset
$$\{(x, u; z) \in S(\mathbb{Q}_p) \mid v_p(x - x_0) \geq e, v_p(u - u_0) \geq e\}.$$
3. Run through S_2. For each element (x_0, u_0, p^e), execute the following operations.
   (a) Test whether the corresponding set is nonempty. If not, delete the element (x_0, u_0, p^e).
   (b) If $e \geq l+1$ and $v_p(x-\mu) \geq l+1$ for some $\mu \in \{0, a, b\}$, and $v_p(u-\nu) \geq l+1$ for a $\nu \in \{0, a', b'\}$, then move (x_0, u_0, p^e) to S_0.
   (c) Test naïvely, using the elementary properties of the Hilbert symbol, whether the elements in the corresponding set all have the same evaluation. If this test succeeds then move (x_0, u_0, p^e) to S_0 or S_1, depending on whether the value is 0 or $\frac{1}{2}$.
   (d) Otherwise, replace (x_0, u_0, p^e) by the p^2 triples $(x_0 + i p^e, u_0 + j p^e, p^{e+1})$ for $i, j \in \{0, \ldots, p-1\}$.
4. If S_2 is empty then output S_0 and S_1 and terminate. Otherwise, go back to step 3.

Example 2.13. Consider the Kummer surface S over $\mathbb{Q}$ given by

$$z^2 = x(x-1)(x-25)u(u+25)(u+36).$$

Then weak approximation is violated on S.

Proof. This is caused by a transcendental Brauer-Manin obstruction. In fact, the matrix (2) is

$$M = \begin{pmatrix} 1 & 25 & 900 & 25 \\ 25 & 1 & -25 & -275 \\ 900 & -25 & 1 & -24 \\ 25 & -275 & -24 & 1 \end{pmatrix} \widehat{=} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -11 \\ 1 & -1 & 1 & -6 \\ 1 & -11 & -6 & 1 \end{pmatrix},$$

and its kernel is $\langle e_1 \rangle$. Hence there is a transcendental Brauer class on S, represented by the quaternion algebra $((x-1)(x-25), (u+25)(u+36))$.

Now the argument is completely elementary. For every $(x, u; z) \in S(\mathbb{Q}_p)$ with $z \neq 0$, one has

$$\sum_p ((x-1)(x-25), (u+25)(u+36))_p = 0,$$

according to the sum formula for the Hilbert symbol. The bad primes of the elliptic curves $y^2 = x(x-1)(x-25)$ and $y^2 = x(x+25)(x+36)$ are $2, 3, 5$, and 11. Hence, the sum is actually only over these four primes.

Our implementation of Algorithm 2.12 shows that the local evaluation map is constant at the primes 2, 3, and 11, but not at 5. Hence, 5-adic points such that $((x-1)(x-25), (u+25)(u+36))_5 = \frac{1}{2}$ may not be approximated by $\mathbb{Q}$-rational ones.

Examples of such 5-adic points include those with $(x, u) = (2, 5)$. Indeed,

$$2 \cdot (2-1) \cdot (2-25) \cdot 5 \cdot (5+25) \cdot (5+36) = -11316 \cdot 5^2$$

is a 5-adic square, but $(2-1) \cdot (2-25) = -23$ is a nonsquare and $v_5((5+25) \cdot (5+36)) = 1$ is odd. □

Remarks 2.14.

(i) The constancy of the local evaluation maps at 3 and 11 and the nonconstancy at 5 also follow from the criterion formulated as Theorem 2.19 below.

(ii) In the coloring obtained on $S(\mathbb{Q}_5)$, all the points such that $x, u \not\equiv 0 \pmod 5$ have color zero. This is rather different from the colorings typically obtained from an algebraic Brauer class. The reader should compare the situation described in [5], where, on the cone over an elliptic curve, three sets of equal sizes appear.

Normal form, ranks, asymptotics. Let k be a field, let a, b, a', and b' be elements of k^* with $a \neq b$ and $a' \neq b'$, and let S be the Kummer surface

$$z^2 = x(x-a)(x-b)u(u-a')(u-b').$$

There are two types of nontrivial Brauer classes $\alpha \in \mathrm{Br}(S)_2 / \mathrm{Br}(k)_2$.

Type 1: *α may be expressed by a single Hilbert symbol.* There are nine cases for the kernel vector of $M_{aba'b'}$. As seen in Observations 2.6(i), a suitable translation of $\mathbf{A}^1 \times \mathbf{A}^1$ transforms the surface into an isomorphic one with kernel vector e_1. Then ab, $a'b'$, and $-aa'$ are squares in k. Note that this implies that $-ba'$, $-ab'$, and $-bb'$ are squares as well.

Type 2: *To express α, two Hilbert symbols are necessary.* There are six cases for the kernel vector of $M_{aba'b'}$. A suitable translation of $\mathbf{A}^1 \times \mathbf{A}^1$ transforms the surface into an isomorphic one with kernel vector $e_2 + e_3$. Then aa', bb', and $(a-b)(a'-b')$ are squares.

Corollary 2.15. *Let p be a prime number, let a, b, a', and b' be elements of $\mathbb{Q}_p^*$ with $a \neq b$ and $a' \neq b'$, and let S be the Kummer surface*

$$z^2 = x(x-a)(x-b)u(u-a')(u-b').$$

Suppose that $v_p(a) \leq v_p(b)$, that $v_p(a') \leq v_p(b')$, and that $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \neq 0$. Then $v_p(aa')$ is even.

Proof. The assertion is that the expression

$$m := \min(v_p(a), v_p(b), v_p(a-b)) + \min(v_p(a'), v_p(b'), v_p(a'-b'))$$

is even as soon as $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \neq 0$. As m is invariant under translations as described in Observations 2.6(i), we may suppose that either e_1 or $e_2 + e_3$ lies in $\ker M_{aba'b'}$. In both cases the assertion is easily checked. Note that either minimum is adopted by at least two of the three valuations. □

Remarks 2.16.

(i) Suppose $k = \mathbb{Q}_p$. Then, by Observations 2.6(ii), we may assume without loss of generality that $a, b, a', b' \in \mathbb{Z}_p$, that $\min(v_p(a), v_p(b)) = 0$, and that $\min(v_p(a'), v_p(b')) = 0, 1$. By Corollary 2.15, the assumption that $M_{aba'b'}$ has a nontrivial kernel ensures that $\min(v_p(a'), v_p(b')) = 0$, too.

(ii) Suppose that $k = \mathbb{Q}$ and that there is a Brauer class of type 1. Reasoning as in the preceding remark, we see that we may suppose that a, b, a', and b' are integers with $\gcd(a, b) = \gcd(a', b') = 1$. Hence there is a normal form with $a > b$, with $a' < b'$, and with $a, b, -a', -b' \in \mathbb{Z} \cap \mathbb{Q}^{*2}$. Up to the involution $(a, b, a', b') \mapsto (-a', -b', -a, -b)$, this normal form is unique. The geometric interpretation of this involution is that it interchanges the two elliptic curves and twists them both by -1.

Proposition 2.17. *Let k be a field of characteristic zero, let $E : y^2 = x(x-a)(x-b)$ and $E' : v^2 = u(u-a')(u-b')$ be elliptic curves over k with all 2-torsion points defined over k, and let $S := \mathrm{Kum}(E \times E')$ be the corresponding Kummer surface. Suppose that E and E' are not isogenous to each other.*

(i) *We have* $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \leq 4$ *and* $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \neq 3$. *Further*, $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 = 4$ *is possible only when* -1 *is a square in* k.

(ii) *Suppose* $k = \mathbb{Q}_p$ *for a prime* p. *If both* E *and* E' *have potential good reduction then* $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2$ *is even.*

(iii) *If* $k = \mathbb{R}$ *then* $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 = 2$.

Proof. All of these assertions will follow from Proposition 2.1. Recall that $M_{aba'b'}$ is a matrix with entries in the $\mathbb{F}_2$-vector space k^*/k^{*2}.

Statement (i): The inequality $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \leq 4$ is clear. If the vector space had dimension three, the matrix $M_{aba'b'}$ would have column rank one. But this is impossible for a symmetric matrix having zeroes on the diagonal. Further, $\dim \mathrm{Br}(S)_2/\mathrm{Br}(k)_2 = 4$ requires $M_{aba'b'}$ to be the zero matrix. In particular, aa' and $-aa'$ both have to be squares in k. This implies that -1 is a square, too.

Statement (ii): Standard considerations (see [23, Proposition VII.5.5], for example) show that the elliptic curve given by $y^2 = x(x-\mu)(x-\nu)$ has potential good reduction if and only if $\mu/\nu \in \mathbb{Z}_p^*$ and $\mu/\nu \not\equiv 1 \pmod{p}$. This implies, in particular, that $p > 2$.

If $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2 = 0$ the assertion is trivially true, so let us assume that $\mathrm{Br}(S)_2/\mathrm{Br}(k)_2 \neq 0$. Then, by Remark 2.16(i), we may assume that the elements a, b, $a-b$, a', b', $a'-b'$ all lie in $\mathbb{Z}_p^*$. But for p-adic units, being a square in $\mathbb{Q}_p$ or not is tested by the Legendre symbol. Thus $M_{aba'b'}$ is essentially an alternating matrix with entries in $\mathbb{F}_2$. Such matrices have even rank.

Statement (iii): After applying one of the translations $\mathbf{A}^1 \times \mathbf{A}^1 \to \mathbf{A}^1 \times \mathbf{A}^1$ given by $(x,u) \mapsto (x-\mu, u-\nu)$ for $\mu \in \{0,a,b\}$ and $\nu \in \{0,a',b'\}$, we may assume that $a > b > 0$ and $a' > b' > 0$. Then

$$M_{aba'b'} = \begin{pmatrix} + & + & + & - \\ + & + & + & + \\ + & + & + & + \\ - & + & + & + \end{pmatrix}$$

has kernel $\langle e_2, e_3 \rangle$. □

Remarks 2.18. We discuss some asymptotic estimates for the number of surfaces with Brauer groups of various types.

(i) Let $N > 0$. Then the number of pairs (a,b) such that a and b are perfect squares, $a < b$, and $a, b-a < N$ is asymptotically CN for

$$C := \tfrac{1}{2}\big(\log(\sqrt{2}+1) + \sqrt{2} - 1\big).$$

Indeed, the Stieltjes integral

$$\int_1^N \big(\sqrt{x+N} - \sqrt{x}\big)\, d\sqrt{x}$$

has exactly this behavior. Assuming that isogenies are rare, we find that the number of surfaces over $\mathbb{Q}$ with integer parameters of absolute value at most N and a 2-torsion Brauer class of type 1 is asymptotically

$$\frac{1}{2}\left(\frac{6}{\pi^2}\right)^2 C^2 N^2 \approx 0.077544 N^2.$$

(ii) On the other hand, a 2-torsion Brauer class of type 2 yields a $\mathbb{Q}$-rational point on the intersection of three quadrics in $\mathbf{P}^6$. The Manin conjecture leads to the naïve expectation of growth of the type $cN \log^d N$ for some integer $d \geq 0$.

(iii) The number of all Kummer surfaces of the form considered and with parameters up to N is $O(N^4)$. Thus, only a very small fraction have a nontrivial 2-torsion Brauer class.

Even fewer surfaces should have odd torsion in their Brauer group. Indeed, for l-torsion, one must have

$$\mathrm{Hom}_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})}(E[l], E'[l]) \neq 0$$

(see [25, Proposition 3.3]). Consequently, $\#E(\mathbb{F}_p) \equiv \#E'(\mathbb{F}_p) \pmod{l}$ for every prime $p \neq l$ that is good for both E and E'. Based on this, our computations show that, up to $N = 200$, no surface has an l-torsion Brauer class for $l \geq 5$. Further, at most eight pairs of j-invariants allow a 3-torsion Brauer class.

(iv) It is possible over $\mathbb{Q}$ to have a 2-dimensional 2-torsion Brauer group. For this, in the normal form of Remark 2.16(ii), one needs that $a - b$ and $b' - a'$ are perfect squares. Further, these surfaces have four normal forms instead of two, as there are two Brauer classes of type 1. These examples correspond to pairs of Pythagorean triples, and we therefore have two Kummer surfaces, differing from each other by a twist by (-1). The asymptotics of Pythagorean triples [1] shows that there are asymptotically

$$\frac{4}{\pi^4}\log^2(1+\sqrt{2})N \approx 0.031899 N$$

surfaces over $\mathbb{Q}$ with integer parameters of absolute value at most N and a Brauer group of dimension two.

(v) Some actual numbers are listed in Table 1. For a precise description of the sample, see Section 4B below.

Trivial evaluation.

Theorem 2.19. *Let $p > 2$ be a prime number and let a, b, a', b' be nonzero elements of $\mathbb{Z}_p$ such that $a \neq b$ and $a' \neq b'$. Let S be the Kummer surface given*

Bound	Dimension 2	Dimension 1, type 1		Dimension 1, type 2	
		Total	Algebraic	Total	Algebraic
50	0	183	1	38	0
100	0	766	2	98	0
200	2	3049	3	367	0
500	12	18825	4	1457	0
1000	20	77249	8	4398	0
2000	42	305812	11	12052	0

Table 1. Number of surfaces with bounded parameters whose Brauer groups have 2-torsion of various types. The first column gives the bound N on the parameters of the surfaces we computed; see Section 4B for a precise description of the parameters allowed. The remaining columns give the number of such surfaces whose Brauer groups have 2-torsion subgroups of dimension 2, of dimension 1 and type 1, and of dimension 1 and type 2. For the 1-dimensional cases, the number of algebraic classes is listed as well.

by $z^2 = x(x-a)(x-b)u(u-a')(u-b')$. *Assume that* e_1 *is a kernel vector of the matrix* $M_{aba'b'}$ *and let* $\alpha \in \mathrm{Br}(S)_2$ *be the corresponding Brauer class.*

(i) *Suppose that either* $a \equiv b \not\equiv 0 \pmod p$ *or* $a' \equiv b' \not\equiv 0 \pmod p$. *Then the evaluation map* $\mathrm{ev}_\alpha : S(\mathbb{Q}_p) \to \mathbb{Q}/\mathbb{Z}$ *is constant.*

(ii) *If* $a \not\equiv b \pmod p$ *and* $a' \not\equiv b' \pmod p$, *and if not all four numbers are* p*-adic units, then the evaluation map* $\mathrm{ev}_\alpha : S(\mathbb{Q}_p) \to \mathbb{Q}/\mathbb{Z}$ *is nonconstant.*

Proof. First step: *Preparations.* We are interested in the Hilbert symbol

$$((x-a)(x-b), (u-a')(u-b'))_p.$$

Recall that a/b, a'/b', and $-bb'$ are all squares in $\mathbb{Q}_p$.

A $\mathbb{Q}_p$-rational point on S corresponds to a pair of points on the elliptic curves $\lambda y^2 = x(x-a)(x-b)$ and $\lambda v^2 = u(u-a')(u-b')$ for a common value of λ. The Hilbert symbol then simplifies to $(\lambda x, \lambda u)_p$.

Second step: 2*-descent.* By 2-descent (see for example [23, Proposition X.1.4]), the elliptic curve $E : Y^2 = X(X-a)(X-b)$ has a point in the square class of x if and only if the system

$$xz_1^2 - tz_2^2 = a$$
$$xz_1^2 - xtz_3^2 = b$$

is solvable. Eliminating t, we obtain $x^2 z_1^2 z_3^2 - xz_1^2 z_2^2 = axz_3^2 - bz_2^2$, which gives

$$(xz_3^2 - z_2^2)(xz_1^2 - b) = (a-b)xz_3^2.$$

Dividing by $-bxz_3^2$ yields

$$\left(1-\frac{z_2^2}{z_3^2}\frac{1}{x}\right)\left(1-\frac{z_1^2}{b}x\right)=1-\frac{a}{b}.$$

In other words, E has a point in the square class of x if and only if the equation $(1-v^2x)(1-w^2x/b)=1-a/b$ is solvable.

Third step: *Application to the Kummer surface* S. As $\lambda y^2=x(x-a)(x-b)$ is equivalent to $y'^2=\lambda x(\lambda x-\lambda a)(\lambda x-\lambda b)$ and b and $-b'$ are squares, we see that S has a point with coordinates in the square classes of x and u if and only if

$$(1-v^2\lambda x)(1-w^2x/b)=1-a/b$$
$$(1-v'^2\lambda u)(1-w'^2x/b')=1-a'/b'$$

has a solution $(v,w,v',w',\lambda)\in(\mathbb{Q}_p^*)^5$.

Proof of (i): Without loss of generality, assume that $a\equiv b\not\equiv 0 \pmod p$ and $a'/b'\in\mathbb{Z}_p$. Let $(x,u;z)\in S(\mathbb{Q}_p)$ be any point such that $z\neq 0$.

Then Lemma 2.21(i) (below) shows that $(u/b',\lambda u)_p=0$. Furthermore, by Lemma 2.21(iii), at least one of x/b and λx is a square in $\mathbb{Q}_p$. In the case $\lambda x\in\mathbb{Q}_p^{*2}$, the assertion $(\lambda x,\lambda u)_p=0$ is clearly true. If $x/b\in\mathbb{Q}_p^{*2}$ then

$$0=(u/b',\lambda u)_p=(-\lambda/b',\lambda u)_p=((\lambda x)/(-bb'),\lambda u)_p=(\lambda x,\lambda u)_p.$$

Proof of (ii): Again without loss of generality, assume that $p^2\mid a$, that b is a unit, and that $a'/b'\in\mathbb{Z}_p$. We claim that, for $\lambda=-b$, there is a point on S such that $x=p$ and $2\mid v_p(u)$, so that $\lambda u=-bu$ is a nonsquare.

Indeed, it is obvious that $-bp(p-a)(p-b)\in\mathbb{Q}_p^{*2}$. Further, by Hensel's Lemma, it suffices to find a pair $(U_1,U_2)\in\mathbb{F}_p^*\times\mathbb{F}_p^*$ of nonsquares such that $(1-U_1)(1-U_2)=1-\overline{a}'/\overline{b}'$. For this, a counting argument applies. In fact, each $U_1\in\mathbb{F}_p\setminus\{0,1,\overline{a}'/\overline{b}'\}$ uniquely determines its partner. As this set contains $(p-1)/2$ nonsquares and only $(p-5)/2$ squares, the assertion follows. □

Remarks 2.20.

(i) It might seem strange to use a descent argument over a local field. It seems to us, however, that a direct argument is neither more elegant nor shorter.

(ii) Using the descent argument above, we also recover the constancy of the evaluation map in the case of good reduction. Indeed, Lemma 2.21(ii) implies that either at least one of x/b and λx is a square, or both have even valuation. The first two cases are dealt with as above. Otherwise, λb is a square, hence $-\lambda/b'$ is a square, too, and one has to show that $2\mid v_p(\lambda u)$. But this is implied by Lemma 2.21(ii) when looking at the second equation.

Lemma 2.21. *Let $p > 2$ be a prime number, let A and B be nonzero elements of $\mathbb{Q}_p$, and let $Q \in \mathbb{Q}_p^*$ be a square. Suppose that the equation $(1-Av^2)(1-Bw^2) = 1 - Q$ is solvable in $\mathbb{Q}_p^* \times \mathbb{Q}_p^*$.*

(i) *We have $(A, B)_p = 0$.*

(ii) *If $Q \in \mathbb{Z}_p^*$ then either $A \in \mathbb{Q}_p^{*2}$, or $B \in \mathbb{Q}_p^{*2}$, or both A and B are of even valuation.*

(iii) *If $Q \in \mathbb{Z}_p^*$ and $Q \equiv 1 \pmod{p}$ then either $A \in \mathbb{Q}_p^{*2}$ or $B \in \mathbb{Q}_p^{*2}$.*

Proof. Statement (i): We have that $Av^2 + Bw^2 - AB(vw)^2$ is a square. When all three summands are of the same valuation, they must be units. The assertion is then clearly true. Otherwise, at most two of the three summands have minimal valuation. Then their sum is a square, too. According to the definition of the Hilbert symbol [2, p. 55], we have either $(A, B)_p = 0$ or $(A, -AB)_p = 0$ or $(B, -AB)_p = 0$. These three statements are equivalent to each other.

Statement (ii): We have $v_p(1 - Q) \geq 0$. On the other hand, if both A and B are nonsquares then $v_p(1 - Av^2), v_p(1 - Bw^2) \leq 0$. This implies equality, hence $Av^2, Bw^2 \in \mathbb{Z}_p$. Both must be units as $Av^2 + Bw^2 - AB(vw)^2$ is, by assumption, a square in $\mathbb{Z}_p^*$.

Statement (iii): If A and B were both nonsquares then $v_p(1 - Av^2) \leq 0$ and $v_p(1 - Bw^2) \leq 0$. As $v_p(1 - Q) > 0$, this is a contradiction. □

Experiments with Algorithm 2.12 show that surprisingly often there are nontrivial Brauer classes with trivial p-adic evaluation. This is partially explained by the following result.

Theorem 2.22. *Let $p > 2$ be a prime number and let $a, b, a', b' \in \mathbb{Q}_p$ be such that $E : y^2 = x(x-a)(x-b)$ and $E' : v^2 = u(u-a')(u-b')$ are elliptic curves. Suppose that E and E' are not isogenous to each other, and let S be the corresponding Kummer surface.*

(i) *If $\dim \mathrm{Br}(S)_2 / \mathrm{Br}(\mathbb{Q}_p)_2 \geq 2$ then there is a nonzero $\alpha \in \mathrm{Br}(S)_2$ such that ev_α is the zero map.*

(ii) *If $\dim \mathrm{Br}(S)_2 / \mathrm{Br}(\mathbb{Q}_p)_2 = 4$ then the subspace of classes with constant evaluation map is of dimension 4 when both E and E' have potential good reduction. The dimension is 3 when neither curve has potential good reduction and 2 in the mixed case.*

Proof. By Remark 2.16(i), we may assume without loss of generality that a and b lie in $\mathbb{Z}_p$ and are not both divisible by p, and that the same holds for a' and b'. The case in which both E and E' have potential good reduction has already been treated in Proposition 2.11.

Statement (ii). If neither curve has potential good reduction, we can apply a translation of $\mathbf{A}^1 \times \mathbf{A}^1$, as in Observations 2.6(i), to reduce to the case $a \equiv b \not\equiv 0 \pmod{p}$ and $a' \equiv b' \not\equiv 0 \pmod{p}$. Then, by virtue of Theorem 2.19(ii), the Brauer classes corresponding to $\langle e_1, e_2, e_3 \rangle$ have constant evaluation maps, but ev_{e_4} is nonconstant.

Further, when only E' has potential good reduction, the same arguments show that the Brauer classes corresponding to $\langle e_1, e_2 \rangle$ have constant evaluation maps, while those of e_3, e_4, and $e_3 + e_4$ are nonconstant.

Statement (i). Only the case that at least one of the curves E and E' does not have potential good reduction requires a proof. Hence, we may assume that $a, b \in \mathbb{Z}_p$ and $a \equiv b \not\equiv 0 \pmod{p}$. Then $ab \in \mathbb{Q}_p^{*2}$.

The upper left 2×2-block of $M_{aba'b'}$ is zero. If the block $\left(\begin{smallmatrix} a'b' & aa' \\ -aa' & a'(a'-b') \end{smallmatrix}\right)$ occurring in the lower left has trivial kernel then the 2×2-block in the upper right is certainly not the zero matrix. Therefore $\dim \ker M_{aba'b'} \leq 1$, a contradiction. Thus, there is a Brauer class represented by a vector from $\langle e_1, e_2 \rangle$. By Theorem 2.19(i), its evaluation map is constant. □

3. A point search algorithm for special Kummer surfaces

The surfaces we are studying are double covers of $\mathbf{P}^1 \times \mathbf{P}^1$, given by equations of the form

$$w^2 = f_{ab}(x, y) f_{a'b'}(u, v).$$

Here, f_{ab} is the binary quartic form $f_{ab}(x, y) := xy(x - ay)(x - by)$. Thus, a point $([x : y], [u : v]) \in (\mathbf{P}^1 \times \mathbf{P}^1)(\mathbb{Q})$ leads to a point on the surface if and only if the square classes of $f_{ab}(x, y)$ and $f_{a'b'}(u, v)$ coincide, or one of them is zero.

We will call the solutions with $f_{ab}(x, y)$ or $f_{a'b'}(u, v)$ zero the *trivial* solutions of the equation. Obviously, there is a huge number of trivial solutions. Our aim is to describe an efficient algorithm that searches for nontrivial solutions and does not care about the trivial ones. In its simplest version, our algorithm works as follows.

Algorithm 3.1 (Point search).

Input: Two sequences $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ of integers and a search bound $B > 0$.

Output: All solutions of the equations

$$w^2 = f_{a_i b_i}(x, y) f_{a_j b_j}(u, v)$$

for which x, y, u, and v are integers with $|x|, |y|, |u|, |v| \leq B$.

1. Compute the bound

$$L := B(1 + \max\{|a_i|, |b_i| \mid i = 1, \ldots, k\})$$

for the linear factors.

2. Store the squarefree parts of the integers in $[1, \dots, L]$ in an array T.
3. Enumerate in an iterated loop representatives for all points $[x : y] \in \mathbf{P}^1(\mathbb{Q})$ with $x, y \in \mathbb{Z}$, $|x|, |y| \le B$, and $x, y \neq 0$.
4. For each point $[x : y]$ enumerated, execute the operations below.
   (a) Run a loop over $i = 1, \dots, k$ to compute the four linear factors x, y, $x - a_i y$, and $x - b_i y$ of f_{a_i, b_i}.
   (b) Store the squarefree parts of the factors in $m_1, \dots, m_4$. (Use the table T to compute the squarefree parts.)
   (c) Put
   $$p_1 := \frac{m_1}{\gcd(m_1, m_2)} \frac{m_2}{\gcd(m_1, m_2)}$$
   $$p_2 := \frac{m_3}{\gcd(m_3, m_4)} \frac{m_4}{\gcd(m_3, m_4)}$$
   $$p_3 := \frac{p_1}{\gcd(p_1, p_2)} \frac{p_2}{\gcd(p_1, p_2)}.$$
   Thus, p_3 is a representative of the square class of $f_{a_i b_i}(x, y)$.
   (d) Store the quadruple $(x, y, i, h(p_3))$ in a list. Here, h is a hash function.
5. Sort the list by the last component.
6. Split the list into parts. Each part corresponds to a single value of $h(p_3)$. (At this point, we have detected all collisions of the hash function.)
7. Run in an iterated loop over all the collisions and check whether
   $$((x, y, i, h(p_3)), (x', y', i', h(p_3')))$$
   corresponds to a solution $([x : y], [x' : y'])$ of the equation
   $$w^2 = f_{a_i b_i}(x, y) f_{a_{i'} b_{i'}}(x', y').$$
   Output all the solutions found.

Remarks 3.2.

(i) For practical search bounds B, the first integer overflow occurs when we multiply $p_1/\gcd(p_1, p_2)$ and $p_2/\gcd(p_1, p_2)$. But we can think of this reduction modulo 2^{64} as being a part of our hash function. Note that the final check of $f_{a_i b_i}(x, y) f_{a_{i'} b_{i'}}(x', y')$ being a square can be done without multiprecision integers by inspecting the gcd's of the eight factors.

(ii) One disadvantage of Algorithm 3.1 is obvious. It requires more memory than is reasonably available by present standards. We solved this problem by the

introduction of what we call a *multiplicative paging*. This is an approach motivated by the simple additive paging as described in [8]. In addition, our memory-optimized point search algorithm is based on the following observation.

Lemma 3.3. *Let p be a good prime. Then, for each pair (x, y) with $\gcd(x, y) = 1$, at most one of the factors x, y, $(x - ay)$, and $(x - by)$ is divisible by p.* □

Algorithm 3.4 (Point search using multivariate paging).

Input: The same as in Algorithm 3.1.

Output: The same as in Algorithm 3.1.

1. Compute the bound L and the square-class representatives as in Algorithm 3.1.
2. Compute the upper bound
$$C := 2 \max\{|a_i|, |b_i| \mid i = 1, \ldots, k\}$$
for the possible bad primes.
3. Initialize an array of boolean variables of length L. Use the value `false` for the initialization. We will call this array the *markers* of the factors already treated.
4. In a loop, run over all good primes below L. Start with the biggest prime and stop when the upper bound C is reached; that is, work in *decreasing* order. For each prime p_p, execute the steps below. We call p_p the *page prime*.
   (a) Run over all multiples m of p_p not exceeding L and such that the p_p-adic valuation is odd. For each m, do the following.
      i. Check whether m is marked as already treated. In this case, continue with the next m.
      ii. Test whether x, y, $x - a_i y$, or $x - b_i y$ can represent this value. Here, use the constraints $|x|, |y| \leq B$ and $i \in \{1, \ldots, k\}$.
      iii. For each possible representation with $\gcd(x, y) = 1$, check to see whether x, y, $x - a_i y$, or $x - b_i y$ is marked as already treated. Otherwise, store the quadruples $(x, y, i, h(p_3))$ into a list.
      iv. Mark the value of m as treated and continue with the next m.
   (b) As in Algorithm 3.1, construct all solutions by inspecting the collisions of the hash function.
5. Up to now, all solutions were found such that w has at least one prime factor bigger than the bad primes bound. To get the remaining ones, use Algorithm 3.1 but skip all values of x, y that are marked as treated factors. Further, break step 4 of Algorithm 3.1 early if m_3 or m_4 is marked as treated.

Remark 3.5. The last step computes all solutions in smooth numbers — that is, points such that the square classes of $f_{ab}(x, y)$ and $f_{a'b'}(u, v)$ are smooth with respect to the bad primes bound C defined in step 2. It is an experimental observation that this step takes only a small fraction of the running time, but gives a large percentage of the solutions. The algorithm may easily be modified such that only the solutions in smooth numbers are found. For this, the markers for treated factors have to be initialized in an appropriate way.

4. Some experiments

4A. ***Coloring by covering — a search for regular colorings.*** As noted in the introduction, on various types of surfaces [3; 11], the (algebraic) Brauer-Manin obstruction leads to very regular colorings. Carrying this knowledge over to the special Kummer surfaces given by

$$S: \quad w^2 = f_4(x, y)g_4(u, v),$$

one is led to test the following: For a $\mathbb{Q}$-rational point with $w \neq 0$, write $\lambda w_1^2 = f_4(x, y)$ and $\lambda w_2^2 = g_4(u, v)$ and expect the color to be given by the square class of λ.

For p-adic points, this defines a coloring with four or eight colors, depending on whether $p > 2$ or $p = 2$. At the infinite place, the color is given by the sign of λ. Motivated by [3; 11], we assume that the p-adic color of a rational point has a meaning only when p divides the conductor of one of the elliptic curves used to construct S. Further, we restricted ourselves to the square classes of even p-adic valuation (for the primes of bad reduction). This does not exclude all rational points reducing to the singular locus at a bad prime.

Thus, we get a coloring of the $\mathbb{Q}$-rational points with 2^{k+1} colors for a surface with k relevant odd primes. Weak approximation would imply that the color map is a surjection. In the case of a visible obstruction, we would expect that at most half of the possible colors are in the image of the color map.

For a systematic test, we used the 184 elliptic curves with odd conductor and $|a|, |b| < 100$. This led to 16,836 surfaces. Table 2 gives an overview of the number of colors that occurred. The table indicates that our result is negative: It seems that there is no obstruction factoring over such a coloring. We expect that one would find $\mathbb{Q}$-rational points of all colors for a sufficiently large search bound.

On one core of an Intel Core 2 Duo E8300 processor, the running times were 18.5 hours for search bound 30,000 and 275 hours for search bound 100,000, but only 51 minutes for smooth solutions with respect to a bad prime bound of 200 and bound 100,000.

#Bad primes	2	3	4	5	6	7	8
#Surfaces	4	182	1678	5777	7409	1726	60
#Colors	8	16	32	64	128	256	512
$B = 1000$	8	15–16	26–32	32–64	33–127	31–157	27– 81
$B = 3000$	8	16	30–32	49–64	67–128	81–226	92–192
$B = 10000$	8	16	32	57–64	93–128	142–254	207–352
$B = 30000$	8	16	32	62–64	109–128	196–256	303–474
$B = 100000$	8	16	32	64	121–128	232–256	387–505
$B = 10000$, smooth	8	16	31–32	54–64	79–128	99–236	113–197
$B = 30000$, smooth	8	16	32	59–64	92–128	146–253	161–300
$B = 100000$, smooth	8	16	32	61–64	108–128	185–256	230–381

Table 2. Number of colors attained by $\mathbb{Q}$-rational points of bounded height on Kummer surfaces of products of elliptic curves, classified by the number of bad primes and the search bound B. The second row of the table indicates the number of surfaces we analyzed with the given number of bad primes. For each number of bad primes and each search bound B, we list the lowest and highest number of colors attained by $\mathbb{Q}$-rational points of height at most B, ranging over the surfaces with the given number of bad primes. For the rows in which the search bound is annotated with the word "smooth", we consider only rational points that are smooth in the sense of Remark 3.5.

4B. ***Investigating the Brauer-Manin obstruction — a sample.*** We determined all Kummer surfaces of the form

$$z^2 = x(x-a)(x-b)u(u-a')(u-b'),$$

with integer parameters of absolute value at most 200, that have a transcendental 2-torsion Brauer class.

More precisely, we determined all $(a, b, a', b') \in \mathbb{Z}^4$ such that

$$\begin{aligned} &\gcd(a,b) = 1, \qquad a > b > 0, \qquad a - b \leq 200, \qquad b \leq 200, \\ &\gcd(a',b') = 1, \qquad a' < b' < 0, \qquad a' - b' \geq -200, \qquad b' \geq -200, \end{aligned}$$

and such that the matrix $M_{aba'b'}$ has nonzero kernel. We made sure that only one of the four equivalent quadruples

$$(a,b,a',b'), \quad (-a',-b',-a,-b), \quad (a,a-b,a',a'-b'), \quad (-a',b'-a',-a,b-a)$$

was on the list, and we ignored the quadruples where (a, b) and (a', b') define geometrically isomorphic elliptic curves.

This led to 3075 surfaces with a kernel vector of type 1 and 367 surfaces with a kernel vector of type 2, together with two surfaces with $\mathrm{Br}(S)_2$ of dimension two. The latter correspond to the quadruples $(25, 9, -169, -25)$ and $(25, 16, -169, -25)$.

#Relevant primes	0	1	2	3	4	5	6
#Surfaces	6	428	1577	1119	276	9	1

Table 3. Number of surfaces with a given number of relevant primes.

Among the 3075 surfaces, 26 actually have $\mathrm{Br}(S)_2 = 0$, due to a $\mathbb{Q}$-isogeny between the corresponding elliptic curves.

The complete list of these surfaces, the exact equations we worked with, and more details are available on both author's web pages in the file `ants_X_data.txt`.

4C. ***The BM-relevant primes — the p-adic point of view.*** We say that a Brauer class $\alpha \in \mathrm{Br}(S)$ *works* at a prime p if the local evaluation map $\mathrm{ev}_{\alpha,p}$ is nonconstant. For every surface in the sample described in Section 4B, we used Algorithm 2.12 and Theorem 2.19 to determine all of the *BM-relevant primes* p — that is, those for which there is a Brauer class working at p.

For the two surfaces with $\mathrm{Br}(S)_2$ of dimension two, the situation is as follows: In the case of the parameter vector $(25, 9, -169, -25)$, one Brauer class works at 2 and 13, another at 5 and 13, and the third at all three. For the surface corresponding to $(25, 16, -169, -25)$, one Brauer class works at 3 and 13, another at 5 and 13, and the last at all three.

Table 3 lists the number of surfaces in our sample set having a given number of relevant primes. The one example with six relevant primes is $(196, 75, -361, -169)$, for which the Brauer class works at 2, 5, 7, 11, 13, and 19.

For three surfaces, it happened that the corresponding elliptic curves were isogenous over a proper extension of $\mathbb{Q}$. In these cases, the Brauer-Manin obstruction is algebraic. For two of the surfaces, it worked at one prime, while for the third it worked at two.

4D. ***The BM-relevant primes — $\mathbb{Q}$-rational points.*** When the Brauer class α works at l primes $p_1, \ldots, p_l$, there are 2^l vectors with entries in $\{0, \frac{1}{2}\}$. By the Brauer-Manin obstruction, half of these vectors cannot be obtained as values of $(\mathrm{ev}_{\alpha,p_1}(x), \ldots, \mathrm{ev}_{\alpha,p_l}(x))$ for $\mathbb{Q}$-rational points $x \in S(\mathbb{Q})$. For every surface in our sample set, and for every vector not forbidden by the Brauer-Manin obstruction, we used Algorithm 3.4 to test whether there is a rational point giving rise to the vector.

It turned out that this was indeed the case. Thus, no further obstruction becomes visible via this coloring. However, in some of the cases rather high search bounds were necessary. Table 4 shows, for the extreme case of six relevant primes, the number of vectors hit for several search bounds. Somewhat surprisingly, the smallest solution for each color was smooth with respect to a bad prime bound of 800.

For the other surfaces in the sample, lower search bounds were sufficient, but the differences were enormous. We summarize our observations in Table 5.

Bound	50	100	200	400	800	1600	3200	6400	12800	25600	50000
#Vecs	5	10	14	20	24	26	28	30	31	31	32

Table 4. Numbers of evaluation vectors obtained from rational points of bounded height for the surface with parameters $(196, 75, -361, -169)$.

		Search bound B								
#Primes	#Surfaces	50	100	200	400	800	1600	3200	6400	12800
2	1577	190	56	22	—	—	—	—	—	—
3	1119	555	187	48	1	—	—	—	—	—
4	262	262	200	127	67	36	24	13	4	—
5	9	9	9	8	8	8	5	3	1	—

Table 5. Search bounds required to obtain all possible evaluation vectors from rational points. For each entry in the first column, we list in the second column the number of surfaces in our sample having that number of relevant primes. For each search bound B in columns 3 through 11, we list the number of these surfaces for which the rational points of height at most B do *not* account for all valuations vectors not forbidden by the Brauer-Manin obstruction.

Remark 4.1. There is the expectation that the behavior of the evaluation map $\mathrm{ev}_{\alpha,p}$ is strongly connected to the type of bad reduction at the prime p. For algebraic Brauer classes, such a connection is well known; for example, see [11]. In the transcendental case, there are only partial results; see for example [13, §4].

For our examples, the reductions S_p are rational surfaces having one or two double lines. Further, $\mathrm{ev}_{\alpha,p}$ is necessarily constant on the set of $\mathbb{Q}$-rational points reducing to a smooth point. The finer structure seems to be complicated; compare Lemma 2.9.

References

[1] Manuel Benito and Juan L. Varona, *Pythagorean triangles with legs less than n*, J. Comput. Appl. Math. **143** (2002), no. 1, 117–126. MR 2003b:11027

[2] Z. I. Borevich and I. R. Shafarevich, *Number theory*, Pure and Applied Mathematics, no. 20, Academic Press, New York, 1966. MR 33 #4001

[3] Martin Bright, *The Brauer-Manin obstruction on a general diagonal quartic surface*, Acta Arith. **147** (2011), no. 3, 291–302. MR 2012f:11118

[4] J. W. S. Cassels and M. J. T. Guy, *On the Hasse principle for cubic surfaces*, Mathematika **13** (1966), 111–120. MR 35 #2841

[5] Jean-Louis Colliot-Thélène, Dimitri Kanevsky, and Jean-Jacques Sansuc, *Arithmétique des surfaces cubiques diagonales*, in Wüstholz [28], 1987, pp. 1–108. MR 89g:11051

[6] Jean-Louis Colliot-Thélène and Alexei N. Skorobogatov, *Good reduction of the Brauer–Manin obstruction*, Trans. Amer. Math. Soc. **365** (2013), no. 2, 579–590. MR 2995366

[7] Sinnou David (ed.), *Number theory*: *Papers from the Séminaire de Théorie des Nombres de Paris*, 1993–94, London Mathematical Society Lecture Note Series, no. 235, Cambridge University Press, 1996. MR 99b:11003

[8] Andreas-Stephan Elsenhans and Jörg Jahnel, *The Diophantine equation* $x^4 + 2y^4 = z^4 + 4w^4$, Math. Comp. **75** (2006), no. 254, 935–940. MR 2007e:11143

[9] ———, *On the Brauer-Manin obstruction for cubic surfaces*, J. Comb. Number Theory **2** (2010), no. 2, 107–128. MR 2907786

[10] ———, *On the order three Brauer classes for cubic surfaces*, Cent. Eur. J. Math. **10** (2012), no. 3, 903–926. MR 2902222

[11] ———, *On the quasi-group of a cubic surface over a finite field*, J. Number Theory **132** (2012), no. 7, 1554–1571. MR 2903170

[12] David Harari, *Obstructions de Manin transcendantes*, in David [7], 1996, pp. 75–87. MR 99e: 14025

[13] Brendan Hassett and Anthony Várilly-Alvarado, *Failure of the Hasse principle on general K*3 *surfaces*, 2011. arXiv 1110.1738 [math.NT]

[14] Brendan Hassett, Anthony Várilly-Alvarado, and Patrick Varilly, *Transcendental obstructions to weak approximation on general K*3 *surfaces*, Adv. Math. **228** (2011), no. 3, 1377–1404. MR 2012i:14025

[15] Evis Ieronymou, *Diagonal quartic surfaces and transcendental elements of the Brauer groups*, J. Inst. Math. Jussieu **9** (2010), no. 4, 769–798. MR 2011g:14053

[16] Evis Ieronymou, Alexei N. Skorobogatov, and Yuri G. Zarhin, *On the Brauer group of diagonal quartic surfaces*, J. Lond. Math. Soc. (2) **83** (2011), no. 3, 659–672. MR 2012e:14046

[17] Ilya Karzhemanov, *One construction of a K*3 *surface with the dense set of rational points*, 2011. arXiv 1102.1873 [math.AG]

[18] Adam Logan, David McKinnon, and Ronald van Luijk, *Density of rational points on diagonal quartic surfaces*, Algebra Number Theory **4** (2010), no. 1, 1–20. MR 2011a:11126

[19] Yu. I. Manin, *Cubic forms*: *algebra*, *geometry*, *arithmetic*, North-Holland Mathematical Library, no. 4, North-Holland Publishing Co., Amsterdam, 1974. MR 57 #343

[20] L. J. Mordell, *On the conjecture for the rational points on a cubic surface*, J. London Math. Soc. **40** (1965), 149–158. MR 30 #58

[21] Bjorn Poonen and Yuri Tschinkel (eds.), *Arithmetic of higher-dimensional algebraic varieties*: *Proceedings of the Workshop on Rational and Integral Points of Higher-Dimensional Varieties held in Palo Alto*, *CA*, *December* 11–20, 2002, Progress in Mathematics, no. 226, Boston, Birkhäuser, 2004. MR 2004h:11001

[22] Thomas Preu, *Transcendental Brauer-Manin obstruction for a diagonal quartic surface*, Ph.D. thesis, Universität Zürich, 2011. http://www.math.uzh.ch/fileadmin/user/preu/publikation/preuThesis.pdf

[23] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Graduate Texts in Mathematics, no. 106, Springer, Dordrecht, 2009. MR 2010i:11005

[24] Alexei Skorobogatov and Peter Swinnerton-Dyer, 2*-descent on elliptic curves and rational points on certain Kummer surfaces*, Adv. Math. **198** (2005), no. 2, 448–483. MR 2006g:11129

[25] Alexei N. Skorobogatov and Yuri G. Zarhin, *The Brauer group of Kummer surfaces and torsion of elliptic curves*, J. Reine Angew. Math. **666** (2012), 115–140. MR 2920883

[26] H. P. F. Swinnerton-Dyer, *Two special cubic surfaces*, Mathematika **9** (1962), 54–56. MR 25 #3413

[27] Olivier Wittenberg, *Transcendental Brauer-Manin obstruction on a pencil of elliptic curves*, in Poonen and Tschinkel [21], 2004, pp. 259–267. MR 2005c:11082

[28] G. Wüstholz (ed.), *Diophantine approximation and transcendence theory*: *Papers from the seminar on number theory held in Bonn*, *May–June* 1985, Lecture Notes in Mathematics, no. 1290, Springer, Berlin, 1987. MR 88j:11036

ANDREAS-STEPHAN ELSENHANS: stephan@maths.usyd.edu.au
School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Sydney, Australia

JÖRG JAHNEL: jahnel@mathematik.uni-siegen.de
Department Mathematik, Universität Siegen, Walter-Flex-Straße 3, D-57068 Siegen, Germany

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.395

msp

Explicit 5-descent on elliptic curves

Tom Fisher

We compute equations for genus-one curves representing nontrivial elements of order 5 in the Tate-Shafarevich group of an elliptic curve. We explain how to write the equations in terms of Pfaffians and give examples for elliptic curves over the rationals both with and without a rational 5-isogeny.

1. Introduction

An explicit descent calculation on an elliptic curve E over a number field K computes the Selmer group (attached to some isogeny) and represents its elements by giving equations for the corresponding covering curves. These curves may be used to help search for generators of the Mordell-Weil group $E(K)$ or to exhibit nontrivial elements of the Tate-Shafarevich group $Ш(E/K)$.

Let C be a smooth curve of genus one representing an element of order n in $Ш(E/K)$. Cassels [8] showed that C admits a K-rational divisor D of degree n. So for $n \geq 3$ we may embed $C \subset \mathbb{P}^{n-1}$ by the complete linear system $|D|$. The result is called a *genus-one normal curve of degree n*. For $n \geq 4$ it is well known (see for example [19; 29]) that the homogeneous ideal of such a curve is generated by a vector space of quadrics of dimension $n(n-3)/2$.

The equations for a genus-one normal curve of degree 5 may conveniently be written as the 4×4 Pfaffians of a 5×5 alternating matrix of linear forms. Over the complex numbers this is a classical fact. In general it is a consequence of the Buchsbaum-Eisenbud structure theorem [7; 6] for Gorenstein ideals of codimension 3. In Section 4 we explain how to compute these matrices of linear forms.

The author has been compiling [22] a list of explicit elements of $Ш(E/\mathbb{Q})[5]$ for elliptic curves $E/\mathbb{Q}$ of small conductor (taken from the Cremona database [10; 11]). The equations are computed using either descent by 5-isogeny, full 5-descent, or visibility. We give details of the first two of these methods in Sections 5 and 6,

MSC2010: primary 11G05; secondary 14H52, 14H25.
Keywords: elliptic curves, descent, Selmer groups, Pfaffians.

expanding on the treatments in [17] and [12; 13; 14]. Our use of visibility is described in [20].

2. Background on descent

Let $\phi : E \to E'$ be an isogeny of elliptic curves over K. A *ϕ-covering* of E' is a pair (C, π) where C is a smooth curve of genus one and $\pi : C \to E'$ is a morphism (both defined over K) such that the diagram

$$\begin{array}{ccc} C & & \\ {\scriptstyle\psi}\downarrow & \searrow^{\pi} & \\ E & \xrightarrow[\phi]{} & E' \end{array}$$

commutes for some isomorphism $\psi : C \to E$ defined over $\overline{K}$.

We write $H^i(K, -)$ as a shorthand for $H^i(\mathrm{Gal}(\overline{K}/K), -)$. Taking Galois cohomology of the short exact sequence of $\mathrm{Gal}(\overline{K}/K)$-modules

$$0 \longrightarrow E[\phi] \longrightarrow E \xrightarrow{\phi} E' \longrightarrow 0$$

gives a long exact sequence of abelian groups

$$\cdots \longrightarrow E(K) \xrightarrow{\phi} E'(K) \xrightarrow{\delta} H^1(K, E[\phi]) \longrightarrow H^1(K, E) \longrightarrow \cdots . \qquad (1)$$

The group $H^1(K, E[\phi])$ parametrises the ϕ-coverings of E', up to isomorphism over K. The subgroup of everywhere locally soluble coverings is the *ϕ-Selmer group* $S^{(\phi)}(E/K)$. Likewise the group $H^1(K, E)$ parametrises the torsors (or principal homogeneous spaces) under E, up to isomorphism over K. The subgroup of everywhere locally soluble torsors is the *Tate-Shafarevich group* $Ш(E/K)$. There is then an exact sequence

$$0 \longrightarrow E'(K)/\phi E(K) \xrightarrow{\delta} S^{(\phi)}(E/K) \longrightarrow Ш(E/K)[\phi_*] \longrightarrow 0$$

where $\phi_* : Ш(E/K) \to Ш(E'/K)$ is the map induced by ϕ.

There are two natural ways to construct a rational divisor class on C. Let m be the smallest positive integer such that $E[\phi] \subset E[m]$. Then $D = \psi^*(m \cdot 0_E)$ and $D' = \pi^*(0_{E'})$ are divisors on C of degrees m and $n = \deg \phi$, respectively. A calculation shows that D is linearly equivalent to all its Galois conjugates, whereas D' is already defined over K. For each $\sigma \in \mathrm{Gal}(\overline{K}/K)$ we pick $h_\sigma \in \overline{K}(C)^\times$ with $\mathrm{div}(h_\sigma) = \sigma D - D$. There is then an obstruction map (see [26; 30; 12])

$$\mathrm{Ob} : H^1(K, E[\phi]) \longrightarrow \mathrm{Br}(K) = H^2(K, \overline{K}^\times)$$

that sends the ϕ-covering (C, π) to the class of the 2-cocycle $(\sigma, \tau) \mapsto \sigma(h_\tau)h_\sigma / h_{\sigma\tau}$. Since $H^1(K, \overline{K}(C)^\times) = 0$ it follows that D is linearly equivalent to a K-rational divisor if and only if (C, π) has trivial obstruction.

If $\#(E[\phi] \cap E[2]) = 1$ or 4 then the elements of $E[\phi]$ sum to 0_E, in which case $\phi^*(0_{E'}) \sim n \cdot 0_E$ and $D' \sim (n/m)D$.

In this paper we are interested in the following two cases, where we may write C as a genus-one normal curve of degree 5 with hyperplane section D:

(i) ϕ is an isogeny of degree 5 and $(C, \pi) \in H^1(K, E[\phi])$.

(ii) ϕ is multiplication-by-5 on E and $(C, \pi) \in S^{(5)}(E/K)$.

The obstruction is trivial in both cases. In the first case this is because $D \sim D'$, whereas in the second case the proof (which we follow in our calculations) uses the local-to-global principle for the Brauer group.

3. Pfaffians

We recall some basic facts about Pfaffians. Let $A = (a_{ij})$ be an $n \times n$ alternating matrix. If $n = 2m$ is even then the *Pfaffian* of A is

$$\mathrm{pf}(A) = \frac{1}{2^m m!} \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) \prod_{i=1}^{m} a_{\sigma(2i-1)\sigma(2i)}. \tag{2}$$

Standard calculations (see [3, §5]) show that $\mathrm{pf}(PAP^T) = \det(P)\,\mathrm{pf}(A)$ and that $\det(A) = \mathrm{pf}(A)^2$. Since $\det(A)$ in an integer coefficient polynomial in the entries of A, the same must be true of $\mathrm{pf}(A)$. This is used to define the Pfaffian over an arbitrary ring.

Pfaffians, just like determinants, may be expanded along a row. We write $A^{\{i,j\}}$ for the matrix obtained from A by deleting the i-th and j-th rows and columns. It may be shown using (2) that

$$\mathrm{pf}(A) = \sum_{j=2}^{n} (-1)^j a_{1j}\,\mathrm{pf}(A^{\{1,j\}}).$$

For example, in the 4×4 case we have

$$\mathrm{pf}\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} \\ & 0 & a_{23} & a_{24} \\ & - & 0 & a_{34} \\ & & & 0 \end{pmatrix} = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}.$$

Definition 3.1. Let A be an $n \times n$ alternating matrix with n odd. The *row vector of submaximal Pfaffians of* A is $\mathrm{Pf}(A) = (p_1, \ldots, p_n)$, where $p_i = (-1)^i\,\mathrm{pf}(A^{\{i\}})$ and $A^{\{i\}}$ is the matrix obtained by deleting the i-th row and column of A.

Lemma 3.2. *If A is an $n \times n$ alternating matrix with n odd then*

(i) $\mathrm{Pf}(A)A = 0$,

(ii) $\mathrm{Pf}(PAP^T) = \mathrm{Pf}(A)\,\mathrm{adj}(P)$,

(iii) $\mathrm{adj}(A) = \mathrm{Pf}(A)^T\,\mathrm{Pf}(A)$.

Proof. Since we only need the case $n = 5$ (which may be checked by a generic computation) we omit the proof. □

4. Computing genus-one models

A *genus-one model* (of degree 5) is a 5×5 alternating matrix of linear forms in variables $x_1, \ldots, x_5$. We write $X_5(K)$ for the space of all genus-one models with coefficients in a field K, and $C_\Phi \subset \mathbb{P}^4$ for the subscheme defined by the 4×4 Pfaffians of $\Phi \in X_5(K)$.

Theorem 4.1. *Let $C \subset \mathbb{P}^4$ be a genus-one normal curve of degree 5 defined over a field K.*

(i) *There exists $\Phi \in X_5(K)$ such that $C = C_\Phi$.*

(ii) *If $\Phi_1, \Phi_2 \in X_5(K)$ with $C = C_{\Phi_1} = C_{\Phi_2}$ then there exist $A \in \mathrm{GL}_5(K)$ and $\mu \in K^\times$ such that $\Phi_2 = \mu A \Phi_1 A^T$.*

Theorem 4.1 is a consequence of the Buchsbaum-Eisenbud structure theorem [7; 6] for Gorenstein ideals of codimension 3. In this section we give a simplified form of the proof and use it to give explicit algorithms for computing Φ and A. These algorithms are needed in our work [19; 20] on the invariant theory of genus-one models.

Example 4.2. Let E be the elliptic curve $y^2 = x^3 + ax + b$. For any $n \geq 3$ we may embed E into $\mathbb{P}^{n-1}$ via the complete linear system $|n \cdot 0_E|$ to give a genus-one normal curve of degree n. If $n = 5$ then the embedding is given by

$$(x_1 : \cdots : x_5) = (1 : x : y : x^2 : xy)$$

and the image is defined by the 4×4 Pfaffians of

$$\begin{pmatrix} 0 & bx_1 & x_5 & x_4 + ax_1 & -x_3 \\ & 0 & -x_4 & -x_3 & x_2 \\ & & 0 & -x_2 & 0 \\ & - & & 0 & -x_1 \\ & & & & 0 \end{pmatrix}.$$

(Since the homogeneous ideal is generated by a 5-dimensional space of quadrics, it suffices to check that the 4×4 Pfaffians are linearly independent and that they vanish on E.)

Let $R = K[x_1, \ldots, x_n] = \bigoplus_{d \geq 0} R_d$ be the polynomial ring with its usual grading by degree. Let $R_+ = \bigoplus_{d \geq 1} R_d$ be the irrelevant ideal.

Definition 4.3. Let M be a finitely generated graded R-module. A *graded free resolution* of M is a complex of graded free R-modules

$$F_\bullet : \quad 0 \longrightarrow F_s \xrightarrow{\varphi_s} F_{s-1} \longrightarrow \cdots \longrightarrow F_2 \xrightarrow{\varphi_2} F_1 \xrightarrow{\varphi_1} F_0 \longrightarrow 0$$

that is exact at all terms except F_0, where we have $F_0/\operatorname{im}(\varphi_1) \cong M$. The resolution $F_\bullet$ is *minimal* if $\phi_i(F_i) \subset R_+ F_{i-1}$ for all i.

We shall need the following two facts.

Lemma 4.4. *Let $F_\bullet$ be a minimal graded free resolution of M. Then any graded free resolution of M is a direct sum of $F_\bullet$ and a trivial complex. In particular, $F_\bullet$ is unique up to isomorphism.*

Proof. See [16, §20.1] or [33, §7]. □

Lemma 4.5 (Buchsbaum-Eisenbud acyclicity criterion). *The complex $F_\bullet$ is acyclic (that is, exact at all terms except F_0) if and only if for all $1 \leq i \leq s$,*

$$\operatorname{rank} F_i = \operatorname{rank} \varphi_i + \operatorname{rank} \varphi_{i+1},$$

and the ideal generated by the $r_i \times r_i$ minors of φ_i (where $r_i = \operatorname{rank} \varphi_i$) has codimension at least i.

Proof. See [5, Theorem 1.4.13] or [16, Theorem 20.9]. We use here that R is Cohen-Macaulay, so that the codimension (also called height) of an ideal is the same as the grade (also called depth). □

We follow the convention that maps of graded R-modules preserve the degree. Let $R(d)$ be R as a graded module over itself with degrees shifted by d, that is, $R(d)_e = R_{d+e}$. We use the same notation for maps of R-modules and the matrices that represent them (with respect to the standard bases).

Theorem 4.6. *Let $C \subset \mathbb{P}^4$ be a genus-one normal curve of degree 5 with homogeneous ideal $I = I(C) \subset R = K[x_1, \ldots, x_5]$.*

(i) *The minimal graded free resolution of R/I takes the form*

$$0 \longrightarrow R(-5) \xrightarrow{Q^T} R(-3)^5 \xrightarrow{\Phi} R(-2)^5 \xrightarrow{P} R \longrightarrow 0. \tag{3}$$

(*This means that $P = (p_1, \ldots, p_5)$ and $Q = (q_1, \ldots, q_5)$ are vectors of quadrics and Φ is a 5×5 matrix of linear forms.*)

(ii) *The K-vector space*

$$\{B \in \operatorname{Mat}_5(K) \mid \Phi B \text{ is alternating}\}$$

is 1-dimensional and contains a nonsingular matrix.

(iii) *If* Φ *is alternating then* P *and* Q *are scalar multiples of* $\mathrm{Pf}(\Phi)$.

Proof. The conclusions of the theorem are unchanged if we extend our field K, so we may assume K is algebraically closed. Then C is an elliptic curve, and up to translation any two divisors on C of the same degree are linearly equivalent. So we may change coordinates on $\mathbb{P}^4$ so that $C = C_\Phi$ where Φ is as given in Example 4.2. (If K has characteristic 2 or 3, we use the more general formula in [19, §6].) By Lemma 3.2(i) there is a complex

$$0 \longrightarrow R(-5) \xrightarrow{P^T} R(-3)^5 \xrightarrow{\Phi} R(-2)^5 \xrightarrow{P} R \longrightarrow 0 \tag{4}$$

with $P = \mathrm{Pf}(\Phi)$. Since P is not identically zero we have $\mathrm{rank}(\Phi) = 4$ and $\mathrm{rank}(P) = 1$. By Lemma 3.2(iii) the ideals generated by the 4×4 Pfaffians of Φ and the 4×4 minors of Φ have the same radical. Since $C \subset \mathbb{P}^4$ has codimension 3, the conditions of Lemma 4.5 are satisfied and so (4) is the minimal graded free resolution of R/I. This proves (i) and shows by Lemma 4.4 that for any resolution (3) there exist $A_1, A_2 \in \mathrm{GL}_5(K)$ such that $A_1 \Phi A_2$ is alternating. Replacing Φ by $\Phi A_2 A_1^{-T}$ we may assume for the proof of (ii) that Φ is alternating.

Suppose that both Φ and ΦB are alternating for some $B \in \mathrm{Mat}_5(K)$. Then $P\Phi = P\Phi B = 0$ and $\Phi P^T = \Phi B P^T = 0$. Since the sequence (3) is exact it follows that P^T and BP^T are scalar multiples of Q^T. Therefore B is a scalar matrix. This proves (ii). To prove (iii) we apply the same argument starting with the identity $\mathrm{Pf}(\Phi)\Phi = 0$. □

Theorem 4.6 not only proves Theorem 4.1(i) but gives the following algorithm for computing a genus-one model Φ with $C = C_\Phi$. We start with a basis $p_1, \ldots, p_5$ for the space of quadrics vanishing on C. We then solve by linear algebra for a matrix Ψ whose columns are a basis for the space of all 5-tuples of linear forms $(\ell_1, \ldots, \ell_5) \in R^5$ satisfying $\sum_{i=1}^5 \ell_i p_i = 0$. Finally we take $\Phi = \Psi B$ where $B \in \mathrm{Mat}_5(K)$ is any nonzero matrix satisfying $\Psi B = -B^T \Psi^T$.

To prove Theorem 4.1(ii) we put $P_1 = \mathrm{Pf}(\Phi_1)$ and $P_2 = \mathrm{Pf}(\Phi_2)$, and note that by Lemma 4.4 there is an isomorphism of complexes

$$\begin{array}{ccccccccccc}
0 & \longrightarrow & R(-5) & \xrightarrow{P_1^T} & R(-3)^5 & \xrightarrow{\Phi_1} & R(-2)^5 & \xrightarrow{P_1} & R & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \mu} & & \downarrow{\scriptstyle A^{-T}} & & \downarrow{\scriptstyle B^T} & & \| & & \\
0 & \longrightarrow & R(-5) & \xrightarrow{P_2^T} & R(-3)^5 & \xrightarrow{\Phi_2} & R(-2)^5 & \xrightarrow{P_2} & R & \longrightarrow & 0
\end{array}$$

for some $A, B \in \mathrm{GL}_5(K)$ and $\mu \in K^\times$. Commutativity of this diagram gives $P_1^T = \mu A^T P_2^T = B P_2^T$ and $\Phi_2 = B^T \Phi_1 A^T$. Since the entries of P_2 are linearly independent it follows that $B = \mu A^T$, and so $\Phi_2 = \mu A \Phi_1 A^T$, as required. The proof shows that $A \in \mathrm{GL}_5(K)$ is uniquely determined up to scalars by the condition $\mathrm{Pf}(\Phi_1) \propto \mathrm{Pf}(\Phi_2)A$. This observation (which also follows by Lemma 3.2(ii)) gives

a convenient way to compute A. If K is algebraically closed then we may scale A so that $\mu = 1$. With this convention A is unique up to sign.

5. Descent by isogeny

We return to working over a number field K. Let $\phi : E \to E'$ be a cyclic isogeny of degree n and let $\widehat{\phi} : E' \to E$ be its dual isogeny. If (C, π) is a ϕ-covering of E' then $(C, \widehat{\phi} \circ \pi)$ is an n-covering of E. In general not all n-coverings of E arise in this way. Instead, an upper bound for the rank is obtained by computing both $S^{(\phi)}(E/K)$ and $S^{(\widehat{\phi})}(E'/K)$.

Since the Weil pairing $E[\phi] \times E'[\widehat{\phi}] \to \mu_n$ is nondegenerate, the action of Galois on $E[\phi]$, $E'[\widehat{\phi}]$, and μ_n is described by three characters

$$\chi^{-1}\omega, \chi, \omega : \mathrm{Gal}(\overline{K}/K) \to (\mathbb{Z}/n\mathbb{Z})^{\times}.$$

Let $L = K(E'[\widehat{\phi}])$ be the fixed field of the kernel of χ, and let $G = \mathrm{Gal}(L/K)$. If n is prime then $[L : K]$ divides $n-1$ and so is coprime to n. By the inflation-restriction exact sequence we have

$$H^1(K, E[\phi]) \cong H^1(L, E[\phi])^G.$$

Since $H^1(L, E[\phi]) \cong H^1(L, \mu_n) \cong L^{\times}/(L^{\times})^n$ it follows (by keeping track of the G-actions) that $H^1(K, E[\phi]) \cong (L^{\times}/(L^{\times})^n)^{\chi}$, where, if A is a G-module, we write

$$A^{\chi} = \{a \in A \mid \sigma(a) = a^{\chi(\sigma)} \text{ for all } \sigma \in G\}.$$

There is an analogue of the exact sequence (1) obtained by replacing K by its completion K_v. Let δ_v be the connecting map in this exact sequence. The Selmer group attached to ϕ is

$$S^{(\phi)}(E/K) = \{\theta \in H^1(K, E[\phi]) \mid \mathrm{res}_v(\theta) \in \mathrm{im}\, \delta_v \text{ for all places } v\}$$

where $\mathrm{res}_v : H^1(K, E[\phi]) \to H^1(K_v, E[\phi])$ is the restriction map. Assuming we can compute the groups

$$L(\mathcal{S}, n) = \{\theta \in L^{\times}/(L^{\times})^n \mid v_{\mathfrak{p}}(\theta) \equiv 0 \bmod n \text{ for all } \mathfrak{p} \notin \mathcal{S}\}$$

for $\mathcal{S}$ a finite set of primes, the problem of computing the Selmer group reduces to that of computing the images of the local connecting maps δ_v. Since we give equations for the covering curves, the images of the δ_v may be computed by working out conditions for these curves to be locally soluble. See for example [17; 18; 9]. Alternatively, as described for example in [23; 28], the images of the δ_v may be computed as the cokernels of the maps $\phi : E(K_v) \to E'(K_v)$.

We take $n = 5$ and split into the cases where χ has order 1, 2 or 4. If χ is trivial then $E[\phi] \cong \mu_5$ and $E'[\widehat{\phi}] \cong \mathbb{Z}/5\mathbb{Z}$ as Galois modules. We recall from [17] that $E \cong C_\lambda$ and $E' \cong D_\lambda$ for some $\lambda \in K$, where C_λ and D_λ are the curves given by

$$\begin{aligned} C_\lambda: &\quad y^2 + (1-\lambda)xy - \lambda y = x^3 - \lambda x^2 + a_4 x + a_6, \\ D_\lambda: &\quad y^2 + (1-\lambda)xy - \lambda y = x^3 - \lambda x^2, \end{aligned} \tag{5}$$

where $a_4 = -5\lambda(\lambda^2 + 2\lambda - 1)$ and $a_6 = -\lambda(\lambda^4 + 10\lambda^3 - 5\lambda^2 + 15\lambda - 1)$.

Theorem 5.1. *If $\lambda_0, \ldots, \lambda_4$ are elements of K such that*

$$\lambda = \prod_{i=0}^{4} \lambda_i \quad \text{and} \quad \theta \equiv \prod_{i=0}^{4} \lambda_i^i \mod (K^\times)^5, \tag{6}$$

then the ϕ-covering of D_λ corresponding to $\theta \in K^\times/(K^\times)^5$ is defined by the 4×4 Pfaffians of

$$\begin{pmatrix} 0 & \lambda_1 x_1 & x_2 & -x_3 & -\lambda_4 x_4 \\ & 0 & \lambda_3 x_3 & x_4 & -x_0 \\ & & 0 & \lambda_0 x_0 & x_1 \\ & - & & 0 & \lambda_2 x_2 \\ & & & & 0 \end{pmatrix}.$$

Proof. See [17, Proposition 2.12]. The analogue of this result for cyclic isogenies of degrees 3 and 4 is given in [18, §1.2]. □

Example 5.2. Taking $K = \mathbb{Q}$ and $(\lambda_0, \ldots, \lambda_4) = (1, 1, 2, 3, 5)$ gives an element of order 5 in $Ш(C_{30}/\mathbb{Q})$.

If χ is a quadratic character then E and E' are the quadratic twists by χ of C_λ and D_λ for some $\lambda \in K$. We write $L = K(\sqrt{d})$.

Theorem 5.3. *If r and s are elements of K, not both zero, then the ϕ-covering of E' corresponding to $\theta = (r + s\sqrt{d})/(r - s\sqrt{d}) \in (L^\times/(L^\times)^5)^\chi$ is defined by the 4×4 Pfaffians of*

$$\begin{pmatrix} 0 & \lambda x_0 & d(x_2 - x_4) & -x_1 + x_3 & -x_3 \\ & 0 & -x_1 - x_3 & x_2 + x_4 & x_4 \\ & & 0 & (r^2 - s^2 d)x_0 & rx_1 + sdx_2 \\ & - & & 0 & sx_1 + rx_2 \\ & & & & 0 \end{pmatrix}.$$

Proof. Let $\alpha = r + s\sqrt{d}$ and $\alpha' = r - s\sqrt{d}$. We apply Theorem 5.1 over L with

$$(\lambda_0, \ldots, \lambda_4) = (\lambda/(\alpha\alpha'), \alpha, 1, 1, \alpha').$$

We then substitute $x_0 \leftarrow -(r^2 - s^2 d)x_0$ and

$$(x_1, \ldots, x_4) \leftarrow (x_1 + \sqrt{d}\,x_2, x_3 + \sqrt{d}\,x_4, x_3 - \sqrt{d}\,x_4, x_1 - \sqrt{d}\,x_2)$$

to give a curve defined over K. Since $[L : K]$ and $\deg \phi$ are coprime to one another, the restriction map $H^1(K, E[\phi]) \to H^1(L, E[\phi])$ is injective. Since the curve we have found and the curve we are looking for are isomorphic over L, they must therefore be isomorphic over K. □

Example 5.4. Taking $K = \mathbb{Q}$ and $\lambda = 11$, $d = 5$, $r = s = 1$ gives an element of order 5 in $Ш(E/\mathbb{Q})$ where E is the elliptic curve 275b3 in Cremona's tables [10; 11].

Remark 5.5. The curve in Theorem 5.1 is defined by the 5 quadrics

$$\lambda_i x_i^2 + x_{i-1}x_{i+1} - \lambda_{i-2}\lambda_{i+2}x_{i-2}x_{i+2} = 0$$

where the subscripts are read modulo 5. If $\lambda_i' = -\lambda_{2i}/(\lambda_{2i-2}\lambda_{2i+2})$ then the curves defined by $\lambda_0, \ldots, \lambda_4$ and $\lambda_0', \ldots, \lambda_4'$ are isomorphic via

$$(x_0 : \cdots : x_4) \mapsto (x_0 : x_2 : x_4 : x_1 : x_3).$$

Taking Jacobians it follows that $C_\lambda \cong C_{-1/\lambda}$. Alternatively this last statement may be checked using the Weierstrass equations (5).

Now suppose χ has order 4. Let σ be the generator of $\mathrm{Gal}(L/K)$ with $\chi(\sigma) = 2$. Then E and E' are isomorphic over L to C_λ and D_λ for some $\lambda \in L$ satisfying $\sigma(\lambda) = -1/\lambda$.

Theorem 5.6. *If $\alpha \in L^\times$ then the ϕ-covering of E' corresponding to*

$$\theta = \alpha^4\sigma(\alpha)^2\sigma^2(\alpha)\sigma^3(\alpha)^3 \in (L^\times/(L^\times)^5)^\chi \tag{7}$$

is isomorphic over L to the curve in Theorem 5.1 with

$$(\lambda_0, \ldots, \lambda_4) = \left(\lambda\sigma(\alpha)\sigma^3(\alpha), \frac{\alpha}{\lambda\sigma(\alpha)\sigma^3(\alpha)}, \frac{\lambda\sigma(\alpha)}{\alpha}, \frac{\lambda\sigma^3(\alpha)}{\sigma^2(\alpha)}, \frac{\sigma^2(\alpha)}{\lambda\sigma(\alpha)\sigma^3(\alpha)}\right).$$

Moreover a model for this curve over K is obtained by substituting

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \leftarrow \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \sigma(\beta_1) & \sigma(\beta_2) & \sigma(\beta_3) & \sigma(\beta_4) \\ \sigma^3(\beta_1) & \sigma^3(\beta_2) & \sigma^3(\beta_3) & \sigma^3(\beta_4) \\ \sigma^2(\beta_1) & \sigma^2(\beta_2) & \sigma^2(\beta_3) & \sigma^2(\beta_4) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

where $\beta_1, \ldots, \beta_4$ is a basis for L over K.

Proof. The first part is clear since we have chosen $\lambda_0, \ldots, \lambda_4$ to satisfy (6). We have also arranged that $\sigma(\lambda_i) = -\lambda_{2i}/(\lambda_{2i-2}\lambda_{2i+2})$. The second part then follows by Remark 5.5. □

Remark 5.7. Since $\rho = 4 + 2\sigma + \sigma^2 + 3\sigma^3 \in \mathbb{F}_5[G]$ is an idempotent satisfying $\sigma\rho = 2\rho$, every element of $(L^\times/(L^\times)^5)^\chi$ is of the form (7).

Example 5.8. Let E and E' be the 5-isogenous elliptic curves

$$\begin{aligned} E = 23808c3: &\qquad y^2 = x^3 - x^2 - 785949x - 271615419, \\ E' = 23808c2: &\qquad y^2 = x^3 - x^2 + 7651x + 676677. \end{aligned}$$

Then $L = \mathbb{Q}(\varepsilon)$ where $\varepsilon = \sqrt{2+\sqrt{2}}$. Moreover $\lambda = (49+41\sqrt{2})/31$ and $\sigma : \varepsilon \mapsto \varepsilon^3 - 3\varepsilon$. We take $\alpha = 1 + \varepsilon$ and $\beta_j = \varepsilon^{j-1}$ for $j = 1, \dots, 4$. After following the construction in Theorem 5.6, the algorithms for minimisation and reduction in [21] suggest the change of coordinates

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \leftarrow \begin{pmatrix} 0 & 0 & 0 & 0 & 62 \\ 0 & 6 & -6 & 14 & 0 \\ 13 & -13 & -7 & -7 & 0 \\ 0 & 1 & -1 & -8 & 0 \\ -3 & 3 & 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

The result is $C \subset \mathbb{P}^4$ defined by the 4×4 Pfaffians of

$$\begin{pmatrix} 0 & x_0 - x_1 + x_3 + 4x_4 & x_1 - x_2 - x_4 & -x_2 - 2x_3 + 4x_4 & x_1 \\ & 0 & -x_2 - 4x_4 & x_1 - x_2 + x_4 & x_3 \\ & & 0 & x_0 - x_1 - x_3 - 4x_4 & x_2 \\ & - & & 0 & x_0 \\ & & & & 0 \end{pmatrix}.$$

Computing the invariants, as described in [19], and using Bruin's programs [4] to check local solubility, we find that C represents an element of $Ш(E/\mathbb{Q})[5]$. It is nontrivial since $E'(\mathbb{Q}) = 0$ and $\theta \notin (L^\times)^5$.

6. An example of full 5-descent

In this section we compute equations for an order-5 element in the Tate-Shafarevich group of the elliptic curve $E/\mathbb{Q}$:

$$6727a1: \quad y^2 + xy = x^3 - x^2 - 202951x - 34841040.$$

Since E has no rational 5-isogenies, our method is to use full 5-descent; that is, descent with respect to the multiplication-by-5 map on E. Further details of the calculation are given in a Magma [2] file available at this article's webpage.

Let $T = (x_T, y_T)$ be a nontrivial 5-torsion point on E. Then $L = \mathbb{Q}(T)$ is a number field of degree 24. Let σ_2 be the automorphism of L with $\sigma_2(T) = 2T$.

We shall write elements of L in terms of u and v where

$$v = -31(2y_T + x_T)/(x_T^2 + 480x_T + 87391)$$

and $u = v/\sigma_2(v)$. Explicitly, u has minimal polynomial

$$X^{12} + 4X^{11} - 6X^{10} - 20X^9 + 15X^8 - 303X^7 + 323X^6 + 303X^5 + 15X^4 + 20X^3 - 6X^2 - 4X + 1$$

and v is a square root of

$$(2u^{11}+9u^{10}-8u^9-46u^8+10u^7-591u^6+343u^5+928u^4+331u^3+60u^2+8u-9)/3.$$

We recall from [15; 34] that there is an injective group homomorphism

$$H^1(\mathbb{Q}, E[5]) \to L^\times/(L^\times)^5$$

whose image is contained in the σ_2-eigenspace

$$\{x \in L^\times/(L^\times)^5 \mid \sigma_2(x) \equiv x^2 \bmod (L^\times)^5\}. \tag{8}$$

The primes of bad reduction for E are 7 and 31, with Tamagawa numbers $c_7 = 1$ and $c_{31} = 2$. Since the Tamagawa numbers are coprime to 5, we have $S^{(5)}(E/\mathbb{Q}) \subset L(\mathcal{S}, 5)$ where $\mathcal{S} = \{\mathfrak{p}_1, \mathfrak{p}_2\}$ is the set of primes of L above 5.

The number field L is too large for an unconditional computation of its class group and units. However according to PARI/GP [32] (which by default makes heuristic assumptions) the class number is 2. We also used PARI/GP to compute a set of fundamental units, and generators for the prime ideals $\mathfrak{p}_1$ and $\mathfrak{p}_2$. This gives a basis for $L(\mathcal{S}, 5) \cong (\mathbb{Z}/5\mathbb{Z})^{15}$. The intersection of $L(\mathcal{S}, 5)$ with the σ_2-eigenspace (8) is 3-dimensional. One of the nontrivial elements is $a = a'/294$, with

$$\begin{aligned} a' = {} & (4600u^{11}+8325u^{10}-72155u^9-50035u^8+289975u^7-1450795u^6 \\ & +4510595u^5-592350u^4-3962957u^3-1755928u^2-811953u-191035)v \\ & +(158985u^{11}+661975u^{10}-836070u^9-3280275u^8+1784950u^7 \\ & -48064875u^6+43645605u^5+52498690u^4+14516335u^3+7628705u^2 \\ & +310520u-311257). \end{aligned}$$

We have $(a) = \mathfrak{c}^5$ for some integral ideal $\mathfrak{c}$, and $a\sigma_2(a)^2 = b^5$ where $b = b'/294$ and

$$\begin{aligned} b' = {} & (452u^{11} + 1935u^{10} - 2186u^9 - 9743u^8 + 4070u^7 - 135379u^6 \\ & + 108106u^5 + 172665u^4 + 54912u^3 + 14840u^2 - 4879u - 12762)v \\ & + (-1983u^{11} - 9082u^{10} + 7240u^9 + 46137u^8 - 7149u^7 + 585937u^6 \\ & - 289205u^5 - 957562u^4 - 338134u^3 - 139997u^2 - 62943u + 7646). \end{aligned}$$

We recall some of the theory from [12; 13; 14]. Let E be an elliptic curve over a field K of characteristic 0. Let R be the K-algebra of all Galois-equivariant maps $E[n] \to \overline{K}$ and let $w : E[n] \to \overline{R}^\times = \mathrm{Map}(E[n], \overline{K}^\times)$ be the map induced by the Weil pairing e_n. If $\sigma \mapsto \xi_\sigma$ is a cocycle representing $\xi \in H^1(K, E[n])$ then by Hilbert's theorem 90 there exists $\gamma \in \overline{R}^\times$ with $\sigma(\gamma)/\gamma = w(\xi_\sigma)$ for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$. We put $\alpha = \gamma^n$ and $\rho = \partial\gamma$ where

$$\partial : \overline{R}^\times \to (\overline{R} \otimes \overline{R})^\times = \mathrm{Map}(E[n] \times E[n], \overline{K}^\times)$$

is given by $(\partial z)(T_1, T_2) = z(T_1)z(T_2)/z(T_1 + T_2)$. Then according to [12, §3] there are group homomorphisms

$$w_1 : H^1(K, E[n]) \to R^\times/(R^\times)^n, \qquad \xi \mapsto \alpha,$$
$$w_2 : H^1(K, E[n]) \to (R \otimes R)^\times/\partial R^\times, \quad \xi \mapsto \rho.$$

The map w_1 is injective for n prime, whereas w_2 is always injective.

Let $\mathrm{Ob} : H^1(K, E[n]) \to \mathrm{Br}(K)$ be the obstruction map as defined in Section 2.

Theorem 6.1. *Assume n is odd. Let $\xi \in H^1(K, E[n])$ and $\rho \in (R \otimes R)^\times$ with $w_2(\xi) = \rho\partial R^\times$. Let $A_\rho = (R, +, *_\rho)$ where the new multiplication $*_\rho$ is defined by*

$$z_1 *_\rho z_2 : T \mapsto \sum_{T_1+T_2=T} e_n(T_1, T_2)^{(n+1)/2}\rho(T_1, T_2)z_1(T_1)z_2(T_2).$$

Then A_ρ is a central simple algebra over K of dimension n^2 representing the class of $\mathrm{Ob}(\xi)$ *in* $\mathrm{Br}(K)$.

Proof. See [12, Lemma 3.11 and §4]. □

Returning to our numerical example, we write α and β for the elements $(1, a)$ and $(1, b)$ in the étale algebra $R = \mathbb{Q} \times L$. To compute ρ exactly (using $\partial\alpha = \rho^5$) we must extract a 5th root in a number field of degree $\frac{1}{2}\#\mathrm{GL}_2(\mathbb{Z}/5\mathbb{Z}) = 240$. This would be the direct analogue of what we do for 3-descent (see [14, §8]), but is clearly not very promising. So instead we write $\rho = \partial\gamma$ and (fixing an embedding $\overline{\mathbb{Q}} \subset \mathbb{C}$) represent $\gamma \in \overline{R} = \mathrm{Map}(E[5], \overline{\mathbb{Q}})$ numerically. Since $\gamma^5 = \alpha$ there are at first sight 5^{25} possibilities for γ. We cut down to just 5^3 choices by requiring that

(i) $\gamma(T)\gamma(2T)^2 = \beta(T)$ for all $T \in E[5]$, and

(ii) $\gamma : E(\mathbb{C})[5] \to \mathbb{C}$ is $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$-equivariant.

To explain these conditions we recall that $\sigma(\gamma)/\gamma = w(\xi_\sigma)$ for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. From this it is easy to see that $T \mapsto \gamma(T)\gamma(2T)^2$ is Galois-equivariant. Since $\alpha(T)\alpha(2T)^2 = \beta(T)^5$, and there are no nontrivial fifth roots of unity in R, this proves (i). Let $\tau \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ be complex conjugation. (Recall that we fixed an embedding $\overline{\mathbb{Q}} \subset \mathbb{C}$.) Since $H^1(\mathbb{R}, E[5]) = 0$ we have $\tau(\gamma)/\gamma = w(\xi_\tau) = w(\tau(S) - S)$

for some $S \in E(\mathbb{C})[5]$. Dividing γ by $w(S)$ now gives (ii). Multiplying γ by $w(T)$ for $T \in E(\mathbb{R})[5]$ does not change $\rho = \partial\gamma$, so in fact we only need to loop over 5^2 choices for γ.

Let T_1, T_2 be a basis for $E[5](\mathbb{C})$ with $\overline{T}_1 = T_1$, $\overline{T}_2 = -T_2$. Then $\zeta = e_5(T_1, T_2)$ is a primitive fifth root of unity. We define

$$h(T_1) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad h(T_2) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \zeta & 0 & 0 & 0 \\ 0 & 0 & \zeta^2 & 0 & 0 \\ 0 & 0 & 0 & \zeta^3 & 0 \\ 0 & 0 & 0 & 0 & \zeta^4 \end{pmatrix},$$

and

$$h : E[5](\mathbb{C}) \to \mathrm{Mat}_5(\mathbb{C}), \quad h(rT_1 + sT_2) = \zeta^{-rs/2} h(T_1)^r h(T_2)^s,$$

where the exponent of ζ is read as an element of $\mathbb{Z}/5\mathbb{Z}$.

We compute the structure constants for A_ρ from the real trivialisation given in [14, §5], that is,

$$A_\rho \otimes \mathbb{R} \xrightarrow{\sim} \mathrm{Mat}_5(\mathbb{R}), \quad z \mapsto \sum_{T \in E[5]} \gamma(T) z(T) h(T).$$

As recommended there, we choose our $\mathbb{Q}$-basis for L to be a $\mathbb{Z}$-basis for $\mathfrak{c}^{-1}$ that is LLL-reduced with respect to the inner product

$$\langle z_1, z_2 \rangle = \sum_{0 \neq T \in E[5]} |\alpha(T)|^{2/5} z_1(T) \overline{z_2(T)}.$$

This makes the structure constants small integers, which are therefore easy to recognise from floating-point approximations. The incorrect choices of γ are quickly discarded since the structure constants do not in general turn out to be integers.

To record our final choice of γ we let T_1, T_2 be the basis for $E(\mathbb{C})[5]$ given (approximately) by

$$T_1 = (1996.32, -87675.66),$$
$$T_2 = (-643.55, 321.77 - 13079.33i).$$

Then γ is the fifth root of α given (approximately) by the following matrix, with entries $\gamma(rT_1 + sT_2)$ for $r, s = 0, \ldots, 4$.

$$\begin{pmatrix} 1.00 & -3.96 + 0.90i & 1.39 - 4.05i & 1.39 + 4.05i & -3.96 - 0.90i \\ -0.92 & 5.87 - 2.18i & 2.39 + 1.96i & 2.39 - 1.96i & 5.87 + 2.18i \\ -2.20 & 4.41 + 3.00i & -3.56 - 4.19i & -3.56 + 4.19i & 4.41 - 3.00i \\ -2.12 & -7.13 - 4.33i & -0.29 + 3.75i & -0.29 - 3.75i & -7.13 + 4.33i \\ 4.44 & 0.14 - 0.12i & -0.96 - 0.48i & -0.96 + 0.48i & 0.14 + 0.12i \end{pmatrix}$$

Although our method for choosing a basis for L as a $\mathbb{Q}$-vector space works well on a computer, the basis vectors (which are elements of $\mathfrak{c}^{-1}$) are extremely messy to write down. To assist in recording some details of the calculation, we replace α and γ by their inverses. Our $\mathbb{Q}$-basis $u_1, \ldots, u_{24}$ for L is now a $\mathbb{Z}$-basis for $\mathfrak{c}$. Its first two elements are $u_1 = u_1'/147$ and $u_2 = u_2'/294$, where

$$\begin{aligned} u_1' = {} & 906u^{11} + 3697u^{10} - 5099u^9 - 18382u^8 + 11847u^7 - 274284u^6 \\ & + 271264u^5 + 284304u^4 + 51522u^3 + 31261u^2 - 4247u - 3174 \end{aligned}$$

and

$$\begin{aligned} u_2' = {} & (640u^{11} + 2621u^{10} - 3565u^9 - 13051u^8 + 8154u^7 - 193589u^6 \\ & + 188894u^5 + 204155u^4 + 40745u^3 + 21338u^2 - 5548u - 2903)v \\ & + (-221u^{11} - 943u^{10} + 1135u^9 + 4972u^8 - 2330u^7 + 65086u^6 \\ & - 53197u^5 - 99488u^4 - 12061u^3 + 13094u^2 + 5473u + 4980). \end{aligned}$$

Then R has basis $r_1, \ldots, r_{25}$, where $r_1 = (1, 0)$ and $r_{i+1} = (0, u_i)$. Let $A_\rho = (R, +, *_\rho)$ with basis $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{25}$ corresponding to $r_1, \ldots, r_{25}$. Note that $\boldsymbol{a}_1$ is the identity. The structure constants turn out to be integers with maximum absolute value 448 and mean absolute value 22.65. As predicted by [14, Lemma 5.2] the order with basis the $\boldsymbol{a}_i$ has discriminant $5^{48} \cdot 7^{16} \cdot 31^{18} = 5^{25} \cdot \mathrm{Disc}(L)$. The basis vectors $\boldsymbol{a}_i$ have minimal polynomials

$$\begin{aligned} & X - 1, \quad X^5 + 435X^3 + 7315X^2 + 835X + 32172, \\ & \qquad\qquad X^5 - 390X^3 - 4885X^2 + 17560X + 1407822, \quad \ldots \end{aligned}$$

If $\alpha \in R^\times/(R^\times)^5$ corresponds to a Selmer group element, then by the local-to-global principle for the Brauer group we have $A_\rho \cong \mathrm{Mat}_5(\mathbb{Q})$. The problem of finding such an isomorphism (called a *trivialisation*) is addressed in [14; 24; 25]. By using Magma to compute a maximal order (and running LLL on the change of basis matrix) we found a basis with minimal polynomials

$$\begin{aligned} & X^2, \quad X^2, \quad X^2, \quad X^4, \quad X^2, \quad X^3, \quad X^2, \quad X^3 - X, \\ & X^5 - 2X^3 + X, \quad X^4 - X^2, \quad X^4 - X^2, \quad X^5 + X^3, \quad X^2, \\ & X^4 - X^2, \quad X^3, \quad X^4 - 2X^2, \quad X^4 - X^2, \quad X^5 - X^3 + X^2 + X, \\ & X^5 - X^3 - 4X^2 + 4X, \quad X^4 - 2X^2 - X, \quad X^4 + X^3 - X^2 - X, \\ & X^5 - X^3, \quad X^5 - 2X^3, \quad X^5 - 5X^2 + X, \quad X^3 + X^2. \end{aligned}$$

Any reducible minimal polynomial gives a zero-divisor in A_ρ, and once we know a zero-divisor it is easy to find a trivialisation. In this way we found a trivialisation τ

that maps $\boldsymbol{a}_1 \mapsto I_5$ and

$$\boldsymbol{a}_2 \mapsto \begin{pmatrix} 13 & -5 & -20 & -20 & -15 \\ -40 & -22 & 40 & 20 & 10 \\ -20 & -35 & 3 & -15 & -15 \\ -15 & 0 & 30 & 13 & 0 \\ 15 & 15 & -10 & -5 & -7 \end{pmatrix}, \quad \boldsymbol{a}_3 \mapsto \begin{pmatrix} -12 & 5 & 0 & 10 & 5 \\ -30 & -42 & 35 & 5 & 0 \\ -50 & -35 & 38 & 20 & 5 \\ -110 & -50 & 65 & 8 & 5 \\ 45 & 45 & -30 & 0 & 8 \end{pmatrix}.$$

These calculations show that the element α of $R^\times/(R^\times)^5$ corresponds to an element of $H^1(\mathbb{Q}, E[5])$ with trivial obstruction. It may therefore be represented by a genus-one normal curve $C \subset \mathbb{P}^4$.

We compute equations for C using the "Hesse pencil method", as described in [12, §5.1]. Let $r_1^*, \ldots, r_{25}^*$ be the basis for R with $\mathrm{Tr}_{R/\mathbb{Q}}(r_i r_j^*) = \delta_{ij}$. It is shown that

$$M = \sum_{i=1}^{25} r_i^* \tau(\boldsymbol{a}_i) \in \mathrm{GL}_5(R) = \mathrm{Map}_{\mathbb{Q}}(E[5], \mathrm{GL}_5(\overline{\mathbb{Q}}))$$

describes the action of $E[5]$ on $C \subset \mathbb{P}^4$. In [20, §12] we gave a practical method for computing all genus-one normal curves $C \subset \mathbb{P}^4$ that have Jacobian E and are invariant under the matrices M_T for $T \in E[5]$. As predicted by [12, Proposition 5.5] there is only one such curve defined over $\mathbb{Q}$. We use the algorithms for minimisation and reduction in [21] to make a final change of coordinates. In this example the model obtained is already minimal, whereas reduction suggests the change of coordinates

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} \leftarrow \begin{pmatrix} -1 & 2 & 1 & -2 & 1 \\ -1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}.$$

The result is $C \subset \mathbb{P}^4$ defined by the 4×4 Pfaffians of

$$\begin{pmatrix} 0 & -x_1+x_2+x_3 & x_1+3x_2+x_4 & -2x_2+x_3+x_5 & 2x_2-2x_3+x_5 \\ & 0 & -x_1-x_2-x_3+x_5 & x_2-x_4+x_5 & -x_2+x_3+x_4 \\ & & 0 & -x_3+x_5 & -x_1+x_3-x_5 \\ & - & & 0 & x_4 \\ & & & & 0 \end{pmatrix}.$$

Computing the invariants, as described in [19], and using Bruin's programs [4] to check local solubility, we find that C represents an element of $Ш(E/\mathbb{Q})[5]$. It is nontrivial since $E(\mathbb{Q})/5E(\mathbb{Q}) = 0$ and $\alpha \notin (R^\times)^5$.

The theory in [12, §3] shows that if $M^5 = \alpha' I_5$ then $\alpha'/\alpha \in (R^\times)^5$. This is a condition we can check exactly. So even though we made use of floating-point approximations (and did not check at the outset that α is in the image of w_1, although methods for doing this are described in [15; 34]), we can be sure that C corresponds to our original choice of α.

Repeating for other choices of α, we found a subgroup of $Ш(E/\mathbb{Q})$ isomorphic to $(\mathbb{Z}/5\mathbb{Z})^2$. For these, and examples for other elliptic curves $E/\mathbb{Q}$ of small conductor, see [22]. The main difficulty in computing further examples is that the computation of class group and units is often prohibitively expensive.

References

[1] Wieb Bosma and John Cannon (eds.), *Discovering mathematics with Magma: Reducing the abstract to the concrete*, Algorithms and Computation in Mathematics, no. 19, Springer, Berlin, 2006. MR 2007h:00016

[2] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system, I: The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[3] N. Bourbaki, *Algèbre. Chapitre* 9: *Formes sesquilinéaires et formes quadratiques*, Actualités Sci. Ind., no. 1272, Hermann, Paris, 1959, reprinted Springer, Berlin, 2007. MR 21 #6384

[4] Nils Bruin, *Some ternary Diophantine equations of signature* $(n, n, 2)$, in Bosma and Cannon [1], 2006, pp. 63–91. MR 2007m:11047

[5] Winfried Bruns and Jürgen Herzog, *Cohen-Macaulay rings*, Cambridge Studies in Advanced Mathematics, no. 39, Cambridge University Press, 1993. MR 95h:13020

[6] David A. Buchsbaum and David Eisenbud, *Algebra structures for finite free resolutions, and some structure theorems for ideals of codimension* 3, Amer. J. Math. **99** (1977), no. 3, 447–485. MR 56 #11983

[7] ———, *Gorenstein ideals of height* 3, Seminar D. Eisenbud/B. Singh/W. Vogel, vol. 2, Teubner-Texte zur Math., no. 48, Teubner, Leipzig, 1982, pp. 30–48. MR 84i:13017

[8] J. W. S. Cassels, *Arithmetic on curves of genus* 1*, IV: Proof of the Hauptvermutung*, J. Reine Angew. Math. **211** (1962), 95–112. MR 29 #1214

[9] Henri Cohen and Fabien Pazuki, *Elementary 3-descent with a 3-isogeny*, Acta Arith. **140** (2009), no. 4, 369–404. MR 2011h:11063

[10] J. E. Cremona, *Algorithms for modular elliptic curves*, 2nd ed., Cambridge University Press, 1997. MR 99e:11068

[11] ———, *Elliptic curve data*, 9999. http://homepages.warwick.ac.uk/staff/J.E.Cremona/ftp/data

[12] J. E. Cremona, T. A. Fisher, C. O'Neil, D. Simon, and M. Stoll, *Explicit n-descent on elliptic curves, I: Algebra*, J. Reine Angew. Math. **615** (2008), 121–155. MR 2009g:11067

[13] ———, *Explicit n-descent on elliptic curves, II: Geometry*, J. Reine Angew. Math. **632** (2009), 63–84. MR 2011d:11128

[14] ———, *Explicit n-descent on elliptic curves*, *III*: *Algorithms*, 2011, to appear in *Math. Comp.* arXiv 1107.3516 [math.NT]

[15] Z. Djabri, Edward F. Schaefer, and N. P. Smart, *Computing the p-Selmer group of an elliptic curve*, Trans. Amer. Math. Soc. **352** (2000), no. 12, 5583–5597. MR 2001b:11047

[16] David Eisenbud, *Commutative algebra: With a view toward algebraic geometry*, Graduate Texts in Mathematics, no. 150, Springer, New York, 1995. MR 97a:13001

[17] Tom Fisher, *Some examples of 5 and 7 descent for elliptic curves over* $\mathbb{Q}$, J. Eur. Math. Soc. (JEMS) **3** (2001), no. 2, 169–201. MR 2002m:11045

[18] ———, *The Cassels-Tate pairing and the Platonic solids*, J. Number Theory **98** (2003), no. 1, 105–155. MR 2003k:11094

[19] ———, *The invariants of a genus one curve*, Proc. Lond. Math. Soc. (3) **97** (2008), no. 3, 753–782. MR 2009j:11087

[20] ———, *The Hessian of a genus one curve*, Proc. Lond. Math. Soc. (3) **104** (2012), no. 3, 613–648. MR 2900238

[21] ———, *Minimisation and reduction of* 5*-coverings of elliptic curves*, Algebra Number Theory **7** (2013), no. 5, 1179–1205.

[22] ———, *Elements of order* 5 *in the Tate-Shafarevich group*, 9999. http://www.dpmms.cam.ac.uk/~taf1000/g1data/order5.html

[23] E. V. Flynn and C. Grattoni, *Descent via isogeny on elliptic curves with large rational torsion subgroups*, J. Symbolic Comput. **43** (2008), no. 4, 293–303. MR 2009c:11080

[24] Gábor Ivanyos, Lajos Rónyai, and Josef Schicho, *Splitting full matrix algebras over algebraic number fields*, J. Algebra **354** (2012), 211–223. MR 2879232

[25] Ádám Lelkes, *Small zero divisors in maximal orders of* $M_n(\mathbf{Q})$, Scientific student conference, Budapest, 2011. http://www.math.bme.hu/~lelkesa/tdk.pdf

[26] Stephen Lichtenbaum, *The period-index problem for elliptic curves*, Amer. J. Math. **90** (1968), 1209–1223. MR 38 #5788

[27] E. Marchionna (ed.), *Questions on algebraic varieties: Lectures given at a Summer School of the Centro Internazionale Matematico Estivo (C.I.M.E.) held in Varenna (Como), Italy, September 7–17, 1969*, C.I.M.E. Summer Schools, no. 51, Berlin, Springer, 2011.

[28] Robert L. Miller and Michael Stoll, *Explicit isogeny descent on elliptic curves*, Math. Comp. **82** (2013), no. 281, 513–529. MR 2983034

[29] David Mumford, *Varieties defined by quadratic equations*, in Marchionna [27], 2011, pp. 29–100. MR 44 #209

[30] Catherine O'Neil, *The period-index obstruction for elliptic curves*, J. Number Theory **95** (2002), no. 2, 329–339, erratum: [31]. MR 2003f:11079

[31] ———, *Erratum to: "The period-index obstruction for elliptic curves"* [*J. Number Theory* **95** (*2002*)*, no. 2, 329–339*], J. Number Theory **109** (2004), no. 2, 390. MR 2005g:11096

[32] The PARI Group, *PARI/GP*, 2012. http://pari.math.u-bordeaux.fr/

[33] Irena Peeva, *Graded syzygies*, Algebra and Applications, no. 14, Springer, London, 2011. MR 2011j:13015

[34] Edward F. Schaefer and Michael Stoll, *How to do a* p*-descent on an elliptic curve*, Trans. Amer. Math. Soc. **356** (2004), no. 3, 1209–1231. MR 2004g:11045

TOM FISHER: T.A.Fisher@dpmms.cam.ac.uk
DPMMS, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, United Kingdom

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.413

msp

On the density of abelian surfaces with Tate-Shafarevich group of order five times a square

Stefan Keil and Remke Kloosterman

Let $A = E_1 \times E_2$ be the product of two elliptic curves over $\mathbb{Q}$, each having a rational 5-torsion point P_i. Set $B := A/\langle (P_1, P_2) \rangle$. In this paper we give an algorithm to decide whether the order of the Tate-Shafarevich group of the abelian surface B is square or five times a square, under the assumptions that we can find a basis for the Mordell-Weil groups of E_1 and E_2 and that the Tate-Shafarevich groups of E_1 and E_2 are finite.

We considered all pairs (E_1, E_2) with prescribed bounds on the conductor and the coefficients in a minimal Weierstrass equation. In total we considered around 20.0 million abelian surfaces, of which 49.16% have Tate-Shafarevich groups of nonsquare order.

1. Introduction

Let A be an abelian variety over a number field K. The Tate-Shafarevich group $Ш(A/K)$ plays an important role in understanding the arithmetic of A. For example, it contains information on the tightness of the upper bound on the Mordell-Weil rank obtained by m-descent. Moreover, the order of this group, which is conjectured to be finite, plays a role in the Birch and Swinnerton-Dyer conjecture.

The Tate-Shafarevich group comes with a pairing, the *Cassels-Tate pairing*, which depends on the choice of a polarization $\lambda : A \to A^\vee$:

$$\langle \cdot, \cdot \rangle_\lambda : Ш(A/K) \times Ш(A/K) \to \mathbb{Q}/\mathbb{Z}.$$

Let $Ш(A/K)_{\mathrm{nd}}$ denote the Tate-Shafarevich group modulo its maximal divisible subgroup. If λ is an isomorphism, that is, A is principally polarized, then the

MSC2010: primary 11G10; secondary 11G40, 14G10, 14K15.
Keywords: Tate-Shafarevich groups, abelian surface, Cassels-Tate equation.

induced pairing on $Ш(A/K)_{\mathrm{nd}}$ is nondegenerate. If moreover this pairing is alternating, then for all primes p the cardinality of the p-primary part $Ш(A/K)_{\mathrm{nd}}[p^\infty]$ is a perfect square; thus, if $Ш(A/K)$ is finite then it is a perfect square.

Tate [18] showed that if λ is an isomorphism and is induced from a K-rational divisor on A, then the Cassels-Tate pairing is in fact alternating, as for example for elliptic curves. However, if $\dim A > 1$ then A may not admit a principal polarization, and even when A is principally polarized this polarization need not be induced by a K-rational divisor on A. Poonen and Stoll [11] showed that in fact there exist genus-2 curves $C/\mathbb{Q}$ such that $\#Ш(J(C)/\mathbb{Q})$ is twice a square. Moreover, they showed that if one assumes that $Ш(J(C)/\mathbb{Q})$ is finite for all genus-2 curves $C/\mathbb{Q}$, then the density of genus-2 curves whose Jacobians have Tate-Shafarevich groups of nonsquare order exists, and is approximately 13%.

For arbitrary abelian varieties Flach [4] showed that if $\#Ш(A/K) = kn^2$, with k square free, then k divides 2 times the degree of every polarization on A. Hence for principally polarized abelian varieties one has that $\#Ш(A/K)$ is either a square or twice a square, if it is finite, but for general abelian varieties there are more possibilities. Stein [17] constructed, for every prime number $p < 25000$ (excluding $p = 2$ and $p = 37$), an example of a $(p-1)$-dimensional abelian variety $A_p/\mathbb{Q}$ such that $\#Ш(A_p) = pn^2$.

We restrict now to the case of $\dim A = 2$. The constructions of Poonen-Stoll and of Stein yield examples of abelian surfaces such that $\#Ш(A/K)$ is a square, twice a square, or three times a square. One might wonder which further possibilities occur. Recently, the first author [6] showed that there exist abelian surfaces such that the Tate-Shafarevich group has order five times a square and seven times a square.

In this paper we will take a closer look at the construction of abelian surfaces with Tate-Shafarevich group of order five times a square. The examples of [6] are members of a two-dimensional family of abelian surfaces with a polarization of degree 5^2. Moreover, one can show that for a general member of this family, every polarization it possesses has degree a multiple of 5; thus they are not a priori excluded by Flach's theorem and might have a Tate-Shafarevich group of order five times a square.

The construction of this family goes as follows. Let (E, O) be an elliptic curve over $\mathbb{Q}$ with a point P of order 5. Then there exists a $d \in \mathbb{Q}^*$ such that $((E, O), P)$ is isomorphic to $((E_d, O), (0, 0))$, where

$$E_d : \ y + (d+1)xy + dy = x^3 + dx^2.$$

Take two numbers $d_1, d_2 \in \mathbb{Q}^*$ and consider $B_{d_1,d_2} := E_{d_1} \times E_{d_2}/\langle (0,0) \times (0,0) \rangle$. Then $A_{d_1,d_2} := E_{d_1} \times E_{d_2} \to B_{d_1,d_2}$ is an isogeny of degree 5. Moreover, if the two elliptic curves are not isogenous, then all polarizations on B_{d_1,d_2} have degree

divisible by 5. The B_{d_1,d_2}'s are the family we consider. In our case we know that $Ш(A_{d_1,d_2}/\mathbb{Q})$ has square order, if it is finite, since it is isomorphic to the product of the two Tate-Shafarevich groups of E_{d_1} and E_{d_2}.

The behavior of the Tate-Shafarevich group under isogenies is well-known. This behavior is part of Tate's proof of the invariance of the Birch and Swinnerton-Dyer conjecture; for more on this see Section 2. The upshot of this is the following: Let $\varphi : A \to B$ be an isogeny and assume that either $\#Ш(A/K)$ or $\#Ш(B/K)$ is finite (which implies that both are finite). Denote by $\varphi^\vee : B^\vee \to A^\vee$ the dual isogeny. For a field $L \supseteq K$ denote by $\varphi_L : A(L) \to B(L)$ the induced map on L-rational points. Let S be a finite set of places containing the primes where A has bad reduction, the infinite places, and the primes dividing the degree of φ. Then the following holds:

$$\frac{\#Ш(A/K)}{\#Ш(B/K)} = \frac{\#\ker\varphi_K \#\operatorname{coker}\varphi_K^\vee}{\#\ker\varphi_K^\vee \#\operatorname{coker}\varphi_K} \prod_{v\in S} \frac{\#\operatorname{coker}\varphi_{K_v}}{\#\ker\varphi_{K_v}}.$$

In Sections 4 and 5 we show that for our choice of abelian surfaces the above-mentioned cardinalities of kernels and cokernels can be determined, provided one has a basis for the Mordell-Weil group of both E_{d_1} and E_{d_2}. (Actually something weaker is enough; see the end of Section 4.) Hence, given bases for the Mordell-Weil groups of both elliptic curves we can determine whether $\#Ш(B/\mathbb{Q})$, if finite, is a square or a nonsquare.

For all pairs (d_1, d_2) with $d_i = u_i/v_i$ where $\max(|u_i|, |v_i|)$ is bounded by $N = 50{,}000$ and where the conductor of E_{d_i} is bounded by $C = 10^6$, we computed this product of cardinalities of kernels and cokernels. There are 2,445,366 such pairs, and 47.01% of these surfaces have a Tate-Shafarevich group of nonsquare order (assuming that $Ш(E_{d_1}/\mathbb{Q})$ and $Ш(E_{d_2}/\mathbb{Q})$ are finite). We also computed these cardinalities for all pairs (d_1, d_2) such that the absolute value of the numerator and denominator of d_i is bounded by $N = 100$. There are 18,522,741 such pairs, and 49.31% of them have Tate-Shafarevich group of nonsquare order. Based on our computations, we expect that the density of abelian surfaces B_{d_1,d_2} with nonsquare Tate-Shafarevich groups exists and is around 50%. For some heuristics see the end of the final section.

The outline of this paper is as follows. In Section 2 we discuss some preliminaries and in Section 3 we explain in more detail the construction of the family of abelian surfaces we consider. In Section 4 we discuss how we can calculate the global quotient and which conditions on E_{d_1} and E_{d_2} are needed for this. In Section 5 we discuss how we calculate the local quotient, which turns out to be a much simpler computation. In Section 6 we sketch the algorithm used for the computations of the densities, and finally in Section 7 we discuss the results we obtain.

2. Preliminaries

Let K be a number field and let G_K be the absolute Galois group $\mathrm{Gal}(\overline{K}/K)$. For a (finite or infinite) place v of K, denote by K_v the completion of K with respect to v and by G_{K_v} the absolute Galois group of K_v.

Let A/K be an abelian variety. Denote by $A^\vee$ the dual abelian variety. Then the *Tate-Shafarevich group* of A/K is defined as

$$Ш(A/K) := \ker\Big(H^1(G_K, A) \to \prod_v H^1(G_{K_v}, A)\Big),$$

where the product is taken over all finite and infinite places of K. Let $\varphi : A \to B$ be an isogeny of abelian varieties. Then the *φ-Selmer group of A/K* is defined as

$$S^\varphi(A/K) := \ker\Big(H^1(G_K, A[\varphi]) \to \prod_v H^1(G_{K_v}, A)\Big).$$

The Tate-Shafarevich group is a torsion group. It is conjectured to be finite, and the φ-Selmer group is known to be finite. The m-torsion subgroup of the Tate-Shafarevich group fits in an exact sequence

$$0 \to A(K)/mA(K) \to S^{[m]}(A/K) \to Ш(A/K)[m] \to 0.$$

That is, it measures the difference between the m-Selmer group and $A(K)/mA(K)$. In theory the m-Selmer group is computable; hence the Tate-Shafarevich group measures the difference between the upper bound on the Mordell-Weil rank obtained by doing m-descent and the actual Mordell-Weil rank of A.

The Tate-Shafarevich group plays also a role in the Birch and Swinnerton-Dyer conjecture:

Conjecture 2.1 (Birch and Swinnerton-Dyer). *Let A/K be an abelian variety and let $L(A, s)$ be its L-series. Set $r := \mathrm{rk}\, A(K)$. Then $Ш(A/K)$ is finite, $L(A, s)$ has a zero of exact order r at $s = 1$, and*

$$\lim_{s\to 1} \frac{L(A, s)}{(s-1)^r} = \frac{2^r \#Ш(A/K) R_A \prod \int_{A(K_v)} |\omega|_v}{\#A(K)_{\mathrm{tor}} \#A^\vee(K)_{\mathrm{tor}}}. \tag{1}$$

The left hand side of (1) is invariant under isogeny. Cassels [2] (for the case $\dim A = 1$) and Tate [18] (for the general case $\dim A \geq 1$) proved that the right hand side is also invariant under isogeny. That is, if $\varphi : A \to B$ is an isogeny then

$$\frac{\#Ш(A/K)}{\#Ш(B/K)} = \frac{R_B \#A(K)_{\mathrm{tor}} \#A^\vee(K)_{\mathrm{tor}} \prod \int_{B(K_v)} |\omega|_v}{R_A \#B(K)_{\mathrm{tor}} \#B^\vee(K)_{\mathrm{tor}} \prod \int_{A(K_v)} |\omega|_v}.$$

This formula was used by Schaefer and the second author [9] to provide examples of elliptic curves with large Selmer groups, by Matsuno [10] and by the second author [8] to provide examples of elliptic curves with large Tate-Shafarevich groups, and by Flynn and Grattoni [5] to compute several Selmer groups.

However, the right hand side of (1) is not well-suited for calculation. One can rewrite the right hand side as follows: For a field $L \supseteq K$, let φ_L denote the group homomorphism $\varphi_L : A(L) \to B(L)$. Then

$$\frac{\#Ш(A/K)}{\#Ш(B/K)} = \frac{\#\ker\varphi_K \#\operatorname{coker}\varphi_K^\vee}{\#\ker\varphi_K^\vee \#\operatorname{coker}\varphi_K} \prod_v \frac{\#\operatorname{coker}\varphi_{K_v}}{\#\ker\varphi_{K_v}}. \tag{2}$$

We will call the first factor (with the φ_K) the *global factor*, and the second factor (with the φ_{K_v}) the *local factor*. If v is a finite prime of good reduction and v does not divides the degree of the isogeny, then $\#\operatorname{coker}\varphi_{K_v} = \#\ker\varphi_{K_v}$; hence the product on the right hand side is a finite product, where only the bad primes, the infinite primes, and the primes dividing the degree of the isogeny need be taken into account.

It is known that if an elliptic curve has analytic rank at most 1, then its Tate-Shafarevich group is finite and its analytic rank is equal to its Mordell-Weil rank. Throughout this paper we will assume that the same is true even for elliptic curves with larger analytic rank.

3. Constructing a family of abelian surfaces

We will construct a two-dimensional family of abelian surfaces B/K, whose members are quotients of products of two elliptic curves E_1, E_2 by an isogeny of degree 5. Therefore $\#Ш(B/K)\cdot 5^a = \#Ш(E_1 \times E_2)$, for some $a \in \mathbb{Z}$. Since $\#Ш(E_1 \times E_2)$ is a square, it follows that $\#Ш(B/K)$ modulo squares is one of $\{1, 5\}$. Additionally, we have that for a general member of this family every polarization has degree divisible by 5. Thus Flach's theorem does not restrict us further.

Let G/K be a group scheme of prime order ℓ. Let E_1, E_2 be two elliptic curves over K such that G is a subgroup scheme of both E_1 and E_2. Let $A = E_1 \times E_2$ and $B = A/G$, where G is embedded diagonally in A. Then the natural isogeny $\varphi : A \to B$ has degree ℓ. Moreover, one can show that either E_1 and E_2 are isogenous or every polarization on B has degree a multiple of ℓ. Hence for general E_1, E_2 we are in the second case.

Consider the case $G = \mathbb{Z}/\ell\mathbb{Z}$; that is, the case in which G is generated by a K-rational point. Since for $\ell > 4$ the functor $Y_1(\ell)$ is representable, one has a universal family of elliptic curves E with a point P of order ℓ. In the case $\ell = 5$ the universal family is given by

$$E_d : \quad y^2 + (d+1)xy + dy = x^3 + dx^2, \quad P = (0,0),$$

for any $d \in K^*$ with $d^2 + 11d - 1 \neq 0$. The four nontrivial 5-torsion points are $(0,0)$, $(-d, d^2)$, $(-d, 0)$, and $(0, -d)$. If we move $(0, -d)$ to $(0,0)$ and bring the curve into standard form we obtain E_d. If we move $(-d, d^2)$ or $(-d, 0)$ to $(0,0)$ and bring the elliptic curve into standard form we obtain $E_{-1/d}$.

We restrict now to the case where $K = \mathbb{Q}$, $\ell = 5$, and G is generated by a $\mathbb{Q}$-rational point. Fix d_1 and d_2 in $\mathbb{Q}^*$ and set $A := E_{d_1} \times E_{d_2}$. The rational 5-torsion subgroup of A has four diagonally embedded subgroups of order 5. Let $G = \mathbb{Z}/5\mathbb{Z}$ be one of those, so that G is the subscheme of A generated by $(0,0) \times [n](0,0)$ for some $n \in \{1, 2, 3, 4\}$. Let $B := A/G$. Then B is a candidate for an abelian surface such that $Ш(B/\mathbb{Q})$ has order five times a square. To actually check whether $Ш(B/\mathbb{Q})$ has nonsquare order we will now calculate both the local and the global factor.

Note that the 16 surfaces $B/\mathbb{Q}$ one obtains by replacing d_i by $-1/d_i$ and using the four values of n break into two sets of 8 isomorphic surfaces. For fixed d_1, d_2 the surfaces corresponding to $n = 1, 4$ lie in one of these isomorphism classes and those for $n = 2, 3$ in the other one. We will see in the next two sections that for fixed d_1, d_2 the size of $Ш(B/\mathbb{Q})$ is independent of n, and so all 16 surfaces will have Tate-Shafarevich groups of the same cardinality. Therefore, for our computations we will only consider the case $d_1, d_2 > 0$ and $n = 1$.

Let A' be the quotient of $E_{d_1} \times E_{d_2}$ by the group scheme generated by $(0,0) \times O$ and $O \times (0,0)$, let E'_{d_i} be the quotient of E_{d_i} by $\langle (0,0) \rangle$, and let η_i be the natural isogeny from E_{d_i} to E'_{d_i}. The natural isogeny $\rho : A \to A'$ factors as $A \to B \to A'$. Consider now the dual picture

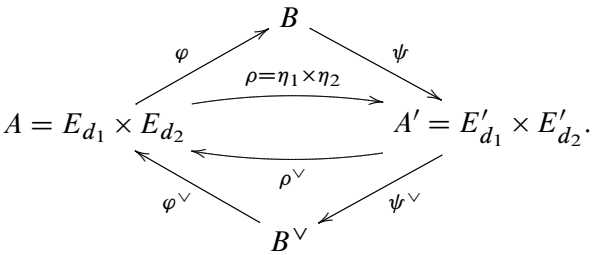
$$(A')^\vee \to B^\vee \to A^\vee.$$

Since A and A' are products of elliptic curves, they are principally polarized. Therefore we have the factorization

$$A' \to B^\vee \to A.$$

The kernel of $A' \to A$ is Cartier dual to the kernel of $A \to A'$, and hence is isomorphic to $(\boldsymbol{\mu}_5)^2$. The kernel of $A' \to B^\vee$ is isomorphic to $\boldsymbol{\mu}_5$ embedded with $(1, -n)$ in $(\boldsymbol{\mu}_5)^2$.

In summary, we have the following diagram:



Lemma 3.1. *Suppose* $L = \mathbb{Q}$. *Then* $\ker \varphi_{\mathbb{Q}} \cong \mathbb{Z}/5\mathbb{Z}$ *and* $\ker \varphi^{\vee}_{\mathbb{Q}} = 0$.

Proof. Since $A[\varphi] = \mathbb{Z}/5\mathbb{Z}$ it follows that $A'[\varphi^{\vee}] = \boldsymbol{\mu}_5$. Taking $\mathbb{Q}$-rational points yields the lemma. □

Lemma 3.2. *Suppose* $L = \mathbb{R}$. *Then* $\ker \varphi_{\mathbb{R}} \cong \mathbb{Z}/5\mathbb{Z}$ *and* $\operatorname{coker} \varphi_{\mathbb{R}} = 0$.

Proof. The first assertion is automatic. The nontrivial element in $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ acts on the fiber of an element of $B(\mathbb{R})$ under $\varphi_{\mathbb{C}}$ either by swapping elements or fixing them. Since the degree of φ is not divisible by 2 at least one element in the fiber is fixed, and hence lies in $A(\mathbb{R})$. □

Let S be the set of primes where A has bad reduction, together with 5. Using the above lemmas it follows that

$$\frac{\#Ш(A/\mathbb{Q})}{\#Ш(B/\mathbb{Q})} = \frac{\#\operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}}{\#\operatorname{coker} \varphi_{\mathbb{Q}}} \prod_{v \in S} \frac{\#\operatorname{coker} \varphi_{\mathbb{Q}_v}}{\#\ker \varphi_{\mathbb{Q}_v}}.$$

In other words, in our situation the global factor from (2) simplifies, and we do not need to consider the local factor at infinity. In the next two sections we will explain how to determine the global and local factors.

4. Determining the global factor

To determine

$$\frac{\#\operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}}{\#\operatorname{coker} \varphi_{\mathbb{Q}}}$$

we assume for the moment that we have a basis for the Mordell-Weil groups $E_{d_1}(\mathbb{Q})$, $E_{d_2}(\mathbb{Q})$, $E'_{d_1}(\mathbb{Q})$, and $E'_{d_2}(\mathbb{Q})$. We will now explain how one can determine $\operatorname{coker} \varphi_{\mathbb{Q}}$ and $\operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}$ from this information.

Using the factorization $\rho^{\vee} = \varphi^{\vee} \circ \psi^{\vee}$ we obtain a surjective homomorphism $\operatorname{coker} \rho^{\vee}_{\mathbb{Q}} \to \operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}$. With Hilbert's Theorem 90 we obtain

$$H^1(G_{\mathbb{Q}}, A'[\rho^{\vee}]) = H^1(G_{\mathbb{Q}}, \boldsymbol{\mu}_5^2) = (\mathbb{Q}^*/\mathbb{Q}^{*5})^2,$$
$$H^1(G_{\mathbb{Q}}, B^{\vee}[\varphi^{\vee}]) = H^1(G_{\mathbb{Q}}, \boldsymbol{\mu}_5) = \mathbb{Q}^*/\mathbb{Q}^{*5}.$$

Under these identifications, the surjection $\operatorname{coker} \rho^{\vee}_{\mathbb{Q}} \to \operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}$ becomes the map $(x, y) \mapsto x^n/y$ from $(\mathbb{Q}^*/\mathbb{Q}^{*5})^2$ to $\mathbb{Q}^*/\mathbb{Q}^{*5}$. One sees immediately that the image of this map is independent of n, so to compute $\operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}$ we may as well set $n = 1$. In order to determine $\operatorname{coker} \varphi^{\vee}_{\mathbb{Q}}$ it suffices to determine a basis in $\mathbb{Q}^*/\mathbb{Q}^{*5}$ for $\operatorname{coker} \eta^{\vee}_{1,\mathbb{Q}}$ and $\operatorname{coker} \eta^{\vee}_{2,\mathbb{Q}}$. By following [15, Exercise 10.1], this can be done quite easily: Suppose that f is a function on E_{d_i} with divisor $5(0,0) - 5O$. Then there exists a unique constant $c \in \mathbb{Q}^*/\mathbb{Q}^{*5}$ such that the map

$$\operatorname{coker} \eta^{\vee}_{i,\mathbb{Q}} \to \mathbb{Q}^*/\mathbb{Q}^{*5}$$

that sends $P \neq (0,0)$, O to $cf(P) \bmod \mathbb{Q}^{*5}$ is a well-defined injective group homomorphism, with image equal to the image of the natural embedding of $\operatorname{coker} \eta_{i,\mathbb{Q}}^{\vee}$ into $H^1(G_{\mathbb{Q}}, E'_{d_i}[\eta_i^{\vee}]) \cong \mathbb{Q}^*/\mathbb{Q}^{*5}$. In our case we can take the function f to be $-x^2 + y + xy$ and the constant c to be 1. The point $(0,0)$ is mapped to d^{-1} and O to 1 by linearity.

An element of $\mathbb{Q}^*/\mathbb{Q}^{*5}$ is determined by its valuations at each prime. Write $d = u/v$ and let S be the set of all primes p dividing five times the minimal discriminant of E_d, that is, $p \mid 5uv(u^2 + 11uv - v^2)$. Define

$$\mathbb{Q}(S,5) := \{x \in \mathbb{Q}^*/\mathbb{Q}^{*5} \mid v_p(x) \equiv 0 \bmod 5 \text{ for all } p \notin S\}.$$

From the same exercise from [15] it follows that $f(\operatorname{coker} \eta_{\mathbb{Q}}^{\vee}) \subset \mathbb{Q}(S,5)$. Hence we can represent an element of $\operatorname{coker} \eta_{\mathbb{Q}}^{\vee}$ by its valuation at each prime number $p \in S$. Once the cokernels of both $\eta_{i,\mathbb{Q}}^{\vee}$ are established, the cokernel of $\varphi_{\mathbb{Q}}^{\vee}$ can be computed easily.

To determine the cokernel of $\varphi_{\mathbb{Q}}$ we use the exact sequence

$$0 \to \ker(\psi_{\mathbb{Q}})/\varphi(\ker \rho_{\mathbb{Q}}) \to \operatorname{coker} \varphi_{\mathbb{Q}} \xrightarrow{\psi} \operatorname{coker} \rho_{\mathbb{Q}} \to \operatorname{coker} \psi_{\mathbb{Q}} \to 0.$$

Note that $\ker(\psi_{\mathbb{Q}}) = \varphi(\ker \rho_{\mathbb{Q}})$. Set $K := \mathbb{Q}(\zeta_5)$, where ζ_5 is a primitive fifth root of unity. Then the restriction map $H^1(G_{\mathbb{Q}}, \mathbb{Z}/5\mathbb{Z}) \to H^1(G_K, \mathbb{Z}/5\mathbb{Z})$ is injective, because its kernel has exponent dividing both $[K : \mathbb{Q}] = 4$ and $\#\mathbb{Z}/5\mathbb{Z}$. Since $A[\varphi]$, $A[\rho]$, and $B[\psi]$ are isomorphic over K to $\boldsymbol{\mu}_5$, $\boldsymbol{\mu}_5 \times \boldsymbol{\mu}_5$, and $\boldsymbol{\mu}_5$ (respectively), we obtain the following commutative diagram, where the vertical maps are embeddings:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \operatorname{coker} \varphi_{\mathbb{Q}} & \xrightarrow{\psi} & \operatorname{coker} \rho_{\mathbb{Q}} & \longrightarrow & \operatorname{coker} \psi_{\mathbb{Q}} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & K^*/K^{*5} & \longrightarrow & (K^*/K^{*5})^2 & \longrightarrow & K^*/K^{*5} & \longrightarrow & 0. \end{array} \tag{3}$$

As above, the third of the lower horizontal maps is just $(x,y) \mapsto x^n/y$. Hence, to determine the cokernel of $\varphi_{\mathbb{Q}}$ it suffices to determine the kernel of x^n/y on $\operatorname{coker} \eta_{1,\mathbb{Q}} \times \operatorname{coker} \eta_{2,\mathbb{Q}} \to \operatorname{coker} \psi_{\mathbb{Q}}$. Again this is independent of n, so we may take $n = 1$. We compute the kernel as follows:

(1) For some $\tilde{d} \in K$ there is a K-isomorphism $\tau : E'_d \to E_{\tilde{d}}$ that sends a generator of $\ker \eta^{\vee}$ to $(0,0)$. The map $f : E'_d \to K^*/K^{*5}$ is then

$$P \mapsto -x(\tau(P))^2 + y(\tau(P)) + x(\tau(P))y(\tau(P)).$$

Hence we have to determine τ. This can be done easily for each individual curve E'_d.

(2) To represent elements in $\operatorname{coker}\eta_{\mathbb{Q}} \subset K^*/K^{*5}$, note that the class number of K^* equals 1. Set

$$K(S,5) := \{x \in K^*/K^{*5} \mid v_{\mathfrak{p}}(x) \equiv 0 \bmod 5 \text{ for all } \mathfrak{p} \notin S\},$$

where S contains all primes $\mathfrak{p}$ of K that are bad primes for E_d or that divide 5; that is, all primes $\mathfrak{p}$ of K lying over a prime p of $\mathbb{Q}$ such that

$$p \mid 5uv(u^2 + 11uv - v^2).$$

From [15, Exercise 10.9] it follows that $f(\operatorname{coker}\eta_{\mathbb{Q}}) \subset K(S,5)$. Hence to represent elements in $\operatorname{coker}\eta_{\mathbb{Q}}$ we have to fix a generator $t_{\mathfrak{p}}$ for each prime $\mathfrak{p} \in S$, and we have to fix generators for the unit group of K modulo fifth powers. The field K is well-understood, and it is easy to see that its unit group is generated by $-\zeta_5$ and $(1+\zeta_5)$. Hence we can write

$$f(P) \equiv \zeta_5^{a_0}(1+\zeta_5)^{a_1} \prod_{\mathfrak{p}\in S} t_{\mathfrak{p}}^{v_{\mathfrak{p}}(f(P))}$$

modulo fifth powers.

Remark. We can weaken the assumption of having a basis for the Mordell-Weil groups $E_{d_1}(\mathbb{Q})$, $E_{d_2}(\mathbb{Q})$, $E'_{d_1}(\mathbb{Q})$, and $E'_{d_2}(\mathbb{Q})$. It is actually sufficient to just have generators of finite-index sublattices of these four groups, such that the indices are not divisible by 5; that is, the generators of infinite order are not divisible by 5 modulo torsion. Such sublattices suffice because their images in the cokernels of $\eta_i^\vee$, respectively η_i, are the entire cokernels. Also, it is sufficient to just know such sublattices for $E_{d_1}(\mathbb{Q})$ and $E_{d_2}(\mathbb{Q})$, because suitable dual sublattices can be easily computed using the isogenies η_i. One only has to calculate the images of the generators under η_i and then check whether their span contains points divisible by 5 modulo torsion.

5. Determining the local factor

We want to calculate

$$\frac{\#\operatorname{coker}\varphi_{\mathbb{Q}_p}}{\#\ker\varphi_{\mathbb{Q}_p}}$$

for all bad primes p and for $p = \deg\varphi = 5$. Since the kernel of $\varphi_{\mathbb{Q}_p}$ is generated by a $\mathbb{Q}$-rational point it follows that $\#\ker\varphi_{\mathbb{Q}_p} = 5$. The size of the cokernel of $\varphi_{\mathbb{Q}_p}$ depends on the reduction of E_{d_1} and E_{d_2}, but turns out to be independent of n.

For $\eta := \eta_i$, we first describe how $\operatorname{coker}\eta_{\mathbb{Q}_p}$ depends on the reduction type of $E := E_{d_i}$. Write $d_i =: u/v$ with $u, v \in \mathbb{Z}$ and $\gcd(u,v) = 1$. Then E has global minimal equation

$$E : y^2 + (u+v)xy + uvy = x^3 + uv^2x^2$$

and discriminant $-(uv)^5(u^2+11uv-v^2)$.

Lemma 5.1. *The elliptic curve E has the following reduction type at a prime p.*

(1) *If $p \mid uv$ then the reduction is split multiplicative and the point $(0,0)$ does not lie on the identity component of the Néron model of E.*

(2) *If $p \mid u^2+11uv-v^2$ then $(0,0)$ lies on the identity component of the Néron model of E and either $p=5$, or $p \equiv \pm 1 \bmod 5$ holds. If $p=5$ the reduction is additive, if $p \equiv 1 \bmod 5$ then the reduction is split multiplicative, and if $p \equiv 4 \bmod 5$ then the reduction type is nonsplit multiplicative.*

Proof. Let $\overline{E}$ be $E \bmod p$ and let $\overline{E}_{\mathrm{ns}}$ be the smooth locus of $\overline{E}$. If $p \mid uv$ then $\overline{E}$ has equation $y^2+\alpha xy = x^3$ for some nonzero $\alpha \in \mathbb{Z}/p\mathbb{Z}$. In particular, $(0,0) \bmod p$ is a node of $\overline{E}$ and the tangent cone is generated by $x=-\alpha y$ and $y=0$, hence the reduction is split multiplicative. Since $(0,0)$ reduces to the singular point of $\overline{E}$ this point does not lie on the identity component of the Néron model of E.

If $p \mid u^2+11uv-v^2$ then the reduction of $(0,0)$ is both on $\overline{E}_{\mathrm{ns}}$ and is nontrivial. In particular the order of the reduction of $(0,0)$, which is 5, divides $\#\overline{E}_{\mathrm{ns}}(\mathbb{F}_p)$. If the reduction is split multiplicative this group has order $p-1$, if the reduction is nonsplit this group has order $p+1$, and if the reduction is additive this group has order p; that is, $p \equiv 1 \bmod 5$, $p \equiv -1 \bmod 5$, and $p=5$ respectively. □

Let $E' := E'_{d_i}$ be the isogenous elliptic curve. Denote by $c_{E,p}$ and $c_{E',p}$ the local Tamagawa numbers, that is, the number of components of the Néron model. We refer to the ratio of $c_{E',p}$ to $c_{E,p}$ as the *Tamagawa quotient*.

Lemma 5.2. *For the Tamagawa quotient we have*

$$\frac{c_{E',p}}{c_{E,p}} = \begin{cases} 1/5 & \text{if } p \mid uv, \\ 5 & \text{if } p \mid u^2+11uv-v^2 \text{ and } p \equiv 1 \bmod 5, \\ 1 & \text{otherwise.} \end{cases}$$

Proof. Since η has degree 5 it follows that $c_{E',p}/c_{E,p} = 5^a$ for some $a \in \mathbb{Z}$. If the reduction is different from split multiplicative then $c_{E,p}$ and $c_{E',p}$ are at most 4, hence $a=0$ and $c_{E,p} = c_{E',p}$.

In [6, Proposition 2.16] it is shown by using Tate curves that if the reduction is split multiplicative then $a \in \{-1,1\}$, depending on whether or not the kernel is on the identity component of the Néron model. □

If $p \nmid \deg \eta = 5$ then from [12, Lemma 3.8] it follows that

$$\frac{\#\operatorname{coker} \eta_{\mathbb{Q}_p}}{\#\ker \eta_{\mathbb{Q}_p}} = \frac{c_{E',p}}{c_{E,p}}.$$

Using this, we easily obtain the following lemma:

Lemma 5.3. *Suppose p is a prime different from 5. We have*

$$\operatorname{coker}\eta_{\mathbb{Q}_p} \cong \begin{cases} \mathbb{Z}/5\mathbb{Z} & \text{if } p \text{ is good for } E, \\ 0 & \text{if } p \mid uv, \\ (\mathbb{Z}/5\mathbb{Z})^2 & \text{if } p \mid u^2 + 11uv - v^2 \text{ and } p \equiv 1 \bmod 5, \\ \mathbb{Z}/5\mathbb{Z} & \text{if } p \mid u^2 + 11uv - v^2 \text{ and } p \equiv 4 \bmod 5. \end{cases}$$

Now $\operatorname{coker}\eta_{\mathbb{Q}_p} \subset H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})$. From Theorem 2 and Proposition 17 of [14, §II.5] it follows that for $p \nmid \deg\eta = 5$ we have

$$\#H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z}) = \#H^0(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})\,\#H^0(G_{\mathbb{Q}_p}, \boldsymbol{\mu}_5) = 5^a,$$

where $a = 1$ if $p \equiv 4 \bmod 5$ and $a = 2$ if $p \equiv 1 \bmod 5$. From this we deduce the following:

Proposition 5.4. *Suppose that $p \neq 5$ is a prime dividing $u^2 + 11uv - v^2$ (so that E has bad reduction at p). Then* $\operatorname{coker}\eta_{\mathbb{Q}_p} = H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})$.

We now return to our abelian surface A. The above proposition enables us to determine $\operatorname{coker}\varphi_{\mathbb{Q}_p}$ for bad primes different from 5.

Proposition 5.5. *Suppose p is a prime of bad reduction for A and $p \neq 5$. Then*

$$\operatorname{coker}\varphi_{\mathbb{Q}_p} \cong \begin{cases} 0 & \text{if } p \mid u_1v_1u_2v_2, \\ (\mathbb{Z}/5\mathbb{Z})^2 & \text{if } p \mid \gcd(u_1^2 + 11u_1v_1 - v_1^2, u_2^2 + 11u_2v_2 - v_2^2) \\ & \quad \text{and } p \equiv 1 \bmod 5, \\ \mathbb{Z}/5\mathbb{Z} & \text{otherwise}. \end{cases}$$

Proof. Recall that

$$\operatorname{coker}\varphi_{\mathbb{Q}_p} = \ker(\operatorname{coker}\eta_{1,\mathbb{Q}_p} \times \operatorname{coker}\eta_{2,\mathbb{Q}_p} \to \operatorname{coker}\psi_{\mathbb{Q}_p}),$$

which equals

$$(\operatorname{coker}\eta_{1,\mathbb{Q}_p} \times \operatorname{coker}\eta_{2,\mathbb{Q}_p}) \cap \ker\big(H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})^2 \to H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})\big).$$

The surjective map $H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})^2 \to H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})$ is given by $(x, y) \mapsto nx - y$. Suppose that $p \mid u_1v_1u_2v_2$. Then by Lemma 5.3 we have $\operatorname{coker}\eta_{i,\mathbb{Q}_p} = 0$ for at least one i, and therefore $\operatorname{coker}\varphi_{\mathbb{Q}_p} = 0$.

Suppose now $p \nmid u_1v_1u_2v_2$. By assumption one of the E_{d_i}, say E_{d_1}, has bad reduction at p. Since $p \nmid 5u_1v_1$ it follows from the above proposition that $\operatorname{coker}\eta_{1,\mathbb{Q}_p} = H^1(G_{\mathbb{Q}_p}, \mathbb{Z}/5\mathbb{Z})$ and hence $\operatorname{coker}\varphi_{\mathbb{Q}_p} \cong \operatorname{coker}\eta_{2,\mathbb{Q}_p}$. Now E_{d_2} has either additive or good reduction. The reduction of E_{d_2} is additive if and only if $p \mid \gcd(u_1^2 + 11u_1v_1 - v_1^2, u_2^2 + 11u_2v_2 - v_2^2)$. Now apply Lemma 5.3 to deduce the structure of $\operatorname{coker}\eta_{2,\mathbb{Q}_p}$, hence the structure of $\operatorname{coker}\varphi_{\mathbb{Q}_p}$. □

It remains to check the case $p = 5$. As before, we first have a look at the elliptic curve E. If $5 \mid uv$ then as above the reduction is split multiplicative and $c_{E',p}/c_{E,p} = 1/5$. Using Tate curves one easily shows that $\operatorname{coker} \eta_{\mathbb{Q}_p} = 0$.

If $5 \mid u^2 + 11uv - v^2$ then the reduction is additive. In particular, the component groups of E and E' have the same order, which is also the case if the reduction is good. Therefore $c_{E',p}/c_{E,p} = 1$. The isogeny $\eta : E \to E'$ can be written as a power series in one variable in a neighborhood of the point O. Again from [12, Lemma 3.8] it follows that

$$\frac{\#\operatorname{coker} \eta_{\mathbb{Q}_5}}{\#\ker \eta_{\mathbb{Q}_5}} = |\eta'(0)|_5^{-1},$$

where $|\eta'(0)|_5$ is the normalized 5-adic absolute value of the leading coefficient of the power series representation of η evaluated at 0. This can be easily computed using Vélu's algorithm [19]. In Lemma 4.1 and Proposition 4.2 of [6] it is shown that in the additive case we have $v_5(u^2 + 11uv - v^2) \in \{2, 3\}$; furthermore, if $v_5(u^2 + 11uv - v^2) = 2$ then $|\eta'(0)|_5 = 1$, while if $v_5(u^2 + 11uv - v^2) = 3$ then $|\eta'(0)|_5 = 1/5$. If E has good reduction at $p = 5$ then it follows that $\#\operatorname{coker} \eta_{\mathbb{Q}_p} = \#\ker \eta_{\mathbb{Q}_p}$, because in this case we also have $|\eta'(0)|_5 = 1$. We summarize as follows.

Lemma 5.6. *We have*

$$\operatorname{coker} \eta_{\mathbb{Q}_p} \cong \begin{cases} \mathbb{Z}/5\mathbb{Z} & \text{if } 5 \text{ is good for } E, \\ 0 & \text{if } 5 \mid uv, \\ (\mathbb{Z}/5\mathbb{Z})^2 & \text{if } 5^3 \mid u^2 + 11uv - v^2, \\ \mathbb{Z}/5\mathbb{Z} & \text{if } 5 \mid u^2 + 11uv - v^2 \text{ and } 5^3 \nmid u^2 + 11uv - v^2. \end{cases}$$

Now we can calculate $\operatorname{coker} \varphi_{\mathbb{Q}_p}$ in the remaining case $p = 5$.

Lemma 5.7. *We have*

$$\#\operatorname{coker} \varphi_{\mathbb{Q}_5} = \begin{cases} 1 & \text{if } 5 \mid u_1 v_1 u_2 v_2, \\ 5^2 & \text{if } 5^3 \mid \gcd(u_1^2 + 11u_1 v_1 - v_1^2, u_2^2 + 11u_2 v_2 - v_2^2), \\ 5 & \text{otherwise.} \end{cases}$$

Proof. If $\operatorname{coker} \eta_{i,\mathbb{Q}_5} = 0$ for one i, then $\operatorname{coker} \varphi_{\mathbb{Q}_5} = 0$. The first condition is equivalent to $5 \mid u_1 v_1 u_2 v_2$.

Suppose now that $\operatorname{coker} \eta_{i,\mathbb{Q}_5} \neq 0$ for both i, which implies that $p = 5$ is additive or good for E_{d_i}. From Proposition 18 and Theorem 5 of [14, §II.5], we find that $H^1(G_{\mathbb{Q}_5}, \mathbb{Z}/5\mathbb{Z}) = (\mathbb{Z}/5\mathbb{Z})^2$ and $H^1_{\mathrm{nr}}(G_{\mathbb{Q}_5}, \mathbb{Z}/5\mathbb{Z}) = \mathbb{Z}/5\mathbb{Z}$. As in the proof of Proposition 5.5, we have that if $\operatorname{coker} \eta_{1,\mathbb{Q}_5} = H^1(G_{\mathbb{Q}_5}, \mathbb{Z}/5\mathbb{Z})$, then $\operatorname{coker} \varphi_{\mathbb{Q}_5} \cong \operatorname{coker} \eta_{2,\mathbb{Q}_5}$ and vice versa. This gives the second case of the lemma, since $\operatorname{coker} \eta_{i,\mathbb{Q}_5} = (\mathbb{Z}/5\mathbb{Z})^2$ if and only if $5^3 \mid u_i^2 + 11u_i v_i - v_i^2$, and $\operatorname{coker} \eta_{i,\mathbb{Q}_5} = \mathbb{Z}/5\mathbb{Z}$ otherwise.

It remains to consider $\operatorname{coker}\eta_{1,\mathbb{Q}_5} = \operatorname{coker}\eta_{2,\mathbb{Q}_5} = (\mathbb{Z}/5\mathbb{Z})$. In this case one can show that $\operatorname{coker}\eta_{i,\mathbb{Q}_5} = H^1_{\mathrm{nr}}(G_{\mathbb{Q}_5}, \mathbb{Z}/5\mathbb{Z})$, for both i; see [6, Propositions 2.10 and 3.5; 13, §3]. Thus the kernel of $\operatorname{coker}\eta_{1,\mathbb{Q}_5} \times \operatorname{coker}\eta_{2,\mathbb{Q}_5} \to \operatorname{coker}\psi_{\mathbb{Q}_5}$, which equals $\operatorname{coker}\varphi_{\mathbb{Q}_5}$, has five elements. This finishes the proof. □

Putting everything together yields the following proposition:

Proposition 5.8. *Let p be a prime. Then*

$$\frac{\#\operatorname{coker}\varphi_{\mathbb{Q}_p}}{\#\ker\varphi_{\mathbb{Q}_p}}$$

is a nonsquare if and only if one of the following occurs:

(1) $p \mid u_1v_1u_2v_2$,

(2) $p \mid \gcd(u_1^2 + 11u_1v_1 - v_1^2, u_2^2 + 11u_2v_2 - v_2^2)$ *and* $p \equiv 1 \bmod 5$, *or*

(3) $p^3 \mid \gcd(u_1^2 + 11u_1v_1 - v_1^2, u_2^2 + 11u_2v_2 - v_2^2)$ *and* $p = 5$.

6. Algorithm

In this section we present the algorithm that we used to produce the databases of abelian surfaces that we studied. Our code was implemented in Sage [16] and is available at [7]. The algorithm consists of two main steps and an initialization step, which we call step 0. In step 1 one creates a database of elliptic curves having a point P of order 5, which are parametrized by two coprime positive integers (u, v). One has to specify which pairs (u, v) one wants to consider. In step 2 one takes such a database of elliptic curves E_d, for $d = u/v$, goes over all pairs of these curves and determines whether the order of the Tate-Shafarevich group of the abelian surfaces $B_{d_1,d_2} = E_{d_1} \times E_{d_2}/\langle(P_1, P_2)\rangle$ is a square. For trivial reasons, pairs of the same elliptic curve are omitted and pairs are considered to be without order.

Algorithm 6.1.

Input: A height bound N and, optionally, a conductor bound C.

Output: The list of all unordered pairs $\{d_1, d_2\}$, where d_1 and d_2 are distinct positive rationals of height at most N such that the elliptic curves E_{d_1} and E_{d_2} have conductor at most C, together with an indication of whether $Ш(B_{d_1,d_2}/\mathbb{Q})$ has square order.

0. *Initialization.* Fix a (large) integer M. For each prime number $p \le M$ determine the prime ideals $\mathfrak{p}$ of $K = \mathbb{Q}(\zeta_5)$ above p and fix an ordering of them. Then fix for each prime ideal $\mathfrak{p}$ a generator $t_{\mathfrak{p}}$.

1. *Creation of a database $\mathcal{D}$ of elliptic curves.* For each pair of coprime positive integers (u, v) such that $\max(u, v) \leq N$, set $E := E_d$, where $d = u/v$. If no conductor bound is given or the conductor of E is at most C, do the following:
   (a) Collect all the primes dividing $5uv(u^2 + 11uv - v^2)$ in a set S.
   (b) Collect all the primes dividing uv in a set T.
   (c) Collect all the primes $p \equiv 1 \bmod 5$ dividing $u^2 + 11uv - v^2$ in a set U.
   (d) If $v_5(u^2 + 11uv - v^2) = 3$, add $p = 5$ to the set U.
   (e) Determine the analytic rank r of E.
   (f) Determine a system of r generators of a sublattice Λ of $E(\mathbb{Q})$, such that the points of infinite order modulo torsion are not divisible by 5. Take the image of Λ in $\mathbb{Q}(S, 5)$ to determine a basis P of $\operatorname{coker} \eta_{\mathbb{Q}}^{\vee} \subset \mathbb{Q}(S, 5)$. The data for each basis element consists of a pair for each prime in S, where the first entry is the corresponding element in S and the second entry is the exponent as an element in $\mathbb{Z}/5\mathbb{Z}$.
   (g) Calculate the image of Λ under η in $E'(\mathbb{Q})$ and determine which image points are divisible by 5 modulo torsion. Divide if possible and determine the nontrivial 5-torsion points of $E'(\mathbb{Q})$ to get a sublattice Λ' of $E'(\mathbb{Q})$, such that the points of infinite order modulo torsion are not divisible by 5. Use this information to compute $\dim \operatorname{coker} \eta_{\mathbb{Q}}$.
   (h) Take the image of Λ' in $K(S, 5)$ to determine a basis Q for $\operatorname{coker} \eta_{\mathbb{Q}} \subset K(S, 5)$. The data for each basis element consists of a pair for each prime in S and a pair for the units. For the primes p in S, the first entry is p and the second entry is a list of elements in $\mathbb{Z}/5\mathbb{Z}$, containing as many entries as there are prime ideals $\mathfrak{p}$ in K over p; for the units, the first element is 1 and the second is the list of exponents of the units.
   (i) Append $((u, v), S, T, U, P, Q)$ to the database $\mathcal{D}$.
2. *Determination of surfaces with* Ш *of nonsquare order.* For each pair

$$((u_1, v_1), S_1, T_1, U_1, P_1, Q_1) \quad \text{and} \quad ((u_2, v_2), S_2, T_2, U_2, P_2, Q_2)$$

   of distinct elements $\mathcal{D}$ (modulo ordering), do the following:
   (a) Set $L := \#(U_1 \cap U_2) - \#(T_1 \cup T_2)$.
   (b) Fix an ordering for $\mathcal{S} := S_1 \cup S_2$.
   (c) Write out the elements from $P_1 \cup P_2$ into a matrix with respect to $\mathcal{S}$. This gives a matrix with entries in $\mathbb{Z}/5\mathbb{Z}$. Calculate the rank of this matrix, which equals the dimension of $\operatorname{coker} \varphi_{\mathbb{Q}}^{\vee}$.
   (d) Write out the elements from $Q_1 \cup Q_2$ into a matrix with respect to the prime ideals $(t_{\mathfrak{p}})$ lying over the primes of $\mathcal{S}$ (and with respect to the units).

This gives a matrix with entries in $\mathbb{Z}/5\mathbb{Z}$. Calculate the rank of this matrix, which equals the dimension of coker $\psi_{\mathbb{Q}}$.

(e) Set $G := \dim \operatorname{coker} \varphi_{\mathbb{Q}}^{\vee} - \dim \operatorname{coker} \eta_{1,\mathbb{Q}} - \dim \operatorname{coker} \eta_{2,\mathbb{Q}} + \dim \operatorname{coker} \psi_{\mathbb{Q}}$. (We have $\dim \operatorname{coker} \varphi_{\mathbb{Q}} = \dim \operatorname{coker} \eta_{1,\mathbb{Q}} + \dim \operatorname{coker} \eta_{2,\mathbb{Q}} - \dim \operatorname{coker} \psi_{\mathbb{Q}}$ from the sequence (3), so $G = \dim \operatorname{coker} \varphi_{\mathbb{Q}}^{\vee} - \dim \operatorname{coker} \varphi_{\mathbb{Q}}$.)

(f) Output $(d_1, d_2, L+G \bmod 2)$, where $d_i = u_i/v_i$.

Remark. The final step is justified as follows: The local factor (without the infinite prime) is a nonsquare if and only if L is odd, and the global factor (without the kernels) is a nonsquare if and only if G is odd. Since the contribution of the infinite prime and the kernels cancel, we have that $Ш(B_{d_1,d_2}/\mathbb{Q})$ has nonsquare order if and only if $L + G$ is odd.

The databases we constructed and the results we obtained are summarized in the following section. To conclude this section, we make some comments on our implementation.

In the cases we considered, Step 0 is not computationally demanding. For example, on a desktop computer it may take some seconds up to a few minutes to compute all generators for all prime ideals of K lying over all primes up to 500,000. Step 2 is also no problem. It consists only of simple set operations and the calculation of the ranks of small matrices with coefficients in $\mathbb{Z}/5\mathbb{Z}$. A few million pairs of elliptic curves can be considered in under an hour.

The computationally demanding part is step 1. There are two main issues. The most problematic calculation is the determination of r generators of a finite index subgroup of the Mordell-Weil group, where r is the analytic rank. We used the standard Sage method `E.point_search(height_limit=18,rank_bound=r)`, and in case this did not come up with enough points we tried some of the remaining curves with `E.gens()`. In several cases these methods did not provide an answer within 48 hours on a single CPU. For these curves we used the method `MordellWeilShaInformation()` in Magma [1], which could handle all our problematic curves in a few seconds each.

The second problematic calculation in the actual code is the computation of the image of coker $\eta_{\mathbb{Q}}$ in $K(S, 5)$. The computation involves factoring ideals of K that are generated by elements of possibly very big norm. For example, the curve E_d, for $d = 1/94$, has analytic rank 1; the numerator and denominator of the image of the point of infinite order in $K(S, 5)$ each have about 600 digits, and Sage was not able to factor the corresponding ideal. As we already knew that the image was trivial, since the dimension of coker $\eta_{\mathbb{Q}}$ was zero, we could skip this calculation. Considering this additional information in the algorithm allowed us to deal with all of the curves we tried. This problem might be avoidable by trying another strategy working modulo primes. The rest of step 1 is not a problem for moderately

chosen $d = u/v$, because it consists mainly of finding the prime factorizations of integers and of rational polynomials of degree 25 (to divide points by 5), as well as calculating isogenies and analytic ranks. In a few hours on a desktop computer, one could produce a database of a few thousand curves.

Remark. At various places in the algorithm we need to assume the Birch and Swinnerton-Dyer conjecture. In step 1(e) we compute the analytic rank of an elliptic curve. To actually compute the analytic rank of a curve E of analytic rank r, we need to assume that the Birch and Swinnerton-Dyer conjecture holds for all elliptic curves with analytic rank at most $r-2$ and that the Mordell-Weil rank of E is at least the analytic rank minus 1. Step 1(f) terminates if and only if the analytic rank of E is at least the Mordell-Weil rank of E.

A second place where we use the Birch and Swinnerton-Dyer conjecture is in the computation of the quantity G in step 2(e). For this we have to assume that for both curves under consideration the analytic rank is precisely the Mordell-Weil rank. However, if we have come this far in the algorithm then we know already that the Mordell-Weil rank is at least the analytic rank.

One may replace steps 1(e) and 1(f) by an algorithm that actually computes a basis for the Mordell-Weil group. This would make the output of the algorithm unconditional. However, in the sample we take below, all elliptic curves have analytic rank at most 3, and for each of them step 1(f) terminated. Hence, to speed up our computations we preferred to determine analytic ranks rather than do full descents.

For the elliptic curves of analytic rank at least 2 we have also to assume that the Tate-Shafarevich group is finite. If this group were infinite then our algorithm would detect whether $\#\ker\varphi^*/\#\operatorname{coker}\varphi^*$ is a square. Here φ^* is the induced morphism on the Tate-Shafarevich groups.

7. Results

Using Algorithm 6.1, in a short time one can produce millions of examples of abelian surfaces over $\mathbb{Q}$ such that the order of the Tate-Shafarevich group is either a square or five times a square. In the cases arising from two elliptic curves each of analytic rank at most 1, the examples are completely unconditional. We constructed two databases of elliptic curves using step 1 of the algorithm. The first database consists of all elliptic curves E_d, where $d = u/v$ for positive integers u and v with $\max(u, v) \leq 50{,}000$, and where the conductor of E_d is bounded by $C = 10^6$. The second database consists of all elliptic curves E_d, where $d = u/v$ for positive integers u and v such that $\max(u, v) \leq 100$.

Database 1 contains 2212 elliptic curves, all of them having analytic rank $r \leq 2$. It is likely that there are no further elliptic curves of conductor at most 10^6 that

		Number of E_d of rank r					Number of E_d of rank r		
N	$\#E_d$	$r=0$	$r=1$	$r=2$	N	$\#E_d$	$r=0$	$r=1$	$r=2$
50,000	2,212	987	1,109	116	800	2,159	956	1,088	115
4,617	2,212	987	1,109	116	700	2,145	951	1,079	115
3,375	2,211	986	1,109	116	600	2,119	941	1,063	115
3,072	2,210	986	1,108	116	500	2,088	921	1,052	115
2,695	2,209	986	1,107	116	400	2,066	912	1,039	115
2,000	2,200	982	1,102	116	300	1,993	872	1,009	112
1,000	2,174	963	1,095	116	200	1,818	786	929	103
900	2,170	961	1,093	116	100	1,391	616	697	78
					50	845	394	405	46

Table 1. Summary of database 1. For each N, we give the number of curves E_d of conductor at most 10^6, where $d>0$ has height at most N. The final three columns give the number of such curves of analytic rank 0, 1, and 2.

have a rational torsion point of order 5, since there is no such curve with $4617 < \max(u, v) \leq 50{,}000$. The database is described in more detail in Table 1, where we state for each analytic rank the number of elliptic curves with conductor at most 10^6 and with $\max(u, v) \leq N$. Database 2 contains 6,087 elliptic curves. All of them have analytic rank $r \leq 3$. See Table 2 for more details. In the following we will present the results of step 2 of the algorithm applied to the two databases described above.

Database 1 yields 2,445,366 abelian surfaces B_{d_1,d_2}. It turns out that 47.01% of these surfaces have Tate-Shafarevich groups of nonsquare order. Database 2 leads to 18,522,741 abelian surfaces. The percentage of the nonsquare case is 49.31. The intersection of the two databases consists of 1,391 curves, hence we considered 966,745 surfaces twice. In total this gives 20,001,362 surfaces, of which 49.16% have a Tate-Shafarevich group of nonsquare order.

		Number of E_d of rank r					Number of E_d of rank r			
N	$\#E_d$	$r=0$	$r=1$	$r=2$	$r=3$	N	$\#E_d$	$r=0$	$r=1$	$r=2$	$r=3$
100	6,087	2,390	3,038	633	26	50	1,547	660	760	123	4
90	4,959	1,987	2,463	490	19	40	979	412	494	70	3
80	3,931	1,597	1,940	380	14	30	555	245	277	33	0
70	2,987	1,235	1,455	287	10	20	255	130	115	10	0
60	2,203	925	1,074	198	6	10	63	40	22	1	0

Table 2. Summary of database 2. For each N, we give the number of curves E_d, where $d>0$ has height at most N. The final four columns give the number of such curves of analytic rank 0, 1, 2, and 3.

rk E_1	rk E_2	$\#B$	$\%(Ш = \square)$	$\%(\mathrm{RE} \equiv \mathrm{rk}\, B)$
0	0	486,591	54.041	100.00
1	1	614,386	58.614	63.51
2	2	6,670	92.039	55.53
0	1	1,094,583	46.634	83.44
0	2	114,492	52.867	47.96
1	2	128,644	74.314	42.48
≤ 1	≤ 1	2,195,560	51.628	81.53

Table 3. Results of experiment 1 for database 1, the curves E_d of conductor at most 10^6 and with $d > 0$ of height at most 50,000. For each pair of ranks, we list the number of surfaces B obtained from elliptic curves in database 1 with those ranks. The fourth column gives the percentage of these surfaces for which Ш has square order, and the fifth column gives the percentage for which the exponent of the regulator quotient is congruent modulo 2 to the rank of the surface.

We did two different experiments with the two databases. In experiment 1 we investigated how the rank influences the squareness of the Tate-Shafarevich group. We list the result in Table 3 for database 1 and in Table 4 for database 2. The first three, respectively four, entries correspond to pairs (E_1, E_2) with the same analytic rank. The following three, respectively six, lines correspond to pairs with different

rk E_1	rk E_2	$\#B$	$\%(Ш = \square)$	$\%(\mathrm{RE} \equiv \mathrm{rk}\, B)$
0	0	2,854,855	48.598	100.00
1	1	4,613,203	48.882	80.91
2	2	200,028	73.031	44.03
3	3	325	98.154	51.08
0	1	7,260,820	51.366	91.02
0	2	1,512,870	50.567	71.36
0	3	62,140	49.891	52.73
1	2	1,923,054	52.717	59.50
1	3	78,988	60.632	46.23
2	3	16,458	84.470	48.23
≤ 1	≤ 1	14,728,878	50.051	89.59

Table 4. Results of experiment 1 for database 2, the curves E_d with $d > 0$ of height at most 100. For each pair of ranks, we list the number of surfaces B obtained from elliptic curves in database 2 with those ranks. The fourth column gives the percentage of these surfaces for which Ш has square order, and the fifth column gives the percentage for which the exponent of the regulator quotient is congruent modulo 2 to the rank of the surface.

C	$\#E$	$\#B$	$\%(Ш = \square)$	$\%(\text{RE} \equiv \text{rk } B)$
1,000,000	2,212	2,445,366	52.990	77.84
800,000	1,966	1,931,595	53.232	77.16
600,000	1,683	1,415,403	53.758	76.06
400,000	1,351	911,925	54.215	75.24
200,000	924	426,426	55.001	73.91
100,000	623	193,753	57.074	74.29
80,000	547	149,331	57.776	74.03
60,000	470	110,215	57.990	72.75
40,000	376	70,500	59.306	73.34
20,000	245	29,890	61.288	71.72
10,000	152	11,476	62.182	72.59
5,000	110	5,995	59.783	71.79
1,000	45	990	65.556	76.77

Table 5. Results of experiment 2 for database 1. For each value of C, we list the number of elliptic curves E_d having conductor at most C and with $d > 0$ of height at most 50,000. In the third column we list the number of abelian surfaces B obtained from pairs of such curves. The fourth column gives the percentage of these surfaces for which Ш has square order, and the fifth column gives the percentage for which the exponent of the regulator quotient is congruent modulo 2 to the rank of the surface.

analytic ranks, and the final line corresponds to pairs with analytic rank $r \leq 1$. If we consider abelian surfaces of fixed analytic rank of at least 4 then the density of the surfaces with square Tate-Shafarevich group seems to be significant larger than 0.5. However the surfaces with rank larger than 2 inside our family are conjectured to have density zero and our database contains very few such cases. The calculations with curves of rank $r \leq 1$ all show that the nonsquare case happens in about 50% of all cases. For both experiments we list how many abelian surfaces B_{d_1,d_2} occur in each of the cases, we state the percentage of the surfaces with square Tate-Shafarevich group, and we give the percentage of in how many cases the parity of the rank of the abelian surface agrees with the parity of the exponent of the regulator quotient (RE). Note that the results are unconditional in case $\text{rk}(E_i) \leq 1$, for both E_i. If one of the analytic ranks is at least 2 then we need to make some assumptions; see the remark at the send of Section 6.

In experiment 2 we looked for the behavior of the distribution of square and nonsquare Tate-Shafarevich group orders for increasing conductor (for database 1) and height (for database 2) of the elliptic curves. For low bounds on the conductor and height, the nonsquare case was less likely. When we increase these bounds the frequency of nonsquares tends to approximately 50%. The results of experiment 2 is given in Table 5 for database 1 and Table 6 for database 2. Note that for

N	$\#E$	$\#B$	$\%(Ш = \square)$	$\%(\mathrm{RE} \equiv \mathrm{rk}\, B)$
100	6,087	18,522,741	50.694	84.14
90	4,959	12,293,361	50.821	83.66
80	3,931	7,724,415	50.941	83.32
70	2,987	4,459,591	51.235	82.51
60	2,203	2,425,503	51.461	82.00
50	1,547	1,195,831	52.211	80.85
40	979	478,731	52.764	79.92
30	555	153,735	54.157	77.12
20	255	32,385	56.384	77.11
10	63	1,953	67.179	74.04

Table 6. Results of experiment 2 for database 2. For each value of N, we list the number of curves E_d with $d > 0$ of height at most N, as well as the number of abelian surfaces B obtained from pairs of such curves. The fourth column gives the percentage of these surfaces for which Ш has square order, and the fifth column gives the percentage for which the exponent of the regulator quotient is congruent modulo 2 to the rank of the surface.

some of the surfaces we assume the weak form of the Birch and Swinnerton-Dyer conjecture mentioned above.

The two ways we ordered the elliptic curves, via conductor and via height, are natural orderings. It is conjectured that the densities obtained with respect to these orderings agree. In both cases the densities seem to exist and are around 0.5. This is in contrast to the results of Poonen and Stoll [11], who showed that the density of nonsquare #Ш for Jacobians of genus-2 curves is about 0.13, while for higher-genus curves the density tends to zero as the genus increases.

We end by giving some heuristics why we expect the density to be 50%. We expect that for a random pair $(d_1 = u_1/v_1, d_2 = u_2/v_2)$ in $\mathbb{Q}^* \times \mathbb{Q}^*$ the global factor is a square for 50% of the abelian surfaces and that the local factor is a square for 50% of them, too. We also expect these distributions to be independent. Using the 18,522,741 pairs obtained from the second database, we get numerical evidence for the independence, as illustrated in Table 7.

Global quotient	Local quotient	Percentage
square	square	26.08
square	nonsquare	24.04
nonsquare	square	25.26
nonsquare	nonsquare	24.61

Table 7. Fraction of surfaces coming from database 2 with square and nonsquare local and global quotients.

$\#(U_1 \cap U_2)$	$\#(T_1 \cup T_2)$	Percentage
even	even	46.71
even	odd	49.55
odd	even	1.80
odd	odd	1.95

Table 8. Fraction of surfaces coming from database 2 with even and odd values of $\#(U_1 \cap U_2)$ and $\#(T_1 \cup T_2)$.

Recall that the exponent of the local quotient equals $\#(U_1 \cap U_2) - \#(T_1 \cup T_2)$, hence one could prove the expected densities for the local quotient by showing that the probability that the set $(T_1 \cup T_2)$ has an even number of elements is independent of the probability that the set $(U_1 \cap U_2)$ has an even number of elements. The corresponding numerical result for database 2 is gathered in Table 8.

The global quotient is harder to control. The exponent of the torsion quotient equals 3 on a density-1 subset of the pairs (d_1, d_2); see [6, Proposition 4.6]. The results of Tables 3–6 suggest that the squareness of the regular quotient, and hence the squareness of the global quotient, is not independent of the parity of the rank. If the ranks of both of the elliptic curves E_1 and E_2 are equal to 0, hence are even, the regulator quotient equals 1, hence is a square. If one elliptic curve is of rank 0 and the other is of rank 1, then the regulator quotient is a nonsquare if and only if coker $\eta_{\mathbb{Q}}$ can be generated by torsion points, where η is the usual isogeny belonging to the elliptic curve of rank 1. In database 2 we have the following situation. For the rank-1 curves it happens in about 91.2% of the cases that $\eta_{\mathbb{Q}}$ is surjective on the free part. In case both ranks are equal to 1, the regulator quotient is a square in about 80.9% of the cases. For the complete second database we get that the parity of the exponent of the regulator quotient agrees with the parity of the rank in 84.14% of the cases. If we consider only all the elliptic curves of rank ≤ 1, then we have that for abelian surfaces B_{d_1,d_2} of even rank the regulator quotient is a square in about 88.2% of the cases, and for abelian surfaces B_{d_1,d_2} of odd rank the regulator quotient is a nonsquare in about 91.0% of the cases; together, this means there is agreement 89.6% of the time. Table 9 gives the situation for the complete database 2.

Regulator quotient	$\mathrm{rk}(B_{d_1,d_2})$	Percentage
square	even	42.067
square	odd	7.931
nonsquare	even	7.927
nonsquare	odd	42.075

Table 9. Fraction of surfaces B_{d_1,d_2} coming from database 2 with square and nonsquare regulator quotient and even and odd rank.

Local quotient	$\mathrm{rk}(B_{d_1,d_2})$	Percentage
square	even	25.670
square	odd	25.675
nonsquare	even	24.324
nonsquare	odd	24.331

Table 10. Fraction of surfaces B_{d_1,d_2} coming from database 2 with square and nonsquare local quotient and even and odd rank.

In contrast to the global quotient, the squareness of the local quotient seems to be independent of the parity of the rank of the abelian surfaces. Table 10 gives the numerical results for database 2.

Acknowledgments

Keil is supported by a scholarship from the Berlin Mathematical School (BMS). Both authors thank Tom Fisher for pointing out the method `MordellWeilSha-Information` in Magma, and the referees for their comments and suggestions.

References

[1] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system. I. The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[2] J. W. S. Cassels, *Arithmetic on curves of genus* 1. *VIII. On conjectures of Birch and Swinnerton-Dyer*, J. Reine Angew. Math. **217** (1965), 180–199. MR 31 #3420

[3] John Cremona, Joan-Carles Lario, Jordi Quer, and Kenneth Ribet (eds.), *Modular curves and abelian varieties*: *Papers from the conference held in Bellaterra*, *July* 15–18, 2002, Progress in Mathematics, no. 224, Birkhäuser, Basel, 2004. MR 2004k:11004

[4] Matthias Flach, *A generalisation of the Cassels-Tate pairing*, J. Reine Angew. Math. **412** (1990), 113–127. MR 92b:11037

[5] E. V. Flynn and C. Grattoni, *Descent via isogeny on elliptic curves with large rational torsion subgroups*, J. Symbolic Comput. **43** (2008), no. 4, 293–303. MR 2009c:11080

[6] Stefan Keil, *Examples of abelian varieties with non-square Tate-Shafarevich group*, 2012. arXiv 1206.1822v1 [math.NT]

[7] ———, *Sage worksheet*: *On the density of abelian surfaces with Tate-Shafarevich group of order five times a square*, 2012. http://www.sagenb.org/home/pub/4330/

[8] Remke Kloosterman, *The p-part of the Tate-Shafarevich groups of elliptic curves can be arbitrarily large*, J. Théor. Nombres Bordeaux **17** (2005), no. 3, 787–800. MR 2006k:11102

[9] Remke Kloosterman and Edward F. Schaefer, *Selmer groups of elliptic curves that can be arbitrarily large*, J. Number Theory **99** (2003), no. 1, 148–163. MR 2003m:11081

[10] Kazuo Matsuno, *Construction of elliptic curves with large Iwasawa λ-invariants and large Tate-Shafarevich groups*, Manuscripta Math. **122** (2007), no. 3, 289–304. MR 2008h:11106

[11] Bjorn Poonen and Michael Stoll, *The Cassels-Tate pairing on polarized abelian varieties*, Ann. of Math. (2) **150** (1999), no. 3, 1109–1149. MR 2000m:11048

[12] Edward F. Schaefer, *Class groups and Selmer groups*, J. Number Theory **56** (1996), no. 1, 79–114. MR 97e:11068

[13] Edward F. Schaefer and Michael Stoll, *How to do a p-descent on an elliptic curve*, Trans. Amer. Math. Soc. **356** (2004), no. 3, 1209–1231. MR 2004g:11045

[14] Jean-Pierre Serre, *Galois cohomology*, Springer, Berlin, 2002. MR 2002i:12004

[15] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 106, Springer, New York, 1986. MR 87g:11070

[16] W. A. Stein et al., *Sage Mathematics Software* (*version* 4.6.2), The Sage Development Team, 2011. http://www.sagemath.org

[17] William A. Stein, *Shafarevich-Tate groups of nonsquare order*, in Cremona et al. [3], 2004, pp. 277–289. MR 2005c:11072

[18] John Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analog*, Séminaire Bourbaki (reprint), vol. 9, Soc. Math. France, Paris, 1995, pp. 415–440, Exp. No. 306. MR 1610977

[19] Jacques Vélu, *Isogénies entre courbes elliptiques*, C. R. Acad. Sci. Paris Sér. A-B **273** (1971), A238–A241. http://gallica.bnf.fr/ark:/12148/bpt6k56191248/f52.image MR 45 #3414

STEFAN KEIL: keil@math.hu-berlin.de
Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

REMKE KLOOSTERMAN: klooster@math.hu-berlin.de
Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.437

msp

Improved CRT algorithm for class polynomials in genus 2

Kristin E. Lauter and Damien Robert

We present a generalization to genus 2 of the probabilistic algorithm of Sutherland for computing Hilbert class polynomials. The improvement over the Bröker-Gruenewald-Lauter algorithm for the genus 2 case is that we do not need to find a curve in the isogeny class whose endomorphism ring is the maximal order; rather, we present a probabilistic algorithm for "going up" to a maximal curve (a curve with maximal endomorphism ring), once we find any curve in the right isogeny class. Then we use the structure of the Shimura class group and the computation of (ℓ, ℓ)-isogenies to compute all isogenous maximal curves from an initial one.

1. Introduction

Cryptographic solutions to provide privacy and security for sensitive transactions depend on using a mathematical group in which the discrete logarithm problem is hard. For example, digital signature schemes or a Diffie-Hellman key exchange may be based on the difficulty of solving the discrete logarithm problem in the group of points on the Jacobian of a genus-2 curve. For this problem to be hard we must ensure that we can choose genus-2 curves over finite fields whose Jacobian have an almost-prime number of points.

One approach to this problem is to construct curves whose Jacobians have a given order using the method of complex multiplication (CM). The CM method works by computing invariants of the curve and then reconstructing the curve using the Mestre-Cardona-Quer [31; 8] algorithm. Invariants are computed by constructing their minimal polynomials, called *Igusa class polynomials*. Computing the invariants is computationally intensive, and there are three known methods for constructing Igusa class polynomials:

MSC2010: primary 14K22; secondary 11Y40, 11Y16, 11G15, 14K02.
Keywords: class field polynomials, CRT, hyperelliptic curve cryptography, isogenies.

(1) the complex analytic method [37; 41; 42; 38];

(2) the Chinese remainder theorem method (CRT) [16; 18; 6]; and

(3) the p-adic lifting method [20; 9; 10].

Currently, the CRT method in genus 2 remains by far the slowest of these three methods, as measured on the small examples that have been computed to date; but the history of the evolution of these three methods in genus 1 gives some hope that the CRT method may be asymptotically competitive with the others. In genus 1, the (explicit) CRT method now holds the record for the best proven bounds on time and space complexity (under GRH), as well as for the size of the largest examples that have been computed [39; 17]. In this paper, we propose numerous improvements to the CRT method for computing genus-2 curves, paralleling improvements made by Sutherland [39] to the CRT method in genus 1.

The CRT method works by computing class polynomials modulo many small primes, and then reconstructing the polynomials with rational coefficients (or modulo a much larger prime number) via the Chinese remainder theorem (respectively, the explicit CRT). The CRT method for computing class polynomials in genus 2 was proposed by Eisenträger and Lauter [16]; they gave sufficient conditions on the CRT primes to ensure correctness and included an algorithm for computing endomorphism rings for ordinary Jacobians of genus-2 curves, generalizing Kohel's algorithm for genus-1 curves. For each small CRT prime p, the algorithm loops through all p^3 possible triples of Igusa invariants of curves, reconstructing the curve and testing for each curve whether it is in the desired isogeny class and whether its endomorphism ring is maximal. The algorithm for computing endomorphism rings from [16] was replaced by a much more efficient probabilistic algorithm in [18], where a number of examples were given for running times of the computations modulo small CRT primes. Bröker, Gruenewald, and Lauter [6] introduced the idea of using computable $(3, 3)$-isogenies to find other curves in the isogeny class once an initial curve was found, but still searched until finding a curve whose Jacobian has endomorphism ring equal to a maximal order (a *maximal* curve). Another improvement described in [6] was a method to construct other maximal curves using $(3, 3)$-isogenies once an initial maximal curve is found.

In this paper we present a generalization to genus 2 of the probabilistic Algorithm 1 in Sutherland [39]. The improvement over the genus-2 algorithm presented in [6] is that we do not need to find a maximal curve in the isogeny class; instead, we present a probabilistic algorithm for "going up" to a maximal curve once we find *any* curve in the right isogeny class. Then we use the structure of the Shimura class group and the computation of (ℓ, ℓ)-isogenies to compute all isogenous maximal curves from an initial one. Although we cannot prove that the going-up algorithm succeeds with any fixed probability, it works well in practice, and heuristically

it improves the running time of the genus-2 CRT method from p^3 per prime p to $p^{3/2}$ per prime p.

Let K denote a primitive quartic CM field, with real quadratic subfield K^+ and ring of integers $\mathcal{O}_K$. Let Φ denote a CM-type of K and let K_Φ denote the reflex CM field. Let TN_Φ denote the type norm associated to the CM-type Φ. Informally, the algorithm is as follows; the individual steps will be explained in subsequent sections.

Algorithm 1.

Input: A primitive quartic CM field K with a CM-type Φ, and a collection of CRT primes P_K for K.

Output: Igusa class polynomials $H_i(x)$, $i = 1, 2, 3$, either in $\mathbb{Q}[x]$ or reduced modulo a prime q.

1. Loop through CRT primes $p \in P_K$:
   (a) Enumerate hyperelliptic curves C of genus 2 over $\mathbb{F}_p$ until a curve in the right $\mathbb{F}_p$-isogeny class (up to a quadratic twist) is found.
   (b) Try to go up to a maximal curve from C; if this step fails, go back to Step 1(a).
   (c) From a maximal curve C, compute all other maximal curves.
   (d) Reconstruct the class polynomials $H_i(x)$ modulo p from the Igusa invariants of the set of maximal curves.
2. Recover $H_i(x)$, $i = 1, 2, 3$, in $\mathbb{Q}[x]$ or modulo q using the (explicit) CRT method once we have computed $H_i(x)$ modulo p for enough primes p.

For the dihedral case, one new aspect of our algorithm is that we extend to the CRT setting the idea of computing the class polynomials associated to only one fixed CM-type Φ for K [38, §III.3]. When K is cyclic, this makes no difference, since all isomorphism classes of abelian surfaces with CM by K arise from one CM-type; but when K is dihedral, two CM-types are needed to find all isomorphism classes of CM abelian surfaces. All three previous versions of the CRT algorithm [16; 18; 6] compute the class polynomials classifying all abelian surfaces with CM by $\mathcal{O}_K$ (with either of the two possible CM-types in the dihedral case). The advantage of our approach is that it computes only a factor of half the degree of the whole class polynomial. The drawback of this approach is that in the dihedral case, each factor of the class polynomials is defined over $\mathcal{O}_{K_\Phi^+}$ rather than over $\mathbb{Z}$. So once we compute the class polynomials modulo $\mathfrak{p}$ as polynomials in $\mathcal{O}_{K_\Phi^+}/\mathfrak{p}$, the CRT step must be performed in $\mathcal{O}_{K_\Phi^+}$.

A *CRT prime* $\mathfrak{p} \subset \mathcal{O}_{K_\Phi^+}$ is a prime such that all abelian surfaces over $\mathbb{C}$ with CM by $(\mathcal{O}_K, \Phi)$ have good reduction modulo $\mathfrak{p}$. By [36, §III.13], $\mathfrak{p}$ is a CRT prime for the CM-type Φ if and only if there exists an unramified prime $\mathfrak{q}$ in $\mathcal{O}_{K_\Phi}$ of degree 1

above $\mathfrak{p}$ of principal type norm (π) with $\pi\overline{\pi} = N_{K/\mathbb{Q}}(\mathfrak{q})$; in particular, this implies that $\mathfrak{q}$ is totally split in the class field corresponding to the abelian surfaces with CM by $(\mathbb{O}_K, \Phi)$. By [21, §3], these surfaces have good reduction modulo $\mathfrak{p}$, and by a theorem of Tate the isogeny class of the reductions modulo $\mathfrak{p}$ is determined by the characteristic polynomial of $\pm\pi$, at least in the case where $\mathbb{O}_K^* = \{\pm 1\}$. For reasons of efficiency, we will work with CRT primes $\mathfrak{p}$ that are unramified of degree one over $p = \mathfrak{p} \cap \mathbb{Z}$. By [21], the reduction to $\mathbb{F}_p$ of the abelian surfaces with CM by $(\mathbb{O}_K, \Phi)$ will then be ordinary. We then make the slight abuse of notation of calling p a CRT prime when there is a CRT prime $\mathfrak{p}$ above it. Note another advantage of restricting to one CM-type: To use p for both CM-types, p needs to split completely into $p = \mathfrak{p}_1\mathfrak{p}_2$ such that both $\mathfrak{p}_1$ and $\mathfrak{p}_2$ are CRT primes, and there are fewer p which satisfy this stronger requirement.

In addition to the two main contributions of the paper — the going-up algorithm to find maximal curves, and an improvement to the algorithm to compute maximal curves from maximal curves — we also give improvements to every step of the CRT algorithm. Here we give a brief outline of the paper and a summary of those improvements.

Step 1(b) of the algorithm (the "going-up" part) is explained in Section 3. We first explain in Section 2 how to compute if a curve is maximal, since this is used in the going-up algorithm. We present some significant improvements over the algorithm from [18]. Step 1(c) (finding all other maximal curves from one maximal curve) is explained in Section 4.

As for Step 1(d), once all maximal abelian surfaces with CM by K are found for a given prime p, it is easy to compute the associated class polynomials modulo p. The class polynomials depend on the choice of Igusa invariants, and we use the invariants recommended in [38, Appendix 3] which give smaller coefficients than those used in [41; 42; 21]. For the dihedral case the class polynomials must be reconstructed over $\mathbb{O}_{K_\Phi^+}$, and we give more details about this step in Section 5.

Section 6 gives a complexity analysis, and explains how each improvement affects the final complexity. The final complexity bound, while still not quasilinear, is a significant improvement compared to [6]. Finally, examples demonstrating significantly improved running times are given in Section 7.

The interested reader will find an extended version of this paper in [28].

2. Checking whether the endomorphism ring is maximal

We recall the algorithm described in [16] for checking whether the endomorphism ring of an abelian surface is maximal, and we describe some improvements. The ideas for computing the endomorphism ring will be used in the going-up phase of Algorithm 1.

2.1. ***The algorithm of Eisenträger, Freeman, and Lauter.*** Let $A/\mathbb{F}_p$ be an ordinary abelian surface with CM by K, let $\mathbb{O} = \mathrm{End}\, A$, and let $\pi \in \mathbb{O}$ be the Frobenius endomorphism. We know that $\mathbb{Z}[\pi] \subseteq \mathbb{Z}[\pi, \bar{\pi}] \subseteq \mathbb{O} \subseteq \mathbb{O}_K$, and our goal is to check whether $\mathbb{O} = \mathbb{O}_K$. First, the Chinese remainder theorem gives us the following proposition:

Proposition 2. *Let* $\{1, \alpha_1, \alpha_2, \alpha_3\}$ *be a basis of* $\mathbb{O}_K$ *as a* $\mathbb{Z}$*-module, and write* $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]] = \prod \ell_i^{e_i}$. *If* $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]] \cdot \alpha_j / \ell_i^{e_i} \in \mathbb{O}$ *for* $j = 1, 2, 3$, *and all* ℓ_i *dividing the index, then* $\mathbb{O} = \mathbb{O}_K$. □

We are then reduced to the following problem: For $\gamma \in \mathbb{O}_K$ such that $\ell^e \gamma \in \mathbb{Z}[\pi, \bar{\pi}]$, check if $\gamma \in \mathbb{O}$.

Proposition 3. *Let* $\mathbb{O} = \mathrm{End}\, A$ *and let* $\gamma \in \mathbb{O}_K$ *be such that* $\ell^e \gamma \in \mathbb{Z}[\pi, \bar{\pi}]$. *There exists a unique integer polynomial* P_γ *of degree less than* 4 *such that* $\ell^e p\gamma = P_\gamma(\pi)$, *and* γ *is in* $\mathbb{O}$ *if and only if* $P_\gamma(\pi) = 0$ *on* $A[\ell^e]$.

Proof. First note that $[\mathbb{Z}[\pi, \bar{\pi}] : \mathbb{Z}[\pi]] = p$ (see [18, p. 38]), so that $\ell^e p\gamma \in \mathbb{Z}[\pi]$, which means we can write $\ell^e p\gamma = P_\gamma(\pi)$ for a unique $P_\gamma \in \mathbb{Z}[x]$ of degree less than 4. Second, since we are dealing with ordinary abelian surfaces over $\mathbb{F}_p$, we have $p \nmid [\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$ by [18, Proposition 3.7], so that $\gamma \in \mathbb{O} \iff p\gamma \in \mathbb{O}$. Lastly, by the universal property of isogenies, we have that $P_\gamma(\pi) = 0$ on $A[\ell^e]$ if and only if $p\gamma \in \mathbb{O}$ (see [16]). Summing up, we only need to check that $P_\gamma(\pi) = 0$ on $A[\ell^e]$ to check that $\gamma \in \mathbb{O}$. □

Remark 4. Since most of the curves in the isogeny class are not maximal, it is more efficient to check the condition $P_\gamma(\pi) = 0$ on $A[\ell]$, $A[\ell^2]$, …, rather than directly on $A[\ell^e]$.

2.2. ***Computing the ℓ^e-torsion.*** The obvious method of using Proposition 3 to test whether an element of $\mathbb{O}_K$ lies in $\mathbb{O}$ involves computing a basis of the ℓ^e-torsion group. The cost of such a computation depends on the degree of the extension where the ℓ^e-torsion points are defined. We have:

Lemma 5. *Let* d *be the degree such that the* ℓ*-torsion points of* A *are defined over* $\mathbb{F}_{p^d}$. *Then* $d \leq \ell^4 - 1$. *Furthermore, the* ℓ^e*-torsion is all defined over* $\mathbb{F}_q$ *with* $q = p^{d\ell^{e-1}}$.

Proof. Let χ_π be the characteristic polynomial of π. Then d is the (multiplicative) order of X in the ring $\mathbb{F}_\ell[X]/\chi_\pi(X)$, so $d \leq \ell^4 - 1$. The second assertion follows from [18, §6]. □

Remark 6. For *maximal* abelian surfaces, [18, Proposition 6.2] gives a better bound for d: In that case we have $d < \ell^3$, and if ℓ is completely split in $\mathbb{O}_K$ we have $d \mid \ell - 1$.

We will use the following algorithm to compute points uniformly in an ℓ-primary group containing $A[\ell^e]$:

Algorithm 7.

Input: An abelian surface $A/\mathbb{F}_p$ and a prime power ℓ^e.

Output: Uniform random points in the group $A(\mathbb{F}_{p^{de}})[\ell^\infty]$, defined below.

1. Precomputation:
   (a) Let d be the (multiplicative) order of X in the ring $\mathbb{F}_\ell[X]/\chi_\pi(X)$ and set $d_e = d\,\ell^{e-1}$.
   (b) Compute $\chi_{\pi^{de}}$ as the resultant in X of $\chi_\pi(Y)$ and $Y^{d_e} - X$, and write $\#A(\mathbb{F}_p^{d_e}) = \chi_{\pi^{de}}(1) = \ell^e\gamma$ with γ prime to ℓ.
2. Repeat as needed:
   (a) Take a random point P (uniformly) in $A(\mathbb{F}_{p^{de}})$.
   (b) Return γP.

Algorithm 4.3 of [18] computes random points in $A(\mathbb{F}_{p^{de}})[\ell^e]$ by taking uniform random points P in $A(\mathbb{F}_{p^{de}})[\ell^\infty]$ and looking at the smallest k such that $\ell^k \cdot P$ is an ℓ^e-torsion point; it generates enough such random points so that the probability that they generate the full ℓ^e-torsion is sufficiently high, and then tests P_γ on these points of ℓ^e-torsion. The algorithm computes how many points are needed so that the probability of generating the full ℓ^e-torsion is greater than $1-\epsilon$ for some $\epsilon > 0$, so the result is not guaranteed (that is, it is a "Monte Carlo" algorithm). This is very inconvenient in our setting since we need to test a lot of curves across different CRT primes p.

To ensure correctness we can check that the subgroup generated by the points obtained is of cardinality ℓ^{4e}, but this is costly. A more efficient way is as follows: $\{P_1, \ldots, P_4\}$ is a basis of the ℓ^e-torsion if and only if $\{\ell^{e-1}P_1, \ldots, \ell^{e-1}P_4\}$ is a basis of the ℓ-torsion. But that can be easily checked by computing the 6 Weil pairings $e_\ell(\ell^{e-1}P_i, \ell^{e-1}P_j)$ for $i < j$ and testing whether the corresponding 4×4 matrix is invertible. Since Weil pairings can be computed in time $O(\log(\ell))$, this is much faster. This is our first improvement, yielding a "Las Vegas" algorithm.

The second drawback of the approach of [18] is that, although the random points in $A(\mathbb{F}_{p^{de}})[\ell^\infty]$ are uniform, this is not always the case for the random points in $A(\mathbb{F}_{p^{de}})[\ell^e]$. To have a high probability of generating the full ℓ-torsion then requires taking many random points in $A(\mathbb{F}_{p^{de}})[\ell^\infty]$: If $A(\mathbb{F}_{p^{de}})[\ell^\infty] = \ell^s$, the algorithm requires $\ell^{s-4e}(-\log(\epsilon))^{1/2}$ random points to succeed with probability greater than $1-\epsilon$. Since generating these points is the most costly part of the algorithm it is best to minimize the number of random points required. Our second improvement is to use an algorithm, due to Couveignes [14] and implemented in the Magma package AVIsogenies [4], to get uniform random points in $A(\mathbb{F}_{p^{de}})[\ell^e]$.

Since the full algorithm is described in more detail in [4], we only give an example to illustrate it here.

Suppose that G is an ℓ-primary group generated by a point P of order ℓ^2 and a point Q of order ℓ. Assume that the first random point chosen is $P = R_1$, which gives an ℓ-torsion point $T_1 = \ell P$. The second random point R_2 chosen will be of the form $\alpha P + \beta Q$. In most cases, $\alpha \neq 0$, so the corresponding new ℓ-torsion point is $T_2 = \alpha \ell P$, a multiple of T_1. However we can correct R_2 by the corresponding multiple: Compute $R_2' = R_2 - \alpha R_1 = \beta Q$. Thus R_2' gives the rest of the ℓ-torsion unless $\beta = 0$. In our setting we can use the Weil pairing to express a new ℓ-torsion point in terms of the generating set already constructed (except when we have an isotropic group, in this case we have to compute the ℓ^2 multiples), and we only need $O(1)$ random points to find a basis. The cost of finding a basis of the ℓ^e-torsion is then $O(d_e \log p + \ell^2)$ operations in $\mathbb{F}_{p^{d_e}}$.

2.3. ***Reducing the degree.*** The complexity of finding the basis is closely related to the degree of the extension d_e. Let d_0 be the minimal integer such that $(\pi^{d_0} - 1) \in \ell \mathbb{O}_K$. Then $d_0 \mid d$, and, as remarked in [18], since we only need to check if $\mathbb{O} = \mathbb{O}_K$, we can first check that $(\pi^{d_0} - 1)/\ell$ lies in $\mathbb{O}$. In other words, we can check that the ℓ-torsion points of A are defined over $\mathbb{F}_{p^{d_0}}$ rather than over $\mathbb{F}_{p^d}$. If this is the case, the ℓ^e-torsion points are then defined over an extension of degree $d_0 \ell^{e-1}$ of $\mathbb{F}_p$, which allows us to work with smaller extensions.

Another improvement we implemented to reduce the degree is to use twists. Let d_0' be the minimal integer such that $((-\pi)^{d_0'} - 1) \in \ell \mathbb{O}_K$. Then there are three possibilities: We have either $d_0' = d_0$, or $d_0' = 2d_0$, or $d_0 = 2d_0'$. In the third case it is to our advantage to replace A by its twist, because the Frobenius of the twist is represented by $-\pi$, and we can therefore compute the points of ℓ^e-torsion by working over extensions of half the degree.

Example 8. Let H be the curve $y^2 = 80x^6 + 51x^5 + 49x^4 + 3x^3 + 34x^2 + 40x + 12$ of genus 2 over $\mathbb{F}_{139}$, and let J be the Jacobian of H. By computing the characteristic polynomial of Frobenius for J we find that

$$(\operatorname{End} J) \otimes \mathbb{Q} \cong \mathbb{Q}\big(i\sqrt{13 + 2\sqrt{29}}\big),$$

and we would like to check whether $\operatorname{End} J$ is maximal. In this example, we compute that $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]] = 3^5$, so we need to compute the points in $J[3^5]$, which live over an extension of degree 81. If we had checked the endomorphism ring of the Jacobian of the twist of H, we would have needed to work over an extension of degree 162.

2.4. ***Reducing the number of endomorphisms to test.*** One last improvement to the algorithm of [18] is to use the fact that $\operatorname{End} A$ is an order; if we know that

$\gamma \in \mathbb{O}$, then we know that the whole ring $\mathbb{Z}[\pi, \bar{\pi}, \gamma]$ is contained in $\mathbb{O}$. For example, suppose $\{1, \alpha_1, \alpha_2, \alpha_3\}$ is a basis for $\mathbb{O}_K$ and $\alpha_3 = \alpha_1\alpha_2 \mod \mathbb{Z}[\pi, \bar{\pi}]^*$. To check that $\mathbb{O} = \mathbb{O}_K$ we only have to check that α_1 and α_2 are in $\mathbb{O}$. In fact, since our algorithm works locally at primes ℓ, we only need the relation between α_3 and $\alpha_1\alpha_2$ to hold locally at ℓ.

We use this idea as follows: Suppose that we have checked that $\{\gamma_1, \ldots, \gamma_k\}$ are endomorphisms lying in $\mathbb{O}$, and we want to check if $\gamma \in \mathbb{O}$. Let N_1 be the order of γ in the $\mathbb{Z}$-module $\mathbb{O}_K/\mathbb{Z}[\pi, \bar{\pi}, \gamma_1, \ldots, \gamma_k]$, and N_2 be the order of γ in $\mathbb{O}_K/\mathbb{Z}[\pi, \bar{\pi}]$. If we write $N_2 = \prod \ell_i^{e_i}$, we only have to check that $(N_2/\ell_i^{e_i})\gamma \in \mathbb{O}$ for $\ell_i \,|\, N_1$. In fact, if the valuation of N_1 at ℓ_i is f_i, then we would only need to check that $(N_1/\ell_i^{f_i})\gamma \in \mathbb{O}$, which means testing if $N_1\gamma = 0$ on the $\ell_i^{f_i}$-torsion, where $N_1\gamma$ is a polynomial in π, $\bar{\pi}$, and the γ_i $(i = 1, \ldots, k)$. We write this polynomial as $N_1/(pN_2)$ times a polynomial in π, so that we still need to compute the $\ell_i^{e_i}$-torsion.

Example 9. Let H be the curve $y^2 = 10x^6 + 57x^5 + 18x^4 + 11x^3 + 38x^2 + 12x + 31$ of genus 2 over $\mathbb{F}_{59}$ and let J the Jacobian of H. We have

$$(\mathrm{End}\, J) \otimes \mathbb{Q} = \mathbb{Q}\big(i\sqrt{29 + 2\sqrt{29}}\big)$$

and we would like to check whether $\mathrm{End}\, J = \mathbb{O}_K$. The ring $\mathbb{O}_K$ is generated as a $\mathbb{Z}$-module by $1, \alpha, \beta, \gamma$, where α has order 2 in $\mathbb{O}_K/\mathbb{Z}[\pi, \bar{\pi}]$, β has order 4, and γ has order 40. The algorithm from [18] would require computing the elements of $J[2^3]$ and $J[5]$. But $(\mathbb{O}_K)_2 = \mathbb{Z}_2[\pi, \bar{\pi}, \alpha]$, so we only need to compute in $J[2]$ and $J[5]$.

2.5. ***The algorithm.*** Incorporating all these improvements yields the following algorithm:

Algorithm 10. Checking that $\mathrm{End}\, A$ is maximal.

Input: An ordinary abelian surface $A/\mathbb{F}_p$ with CM by K.

Output: *True* or *false*, depending on whether or not $\mathrm{End}\, A = \mathbb{O}_K$.

1. Choose a basis $\{1, \alpha_1, \alpha_2, \alpha_3\}$ of $\mathbb{O}_K$ and a basis $\{1, \beta_1, \beta_2, \beta_3\}$ of $\mathbb{Z}[\pi]$ such that $\beta_1 = c_1\alpha_1$, $\beta_2 = c_2\alpha_2$, $\beta_3 = c_3\alpha_3$ and $c_1, c_2, c_3 \in \mathbb{Z}$ with $c_1 \,|\, c_2 \,|\, c_3$.
2. (Checking where the ℓ-torsion lives.) For each $\ell \,|\, [\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$ do:
   (a) Let d be the smallest integer such that $\pi^d - 1 \in \ell\mathbb{O}_K$, and d' be the smallest integer such that $(-\pi)^{d'} - 1 \in \ell\mathbb{O}_K$. If $d' < d$, switch to the quadratic twist.
   (b) Compute a basis of $A[\ell](\mathbb{F}_{p^d})$ using the algorithm from [4].
   (c) If this basis is of cardinality (strictly) less than 4, return *false*.
   (d) (Checking the generators of $\mathbb{O}_K$.) For $i = 1, 2, 3$ do:
      i. Let N_1 be the order of α_i in $\mathbb{O}_K/\mathbb{Z}[\pi, \bar{\pi}, \alpha_j \,|\, j < i]$ and N_2 the order of α_i in $\mathbb{O}_K/\mathbb{Z}[\pi, \bar{\pi}]$.

ii. If $\ell \mid N_1$, let e be the ℓ-valuation of N_2 and write $pN_2\alpha_i$ as a polynomial $P(\pi)$.
iii. Compute a basis of $A(\mathbb{F}_{p^{d\ell^{e-1}}})[\ell^e]$.
iv. If $P(\pi) \neq 0$ on this basis, return *false*.

3. Return *true*.

2.6. ***Complexity.*** We will measure complexity in terms of operations in the base field $\mathbb{F}_p$, and we will neglect factors of $\log(p)$. Since the index $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ is bounded by a polynomial in p by [18, Proposition 6.2], evaluating the polynomials $P(\pi)$ (of degrees at most 3) is done in logarithmic time. The most expensive part of the algorithm is then the computation of $A[\ell^e]$, for the various ℓ dividing the index $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ where e is at most the ℓ-valuation of the index. According to Lemma 5 and Remark 6, the ℓ^e-torsion points live in an extension of degree at most $d = \ell^{e+3}$. Since $\#A(\mathbb{F}_{p^d}) = p^{2d(1+\epsilon)}$, computing a random point in $A(\mathbb{F}_{p^d})[\ell^e]$ takes $\widetilde{O}(d^2)$ operations in $\mathbb{F}_p$. Correcting this random point requires some pairing computations, and costs at most $O(\ell^2)$ (in case the first points give an isotropic group). Since we need $O(1)$ such random points, the global cost is given by the following proposition (we will only need a very rough bound for the complexity analysis in Section 6):

Proposition 11. *Let* $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] = \prod \ell_i^{e_i}$ *be the decomposition of the index into powers of primes. Then checking if an abelian surface in the isogeny class is maximal can be done in time* $\sum \widetilde{O}(\ell_i^{2e_i+6})$.

Remark 12. One can compare to [18, Proposition 4.6] to see the speedup we gain in the endomorphism ring computation. We note that our method is exponential in the discriminant, while in [3] one can find a subexponential algorithm to compute the endomorphism ring of an ordinary abelian surface. In ongoing work with Gaetan Bisson, we have developed a method that combines the going-up algorithm of the next section with his endomorphism ring algorithm. Since we still need to take ℓ-isogenies for $\ell \mid [\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$ in the going-up step, this approach is mainly interesting when the index is divisible by a power of a prime.

3. Going up

"Going up" is the process of finding genus-2 curves with maximal endomorphism ring by moving from *any* curve in the isogeny class to a maximal one via isogenies. This is not always possible and we will explain some of the obstructions. One difficulty was already illustrated in [6, Example 8.2], where it was shown that there can be cycles in the isogeny graph involving only nonmaximal curves. Clearly, when trying to "go up", the algorithm should avoid making cycles in the graph, and we propose one method to avoid that. Further difficulties arise from the fact

that the graph of rational (ℓ, ℓ) isogenies can be disconnected, and can even have isolated nodes. This is an important caveat, as this means that our method for going up will not always succeed, so we only have a probabilistic algorithm; furthermore, we cannot currently estimate the probability of failure.

As noted in [18], for the type of fields we can deal with via the CRT method, the cost of going through p^3 Jacobians is dominant compared to checking if the endomorphism ring is maximal. (This imbalance is magnified in our case due to our faster algorithm to compute the ℓ^e-torsion.) In our algorithm, we try to find a random curve in the isogeny class, and we try to select p so that the probability of finding a curve in the right isogeny class is of magnitude $p^{3/2}$. In practice, finding one such curve is still the dominant aspect, which explains why we can afford to spend a lot of effort on going up from this curve.

The algorithm we propose for going up is made possible by the techniques developed in [30; 13; 33] for computing rational (ℓ, ℓ)-isogenies between abelian surfaces over finite fields. If A is an ordinary abelian surface with CM by K, then for each ℓ dividing the index $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$, we try to find an (ℓ, ℓ)-isogeny path starting from A and going to A' such that $(\mathbb{O}_K)_\ell = (\operatorname{End} A')_\ell$. If this is possible, we let $A = A'$ in the next step (going to the next ℓ). A rather inefficient method for finding A' would be to use the algorithm for computing endomorphism rings which was detailed in the preceding section (modified to handle the case of nonmaximal orders), compute the endomorphism ring of End A and the (ℓ, ℓ)-isogenous surfaces A', and keep A' if its endomorphism ring is bigger than that of A. In this section we will describe a more efficient algorithm, which combines the endomorphism ring checks of the preceding section with a going-up phase. Since we are working locally in ℓ, we may as well suppose that we are working over $\mathbb{Z}_\ell$.

3.1. ***Going up for one endomorphism.*** In this section, we suppose that we have an element $\alpha' \in \mathbb{O}_K$ such that $\alpha := \gamma \ell^e \alpha'$ lies in $\mathbb{Z}[\pi]$ for some $\gamma \in \mathbb{O}_K$ prime to ℓ. Starting from an abelian surface A in the isogeny class, we want to find an abelian surface A' such that $\alpha/\ell^e \in \operatorname{End} A'$ (or equivalently that $\alpha' \in \operatorname{End} A'$ locally at ℓ).

We saw in Section 2 that α/ℓ^e is in the endomorphism ring of A if and only if $\alpha(A[\ell^e]) = 0$, and we know how to compute this subgroup. More generally, we let $N = \#\alpha(A[\ell^e])$. We think of N as a way to measure the "obstruction" to α/ℓ^e being an element of End A. Our algorithm is as follows: For each (ℓ, ℓ)-isogenous surface A', we let $N' = \#\alpha(A'[\ell^e])$ and we replace A by A' if $N' < N$. We iterate this process until $N = 1$, in which case we have succeeded, or until we are stuck, in which case we try to find a new random abelian surface in the right isogeny class.

Rather than directly computing the obstruction $N = \#\alpha(A[\ell^e])$, we can compute the partial obstructions $N(\epsilon) := \#\alpha(A[\ell^\epsilon])$ for $\epsilon \leq e$. Starting from $\epsilon = 1$, we take isogenies until we find an abelian surface A with $N(\epsilon) = 1$, which means that

$\alpha/\ell^\epsilon \in \mathrm{End}\, A$. We will now try to take isogenies to reduce the obstruction of higher degree $N(\epsilon+1)$. Let $k = \alpha(A[\ell^{\epsilon+1}]) \subseteq A[\ell]$. The following lemma helps us select the isogeny we are looking for:

Lemma 13. *With notation and assumptions as above, let A' be an abelian surface isogenous to A such that $\#\alpha(A'[\ell^{\epsilon+1}]) < \#\alpha(A[\ell^{\epsilon+1}])$. Then the kernel of the isogeny $A \to A'$ intersects nontrivially with $k = \alpha(A[\ell^{\epsilon+1}])$.*

Proof. Let $f : A \to A'$ be a rational isogeny between A and A'. Then since α is a polynomial in the Frobenius, we have $\alpha \circ f = f \circ \alpha$. In particular, f maps $\alpha(A[\ell^{\epsilon+1}])$ to $\alpha(A'[\ell^{\epsilon+1}])$. If $\#\alpha(A'[\ell^{\epsilon+1}]) < \#\alpha(A[\ell^{\epsilon+1}])$ then there exists $x \in \mathrm{Ker}\, f \cap \alpha(A[\ell^{\epsilon+1}])$. □

This gives the following algorithm:

Algorithm 14. Going up for one endomorphism α/ℓ^ϵ.

Input: An ordinary abelian surface $A/\mathbb{F}_p$ with CM by K, a prime power ℓ^e, and an $\alpha \in \ell^e \mathbb{O}_K$.

Output: An abelian surface $A'/\mathbb{F}_p$ isogenous to A such that $\alpha/\ell^\epsilon \in \mathrm{End}\, A'$, or *fail*.

1. Set $\epsilon = 1$.
2. Compute $N(\epsilon) = \#\alpha(A[\ell^\epsilon])$.
3. If $N(\epsilon) = 1$, do:
   (a) If $\epsilon = e$ then return A.
   (b) Otherwise, set $\epsilon := \epsilon + 1$, and go back to Step 2.
4. At this point, $N(\epsilon) > 1$. Let $\mathcal{L}$ be the list of all rational maximal isotropic subgroups of $A[\ell]$ which intersect nontrivially with $\alpha(A[\ell^\epsilon])$. For $k \in \mathcal{L}$ do:
   (a) Compute $A' = A/k$.
   (b) Let $N'(\epsilon) = \#\alpha(A'[\ell^\epsilon])$.
   (c) If $N'(\epsilon) < N(\epsilon)$, set $A = A'$ and go back to Step 2.
5. Return *fail*.

Remark 15. As in Section 2 we let d_0 be the minimal integer such that $(\pi^{d_0} - 1) \in \ell\mathbb{O}_K$ and d the minimal integer such that $(\pi^d - 1) \in \ell\mathbb{Z}[\pi]$. Then the ℓ^ϵ-torsion points of A are defined over an extension of degree $d\,\ell^{\epsilon-1}$. If moreover $(\pi^{d_0} - 1)/\ell \in \mathrm{End}\, A$ they are actually defined over an extension of degree $d_0\ell^{\epsilon-1}$.

Therefore when we try to go up globally for all endomorphisms α, the first step is to try to go up for the endomorphism $(\pi^{d_0} - 1)/\ell$. During the algorithm, the obstruction N is given by the size of the kernel of $\pi^{d_0} - 1$, whose rank is 4 minus the rank of the ℓ-torsion points defined over $\mathbb{F}_{p^{d_0}}$. So we compute the size of a basis of $A[\ell](\mathbb{F}_{p^{d_0}})$ and take isogenies, where this size increases until we find the full rank.

3.2. ***Going up globally.*** Let $\{1, \alpha_1/\ell^{e_1}, \alpha_2/\ell^{e_2}, \alpha_3/\ell^{e_3}\}$ be a generating set for the maximal order $(\mathbb{O}_K)_\ell$ over the subring $\mathbb{Z}_\ell[\pi, \bar{\pi}]$, where $\alpha_i \in \mathbb{Z}_\ell[\pi, \bar{\pi}]$. Starting from an abelian surface A in the isogeny class, we want to find an abelian surface which is maximal at ℓ.

We could apply Algorithm 14 for each α_i/ℓ^{e_i}, but the algorithm does not guarantee that the endomorphisms already defined on A stay defined during the process, so we would observe loops on nonmaximal abelian surfaces with this method. Moreover we want to reuse the computations of $A[\ell^\epsilon]$, which are the expensive part of the process.

If $N_i = \#\alpha_i(A[\ell^{d_i}])$ for $i = 1, 2, 3$ is the obstruction corresponding to α_i, we define N to be the global obstruction $N = \sum N_i$. We can then adapt the same method: For each (ℓ, ℓ)-isogenous A', if $N_i' = \#\alpha_i(A'[\ell^{d_i}])$, then we replace A by A' if $\sum N_i' < \sum N_i$. We iterate this process until all the $N_i = 1$, in which case we go to the next ℓ, or until we are stuck, in which case we try to find a new random abelian surface in the right isogeny class.

As before, if $e = \max(e_1, e_2, e_3)$ we first compute $A[\ell^\epsilon]$ and the partial obstructions $N_i(\epsilon) = \#A[\ell^{\min(\epsilon, e_i)}]$ (for $i = 1, 2, 3$). We do the same for the (ℓ, ℓ)-isogenous abelian surfaces, and switch to the new one if $\sum N_i(\epsilon)$ decreases (strictly). This allows working with smaller torsion in the beginning steps.

The level ϵ of the individual obstruction we are working on depends on the endomorphism considered, so if we get stuck on level ϵ, we may have to look at level $\epsilon + 1$ even if not all endomorphisms α_i/ℓ^ϵ are defined yet. For instance, in the case where we are only dealing with two generators, there are examples where $N_1(\epsilon) = 1$, $N_2(\epsilon) \neq 1$ and $N_1'(\epsilon) = 1$, $N_2'(\epsilon) = N_2(\epsilon)$ for all (ℓ, ℓ)-isogenous abelian surfaces A', so we are stuck on level ϵ. However we can still find an isogenous A' such that $N_1'(\epsilon + 1) < N_1(\epsilon + 1)$.

Finally, as in Remark 15, we first try to go up in a way that increases the size of $A(\mathbb{F}_{p^{d_0}})[\ell]$. If we are unlucky and get stuck, we switch to the computation of the full ℓ-torsion over $\overline{\mathbb{F}}_p$. This method allows working over the smallest extension to compute $A[\ell^e]$ as soon as possible.

A summary of the algorithm with the notation from above is given below:

Algorithm 16. Going up.

Input: An ordinary abelian surface $A/\mathbb{F}_p$ with CM by K, and a prime ℓ.

Output: An abelian surface $A'/\mathbb{F}_p$ with $\operatorname{End} A = \mathbb{O}_K$ (locally at ℓ), or *fail*.

1. (Special case for the endomorphism $(\pi^{d_0} - 1)/\ell$.) Compute a basis B of $A(\mathbb{F}_{p^{d_0}})[\ell]$. If $\#B < 4$, compute a basis B' of $A'(\mathbb{F}_{p^{d_0}})[\ell]$ for each (ℓ, ℓ)-isogenous abelian surface A'. If $\#B' > \#B$, restart the algorithm with $A' = A$. If $\#B = 4$ or we get stuck, go to the next step.
2. Set $\epsilon = 1$.

3. Compute[1] $N_i(\epsilon) = \#\alpha_i(A[\ell^{\min(\epsilon,e_i)}])$ for $i = 1, 2, 3$.
4. If $\{N_i : i = 1, 2, 3\} = \{1\}$, do:
   (a) If $\epsilon = \max(e_i : i = 1, 2, 3)$ then return A.
   (b) Otherwise, set $\epsilon := \epsilon + 1$ and go back to Step 3.
5. Let $\mathcal{L}$ be the list of all rational maximal isotropic kernels of $A[\ell]$ which intersect nontrivially with one of the $\alpha_i(A[\ell^{\min(\epsilon,e_i)}])$. For $k \in \mathcal{L}$ do:
   (a) Compute $A' = A/k$.
   (b) Let $N_i'(\epsilon) = \#\alpha_i(A'[\ell^{\min(\epsilon,e_i)}])$.
   (c) If $\sum N_i'(\epsilon) < \sum N_i(\epsilon)$, restart the algorithm with $A = A'$ (but do not reinitialize ϵ in Step 2).
6. If we get stuck and $\epsilon < \max(e_i : i = 1, 2, 3)$, set $\epsilon := \epsilon + 1$ and go back to Step 3.
7. Return *fail*.

3.3. ***Cost of the going-up step.*** We will see in the examples that the going-up step is a very important part in speeding up the CRT algorithm in practical computations. However, since it is doomed to fail in some cases (see Remark 18), we need to check that it will not dominate the complexity of the rest of the algorithm, so that in theory there will be no drawback to using it. Thus we need to estimate the cost of the going-up step.

The going-up phase is a mix of endomorphism testing and isogeny computations. We already analyzed the cost of the endomorphism testing in the preceding section. For the isogeny computation, the points in the kernel of rational (ℓ, ℓ)-isogenies live in an extension of degree at most $\ell^2 - 1$. Transposing the analysis of Section 2.6 to this case shows that the computation of all of the points in these kernels takes at most $\widetilde{O}(\ell^4)$ operations in $\mathbb{F}_p$. There are at most $O(\ell^3)$ such kernels, and each isogeny computation takes at most $\widetilde{O}(\ell^4)$ operations in the extension. The final cost is at most $\widetilde{O}(\ell^9)$ operations in $\mathbb{F}_p$ for computing all isogenies. For each of the $O(\ell^3)$ isogenous abelian surfaces we do (part of) the endomorphism ring computation, which takes $\widetilde{O}(\ell^{2e+6})$ operations, according to Section 2.6. Since the global obstruction computed is of size $O(\ell^e)$, we do at most $O(e)$ steps. The global complexity is then given as follows:

Proposition 17. *Let* $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] = \prod \ell_i^{e_i}$ *be the decomposition of the index into prime factors. Then the going-up phase either fails or is done in at most* $\widetilde{O}(\sum \ell_i^{2e_i+9})$ *operations in the base field.*

[1]The degree of the extension where the full ℓ^ϵ-torsion is defined depends on whether Step 1 succeeded.

Remark 18. It is important to note that the going-up phase does not always succeed. We will give some examples of that in Section 7. First, as noted in the introduction of this section, the (ℓ, ℓ)-isogeny graph is not always connected, so if we start with a curve not in the same component as a maximal curve, there is no way to find the maximal curves using only (ℓ, ℓ)-isogenies. Second, even if the curve is in the same component as a maximal curve, finding a maximal curve may involve going through isogenous curves that increase the global obstruction, so the going-up algorithm would not find it.

In practical computations we observed the following behavior: In the very large majority of the cases where we were not able to go up, there actually did not exist any rational (ℓ, ℓ)-isogenies for any curve in the isogeny class. If χ_π is the characteristic polynomial, this can be detected by the fact that χ_π does not factor modulo ℓ as $\chi_\pi = P\,\overline{P} \pmod{\ell}$ (where $\overline{P}$ is the conjugate of P under the action $\pi \to p/\pi$, which sends the Frobenius to the Verschiebung). In this situation, there is no way to go up even locally at ℓ. This gives a criterion for estimating whether one can go up for this ℓ.

4. Computing maximal curves from maximal curves

Once a maximal curve in the isogeny class has been found via the random search and going-up steps, we use isogenies to find the other maximal curves. The set of maximal curves in the isogeny class corresponding to a fixed CM-type Φ is a principal homogeneous space under the action of the Shimura class group

$$\mathfrak{C} = \{(I, \rho) \,|\, I \text{ a fractional } \mathbb{O}_K\text{-ideal with } I\,\bar{I} = (\rho),\ \rho \in K^+ \text{ totally positive}\}/K^*,$$

associated to the primitive quartic CM field K, which acts by isogenies (see for instance [6, §3]).

However, using the Magma package AVIsogenies we can only compute isogenies with a maximal isotropic kernel. The lemma below show that in terms of the Shimura class group, this means that we can only compute the action corresponding to (equivalences classes) of elements of the form (I, ℓ), where I is an ideal in K and ℓ is a prime number.

Lemma 19. *Let (I, ρ) be an element of the Shimura class group $\mathfrak{C}$ and let ℓ be a prime. Then the action of (I, ρ) on a maximal abelian surface A corresponds to an isogeny with maximal isotropic kernel in $A[\ell]$ if and only if $\rho = \ell$ (so if and only if I has relative norm ℓ).*

Proof. This follows from the construction of the action of $\mathfrak{C}$ on the set of maximal abelian surfaces. The action is given by the isogeny $f : \mathbb{C}^2/\Lambda \to \mathbb{C}^2/I\Lambda$ and moreover the action of $\bar{I}$ corresponds to the dual isogeny $\widehat{f}$ (here we identify the abelian surface A with its dual $\widehat{A}$ via the principal polarization induced from the

CM data). Since ℓ is prime, the isogeny corresponding to I is an (ℓ, ℓ) isogeny if and only if $I\bar{I} = (\rho) = (\ell)$. □

Therefore to ensure that we can find all other maximal curves using this type of isogeny we make the following heuristic assumption.

Assumption. There is a polynomial P such that for every primitive quartic CM field K, the Shimura class group associated to K is generated by elements of the form (I, ℓ), where ℓ ranges over the prime numbers less than $P(\log \Delta)$ and where Δ is the discriminant of K.

Justification. We have tested this assumption on numerous examples, using the bound $12 \log \Delta'$, where Δ' is the discriminant of the reflex field, which is itself $O(\Delta^2)$. The assumption on the size of the isogenies will be used in the complexity analysis. At worst, we know (under GRH) that the class group of the reflex field is generated by prime ideals of degree one and of norm polynomial in $\log \Delta$ [1, Theorem 1]. But if I is such an ideal of $\mathbb{O}_{K_\Phi}$ of norm prime to p, then the element $(\mathrm{TN}(I), N(I))$ will give a horizontal isogeny. So we will at least be able to compute all the maximal curves that are deduced from the first one by an action coming from the type norm. As we will see in the complexity analysis in Section 6, this is sufficient for most discriminants Δ. □

Lemma 20. *Let A be an ordinary abelian surface with* $(\mathrm{End}\, A) \otimes \mathbb{Q} = K$, *and let $f : A \to B$ be an isogeny of degree prime to* $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$. *Then* $\mathrm{End}\, A = \mathrm{End}\, B$.

Proof. Let d be the smallest integer that factorizes through f, so $d = f\tilde{f}$ for some isogeny $\tilde{f} : B \to A$. By assumption d is prime to the index. If $\alpha \in \mathrm{End}\, A$, then $f \circ \alpha \circ \tilde{f} = d\alpha$ is an endomorphism of B. Since $[\mathbb{O}_K : \mathrm{End}\, B]$ is prime to d, we have that $\alpha \in \mathrm{End}\, B$. The same argument shows that $\mathrm{End}\, B \subseteq \mathrm{End}\, A$, so $\mathrm{End}\, A = \mathrm{End}\, B$. □

Note that we can precompute generators of the Shimura class group since this data does not depend on the current prime p. We want to find generators of relative norm a prime $\ell \in \mathbb{Z}$ with ℓ as small as possible, since the size of ℓ will directly influence the time spent to find the other maximal curves.

Now for a CRT prime p, there may exist among the generators we have chosen some that divide the index $[\mathbb{O}_K : \mathbb{Z}[\pi, \bar{\pi}]]$. We can either find other generators (whose norm will be bigger), or still try to use the precomputed generators. In this case, if such a generator has norm ℓ, then not all new (ℓ, ℓ)-isogenous abelian surfaces will be maximal, so we have to use Algorithm 10 to test which of them is maximal. In that case, after the isogeny is applied, the ℓ^e-torsion (in the notation of Section 3) must again be computed, along with the action of the generators of $(\mathbb{O}_K)_\ell$ over $\mathbb{Z}[\pi, \bar{\pi}]_\ell$. The trade-off depends then on the degree of the extension field required to compute the ℓ^e-torsion for small ℓ dividing the index versus the

degree of the field of definition for the points in the kernel of the ℓ-isogeny for ℓ not dividing the index.

Finally, we can also use the group structure of the Shimura class group as follows: Suppose that we have computed maximal curves corresponding to the action of $\alpha_1, \ldots, \alpha_t \in \mathfrak{C}$, and we want to find new maximal curves by computing (ℓ, ℓ)-isogeny graphs starting from these curves. Then if $\mathfrak{C}(\ell)$ is the set of elements of the form (I, ℓ) in $\mathfrak{C}$, then the number of maximal curves that we can find in this way is the cardinality of the subgroup generated by the α_i and $\mathfrak{C}(\ell)$. In particular, as soon as we reach this number, we can stop the computation since it will not yield any new maximal curves. This is particularly useful when ℓ divides the index, because then we avoid some endomorphism tests. In the isogeny graph computation done by AVIsogenies, each node is computed twice since there are two edges between adjacent nodes (corresponding to the isogeny and the dual). Here, since we know the number of nodes, we can abort the computation early.

We thus obtain the following algorithm:

Algorithm 21. Finding all maximal curves from one maximal curve.

Input: An ordinary abelian surface $A/\mathbb{F}_p$ with CM by $(\mathbb{O}_K, \Phi)$.

Output: All abelian surfaces over $\mathbb{F}_p$ with CM by $(\mathbb{O}_K, \Phi)$.

1. Precomputation: Compute a set of generators of the Shimura class group with relative norm ℓ as small as possible. (The set is not chosen to be minimal; on the contrary, we want some redundancy.) For each of the generators, compute the extension degree of the field of definition of the geometric points of the kernel corresponding to this generator.
2. For each generator of (relative) norm ℓ dividing the index, replace the previous degree by the degree of the extension where the ℓ^e-torsion lives. (Usually e is the ℓ-valuation of the index, but the tricks from Section 2 can sometimes reduce it.)
3. Sort the generators by the corresponding degrees to get a list $(g_1, \ldots, g_n)$.
4. For each generator g_i on the list, let ℓ_i be its norm and do:
   (a) Compute the surfaces (ℓ_i, ℓ_i)-isogenous to the one already found. If ℓ_i divides the index, then do an endomorphism ring computation from Section 2 and keep only the maximal curves.
   (b) Repeat until the number of maximal abelian surfaces is $\#\langle \mathfrak{C}(\ell_1), \ldots, \mathfrak{C}(\ell_i)\rangle$.

5. The CRT step

In contrast to the elliptic curve case, the coefficients of the class polynomials in genus 2 are rational numbers, not integers. We estimate the denominators of these

rational numbers by using the Bruinier-Yang conjectural formula [7] (proved only for special cases [43; 44]), together with minor adjustments from [22]; since we are using the invariants from [38, Appendix 3], we must also alter the denominator formulas by small powers of 2. A formula for the factorization of the denominators that holds for general primitive quartic CM fields was recently given in [26]; this formula produces a multiple of the denominators, because it allows for cancellation with the numerators and for the case where K^+ does not have class number 1. As in [16, Theorem 3], we can multiply coefficients by their denominators, and then use the CRT to reconstruct the polynomials.

5.1. ***Sieving the CRT primes.*** To determine whether to use a CRT prime in the CRT algorithm, we check if the corresponding isogeny class is large enough. There are approximately $2p^3$ isomorphism classes of genus-2 curves over $\mathbb{F}_p$ (see [5, Proposition 7.1]), and since the area of Figure 10.1 in [29] is $32/3$, there are approximately $(32/3)p^{3/2}$ isogeny classes. We keep p if the size of the isogeny class is of size roughly $p^{3/2}$. We could compute the size of this isogeny class by using Lemma 6.3 in [29] for each order (stable by conjugation) between $\mathbb{Z}[\pi, \bar{\pi}]$ and $\mathbb{O}_K$, but since computing the lattice of orders is quite costly, we instead use a heuristic derived from this formula. More details on this heuristic are given in [28, §7.2].

In practice, we are only interested in the number of curves from which we can go up. This is harder to estimate, but numerous computations showed that the main obstruction to going-up occurs when there are no (ℓ, ℓ)-isogenies with rational kernel at all. But this case is easy to detect (see [4]). So in the previous estimate, we discount the orders whose index is divisible by such an ℓ.

Finally, we use a dynamic approach for the prime selection: We use a prime if the probability of finding a maximal curve with the going-up algorithm is better than a certain threshold (depending on the size of the prime), but we go back to previously discarded (smaller) primes if they satisfy the threshold for the current size of primes we are considering.

5.2. ***The CRT.*** In the cyclic case, we compute the class polynomials modulo small integer primes, and we use the CRT to get the result modulo the product P (the "precision") of these small primes. Once the precision is large enough, we can recover the polynomials over $\mathbb{Z}$ by lifting each coefficient to an integer in the interval $[-P/2, P/2]$.

In the dihedral case, the primes are in $\mathbb{O}_{K_\Phi^+}$, and so is the precision ideal P. Here we explain how to lift a coefficient x mod P to $\mathbb{O}_{K_\Phi^+}$. Take the Minkowski embedding of a lift of x, and find the closest vector c_x in the lattice associated to P in the Minkowski embedding. Then c_x corresponds to an element of the ideal P, and our final lift is $x - c_x$. We note that the lattice is of rank 2, so we can directly compute the closest vector rather than doing an LLL approximation.

5.3. ***Lifting without denominators.*** We note that in the dihedral case, the denominator from the formulas in [7; 22; 26] is too large, as it takes into account both CM-types. This increases the size of the coefficients we compute, so that using those denominator formulas does not actually give better results than doing a rational reconstruction directly.

With the notation from above, from $x \bmod P$ we want to do a rational lift of x. This time we embed the lattice associated to P into the lattice of rank 3 obtained by adjoining the vector $[Cx_1, Cx_2, C]$ where x_1 and x_2 are the two real embeddings of (a lift of) x and C is a constant accounting for how skewed we expect the size of the denominator to be compared to the numerators. A minimal vector in this lattice will correspond to an element $N = c + Dx$ where $c \in p$ and D is an integer. We then take N/D as our lift for x.

This solution requires the precision to be the sum of the bit sizes of the numerators and denominator, so it can be even better than using the denominator formulas for small denominators, where there may be cancellation with the numerators.

6. Complexity

In this section, we give a mostly heuristic analysis of how Algorithm 16 (the going-up algorithm) and Algorithm 21 (the algorithm to find all maximal curves from one maximal curve) affect the asymptotic complexity of Algorithm 1. We will sometimes call the isogenies we compute in the going-up algorithm *vertical steps*, and the isogenies we compute in Algorithm 21 *horizontal steps*; this is in analogy with the corresponding terminology in the elliptic curve case.

We begin with a quick reminder of the rough complexity analysis of the CRT method in the elliptic curve case, where K is a quadratic imaginary field. In this case there is only one class polynomial H, whose degree is the class number of $\mathbb{O}_K$, and classical bounds give that $\deg H = \widetilde{O}(\sqrt{\Delta})$, where Δ is the discriminant of $\mathbb{O}_K$. Likewise, the coefficients of H have size $\widetilde{O}(\sqrt{\Delta})$. So the whole class polynomial is of size $\widetilde{O}(\Delta)$.

Each CRT prime p gives $\log(p)$ bits of information, so neglecting logarithmic factors, we need about $\sqrt{\Delta}$ primes. CRT primes split completely in the Hilbert class field of K, whose Galois group is $\mathrm{Cl}(\mathbb{O}_K)$, so by the Cebotarev theorem the density of CRT primes is roughly $1/\#\mathrm{Cl}(\mathbb{O}_K) \simeq 1/\sqrt{\Delta}$. Neglecting logarithmic factors again, we therefore expect the biggest prime p to be of size $\widetilde{O}(\Delta)$.

Now there are $O(p)$ isomorphism classes of elliptic curves, and $\widetilde{O}(\sqrt{\Delta})$ maximal curves, so one is found in time $\widetilde{O}(p/\sqrt{\Delta}) = \widetilde{O}(\sqrt{p})$. Once one maximal curve is found, all others can be obtained using isogenies of degree logarithmic in Δ, so one can recover all maximal elliptic curves over $\mathbb{F}_p$ in time $\widetilde{O}(\sqrt{p}) = \widetilde{O}(\sqrt{\Delta})$.

We need $\sqrt{\Delta}$ CRT primes, so the total cost is $\widetilde{O}(\Delta)$. The CRT reconstruction can be done in quasilinear time too, so in the end the algorithm is quasilinear, even without using a vertical step. If we had not used horizontal steps, the complexity would have been $\widetilde{O}(\Delta^{3/2})$.

Now consider the genus-2 case. Let $\Delta_0 = \Delta_{K^+/\mathbb{Q}}$ and $\Delta_1 = N_{K^+/\mathbb{Q}}(\Delta_{K/K^+})$, so $\Delta = \Delta_{K/\mathbb{Q}} = \Delta_1\Delta_0^2$. Then the degree of the class polynomials is $\widetilde{O}(\Delta_0^{1/2}\Delta_1^{1/2})$, while the height of their coefficients is bounded by $\widetilde{O}(\Delta_0^{5/2}\Delta_1^{3/2})$ (see [38, §II.9] and [21]). In practice, we observe [38, Appendix 3] that the coefficient height is bounded by $\widetilde{O}(\Delta_0^{1/2}\Delta_1^{1/2})$, and we will use this observed bound in the following analysis. According to [6, §6.4], the smallest prime is of size $\widetilde{O}(\Delta_0\Delta_1)$. We need $\widetilde{O}(\Delta_0^{1/2}\Delta_1^{1/2})$ CRT primes, and an analysis using [24], as in [2, §5, Lemma 3], shows that the largest prime is also $\widetilde{O}(\Delta_0\Delta_1)$. We remark that the sieving phase does not affect the size of the largest prime (apart from the constant in the big O) as long as we sieve a positive density of CRT primes.

For the horizontal step, the isogeny computation involves primes of size logarithmic in Δ, so the cost of this step is quasilinear in the number $\widetilde{O}(\Delta_0^{1/2}\Delta_1^{1/2})$ of maximal curves. This is under the Assumption from Section 4. Without this assumption, what we know is that for each ideal I in $\mathbb{O}_{K_\Phi}$ of norm prime to p, the element $(\mathrm{TN}(I), N(I))$ is an element of the Shimura class group whose action is given by a maximally isotropic kernel. In the horizontal step, we can then compute the action of $\mathrm{TN}(\mathrm{Cl}(\mathbb{O}_{K_\Phi}))$ by isogenies of size logarithmic in Δ. By Lemma 6.5 of [6], the cofactor is bounded by $2^{6w(D)+1}$, where $w(D)$ is the number of prime divisors of D. This gives a bound on the number of horizontal isogeny steps we need to take. As remarked in [6, p. 516], we have $w(n) < 2\log\log n$ outside a density-0 subset of very smooth integers, so the corresponding factor can be absorbed into the $\widetilde{O}$-notation.

In contrast, the complexity of the endomorphism ring computation and the going-up phase involves the largest prime power dividing the index $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]]$. According to Proposition 6.1 of [18] we have that $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] \leq 16p^2/\sqrt{\Delta}$. For the size of the CRT prime we are considering, we see that $[\mathbb{O}_K : \mathbb{Z}[\pi, \overline{\pi}]] = \widetilde{O}(\Delta_0\Delta_1^{3/2})$. We fix $\epsilon = 1/2$. Assuming that the index is uniformly distributed, [15] showed that there is a positive density of CRT primes where the largest prime power dividing the index is $O(\Delta_0^{\epsilon/100}\Delta_1^{\epsilon/100})$. By the complexity analysis of Sections 2.6 and 3.3, we see then that there is a positive density of primes where these algorithms take time at most $O(\Delta_0^{\epsilon}\Delta_1^{\epsilon})$.

We then let $p = \widetilde{O}(\Delta_0\Delta_1)$ be a CRT prime. There are $O(\sqrt{p})$ maximal curves, so we expect the isogeny class to be of size $\Theta(p^{3/2})$ (see Heuristic 6.6 in [6]). Up to isomorphism over the algebraic closure, there are p^3 genus-2 curves over $\mathbb{F}_p$. The original CRT algorithm of [16; 18] looped through all p^3 geometric isomorphism classes of curves and tested whether the corresponding endomorphism ring

is maximal. This takes time $\widetilde{O}(\Delta_0^3\Delta_1^3) + O(\Delta_0^{3/2+\epsilon}\Delta_1^{3/2+\epsilon})$ per CRT prime. Since $\widetilde{O}(\Delta_0^{1/2}\Delta_1^{1/2})$ CRT primes are needed, we find a total cost of $\widetilde{O}(\Delta_0^{7/2}\Delta_1^{7/2})$, given our choice of ϵ.

The approach of [6] is to search for only one maximal curve, and then to use horizontal isogenies to find the others. With the improvements proposed in this paper (using *all* horizontal isogenies and not just those coming from the type norm, and the improved endomorphism ring computation), we find a cost of $\widetilde{O}(\Delta_0^{5/2}\Delta_1^{5/2}) + O(\Delta_0^{3/2+\epsilon}\Delta_1^{3/2+\epsilon})$ per CRT prime. The total cost is then $\widetilde{O}(\Delta_0^3\Delta_1^3)$.

With our method, we need to find a curve in the isogeny class where the going-up step yields a maximal curve. Finding a curve in the isogeny class takes time $O(p^{3/2})$. If X is the number of going-up steps we need to try on average, the cost per CRT prime is then $\widetilde{O}(X(\Delta_0^{3/2}\Delta_1^{3/2} + \Delta_0^{\epsilon}\Delta_1^{\epsilon}))$. At best, $X = O(1)$, and we have a total cost of $\widetilde{O}(\Delta_0^2\Delta_1^2)$ from CRT primes. So at best we have a quasiquadratic complexity, while the CRT itself is quasilinear, and thus negligible. We see that we are still far from quasilinearity achieved by the analytic method. At worst, $X = O(p)$ (number of random tries in the isogeny class until we find a maximal one directly), and we recover the quasicubic complexity of the previous method.

To improve the complexity, there are two possibilities. The first is to increase the probability of success of the going-up method. This requires an algorithm to compute isogenies with cyclic kernels. But even with that, we achieve at most quasiquadratic complexity because the size of the isogeny class is too small compared to the size of the search space. This is the case because the algorithm computes the class polynomials (a scheme of dimension 0) directly from the moduli space of dimension 3 of all abelian surfaces. In contrast, in the elliptic curve case, the algorithm searches a space of dimension 1 for elements of a space of dimension 0. It would be interesting to find convenient subspaces of the moduli space of smaller dimension, and to work over them. One example would be to use Humbert surfaces, which are of dimension 2, and Gundlach invariants, as proposed in [27].

7. Examples

7.1. ***Improvements due to the going-up phase.*** We first look at improvements due solely to the going-up phase. The new timings for the case $K = \mathbb{Q}(i\sqrt{29 + 2\sqrt{29}})$, $\mathfrak{C}(\mathbb{O}_K) = \{0\}$, are given in Table 1, compared to old timings from [18, §9]. This is a cyclic Galois example with class number one, so there is only one maximal curve and the algorithm from Section 4 is not used.

Note that much less time is spent exploring curves with the new algorithm, due to the going-up algorithm. Also note that, even though the going-up phase is more complicated, it is still less costly than the computation of the endomorphism rings in the old algorithm, due to the improvements described in Section 2 and the fact that the new version calls it less often.

					Timings (in seconds)	
p	l^d	α_d	Curves	Estimate	Old	New
7	—	—	1	1	0.3 + 0.0	0.1 + 0.0
23	**13**	84	15	2 (16)	9 + 70.7	0.4 + 24.6
53	7	3	7	7	105 + 0.5	7.7 + 0.5
59	2, **5**	1, 12	322	48 (286)	164 + 6.4	1.4 + 0.6
83	3, 5	4, 24	77	108	431 + 9.8	2.4 + 1.1
103	*67*	*1122*				
107	7, **13**	3, 21	105	8 (107)	963 + 69.3	
139	$\mathbf{5^2}$, 7	60, 2	259	9 (260)	2189 + 62.1	
181	3	1	161	135	5040 + 3.6	4.5 + 0.2
197	5, 109	24, *5940*				
199	$\mathbf{5^2}$	60	37	2 (39)	6360 + 1355.3	
223	2, *23*	1, 11	1058	39 (914)	10440 + 35.1	
227	109	*1485*				
233	5, 7, **13**	8, 3, 28	735	55 (770)	11580 + 141.6	88.3 + 29.4
239	7, 109	6, *297*				
257	3, 7, **13**	4, 6, 84	1155	109 (1521)	17160 + 382.8	
313	3, **13**	1, 14		146 (2035)		165.0 + 14.7
373	5, 7	6, 24		312		183.4 + 3.8
541	2, 7, **13**	1, 3, 14		294 (4106)		91.0 + 5.5
571	3, **5**, 7	2, 6, 6		1111 (6663)		96.6 + 3.1
Total time for calculating class polynomials:					56585	776

Table 1. Timings and other information for the old and new algorithms to compute the Igusa class polynomials for the field $\mathbb{Q}(i\sqrt{29+2\sqrt{29}})$, using a 2.39 GHz AMD Opteron with 4 GB of RAM. The first column gives the possible CRT primes; an entry in the "Timings" column indicates whether this CRT prime was used in the calculation. The second column lists the ℓ^d-torsion subgroups required to compute whether a curve is maximal; bold entries indicate that there are no rational (ℓ, ℓ)-isogenies (so that "going up" is not possible), and italic entries indicate that (ℓ, ℓ)-isogenies are too expensive to compute. The third column gives the degree of the field extensions where the points of these subgroups live; the degree is italicized when it is so large that computing the ℓ^d-torsion would be too expensive. The fourth column indicates the total number of curves in the isogeny class, computed via the algorithm from [18]. The fifth gives an estimate, obtained as explained in Section 5.1, for the number of curves from which we can go up, and, in parentheses, for the total number of curves in the isogeny class. The last two columns give the timings of the old and new algorithms, split into "Time exploring curves" + "Time spent computing endomorphism rings/Time spent going up". The old timings are obtained from [18, Table 3]. The total times listed on the last line include some overhead not accounted for elsewhere.

The trade-offs in the going-up step depend on the discriminant of the CM field K. The more CRT primes we need, the bigger the isogenies and the bigger the degrees in the endomorphism ring computations we allow. Note that computing (ℓ, ℓ)-isogenies requires $\widetilde{O}(\ell^2)$ operations in the field where the points of the kernel are defined when ℓ is congruent to 1 (mod 4), but $\widetilde{O}(\ell^4)$ when ℓ is congruent to 3 (mod 4). So in the above example, we computed the $(109, 109)$-isogenies faster than the $(23, 23)$-isogenies.

7.2. ***Dihedral examples.*** Here we illustrate our new CRT algorithm for dihedral fields, for $K = \mathbb{Q}(X)/(X^4 + 13X^2 + 41)$ with $\mathfrak{C}(K) \simeq \{0\}$.

We first compute the class polynomials over $\mathbb{Z}$ using Spallek's invariants, and obtain the following polynomials in 5956 seconds:

$$H_1 = 64X^2 + 14761305216X - 11157710083200000,$$
$$H_2 = 16X^2 + 72590904X - 8609344200000,$$
$$H_3 = 16X^2 + 28820286X - 303718531500.$$

Next we compute them over the real subfield and use the invariants from [38, Appendix 3]. We get:

$$H_1 = 256X - 2030994 + 56133\alpha,$$
$$H_2 = 128X + 12637944 - 2224908\alpha,$$
$$H_3 = 65536X - 11920680322632 + 1305660546324\alpha,$$

where α is a root of $X^2 - 3534X + 177505$, so that $\mathbb{O}_{K_0^+} = \mathbb{Z}[\alpha]$. This computation took 1401 seconds, so in this case, the speedup due to using better invariants and computing over the real subfield is more than 4-fold.

References

[1] Eric Bach, *Explicit bounds for primality testing and related problems*, Math. Comp. **55** (1990), no. 191, 355–380. MR 91m:11096

[2] Juliana Belding, Reinier Bröker, Andreas Enge, and Kristin Lauter, *Computing Hilbert class polynomials*, in van der Poorten and Stein [40], 2008, pp. 282–295. MR 2009j:11200

[3] Gaetan Bisson, *Endomorphism rings in cryptography*, Ph.D. thesis, Technische Universiteit Eindhoven and Institut National Polytechnique de Lorraine, 2011. http://repository.tue.nl/714676

[4] Gaetan Bisson, Romain Cosset, and Damien Robert, *AVIsogenies, a library for computing isogenies between abelian varieties*, 2012. http://avisogenies.gforge.inria.fr

[5] Bradley W. Brock and Andrew Granville, *More points than expected on curves over finite field extensions*, Finite Fields Appl. **7** (2001), no. 1, 70–91. MR 2002d:11070

[6] Reinier Bröker, David Gruenewald, and Kristin Lauter, *Explicit CM theory for level* 2*-structures on abelian surfaces*, Algebra Number Theory **5** (2011), no. 4, 495–528. MR 2870099

[7] Jan Hendrik Bruinier and Tonghai Yang, *CM-values of Hilbert modular functions*, Invent. Math. **163** (2006), no. 2, 229–288. MR 2008b:11053

[8] Gabriel Cardona and Jordi Quer, *Field of moduli and field of definition for curves of genus* 2, in Shaska [35], 2005, pp. 71–83. MR 2006h:14036

[9] Robert Carls, David Kohel, and David Lubicz, *Higher-dimensional* 3*-adic CM construction*, J. Algebra **319** (2008), no. 3, 971–1006. MR 2010e:14042

[10] Robert Carls and David Lubicz, *A p-adic quasi-quadratic time point counting algorithm*, Int. Math. Res. Not. **2009** (2009), no. 4, 698–735. MR 2010c:14020

[11] Jean Chaumine, James Hirschfeld, and Robert Rolland (eds.), *Algebraic geometry and its applications*: *Proceedings of the* 1*st Symposium* (*SAGA*) *held in Papeete*, *May* 7–11, 2007, Series on Number Theory and its Applications, no. 5, Hackensack, NJ, World Scientific, 2008. MR 2009h:14003

[12] Alina-Carmen Cojocaru, Kristin Lauter, Rachel Pries, and Renate Scheidler (eds.), *WIN—women in numbers*: *Research directions in number theory*, *including the proceedings of the Banff International Research Station* (*BIRS*) *Workshop held in Banff*, *AB*, *November* 2–7, 2008, Fields Institute Communications, no. 60, American Mathematical Society, Providence, RI, 2011. MR 2012g:11005

[13] Romain Cosset and Damien Robert, *Computing* (ℓ, ℓ)*-isogenies in polynomial time on Jacobians of genus* 2 *curves*, Cryptology ePrint Archive, Report 2011/143, 2011. http://eprint.iacr.org/2011/143

[14] J.-M. Couveignes, *Linearizing torsion classes in the Picard group of algebraic curves over finite fields*, J. Algebra **321** (2009), no. 8, 2085–2118. MR 2010e:14019

[15] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude*, Ark. Mat. Astr. Fys. **22A** (1930), no. 10, 1–14.

[16] Kirsten Eisenträger and Kristin Lauter, *A CRT algorithm for constructing genus* 2 *curves over finite fields*, in Rodier and Vladut [34], 2010, pp. 161–176, preprint version at arXiv:math/0405305 [math.NT]. MR 2856565

[17] Andreas Enge and Andrew V. Sutherland, *Class invariants by the CRT method*, in Hanrot et al. [23], 2010, pp. 142–156. MR 2012d:11246

[18] David Freeman and Kristin Lauter, *Computing endomorphism rings of Jacobians of genus* 2 *curves over finite fields*, in Chaumine et al. [11], 2008, pp. 29–66. MR 2010a:14042

[19] A. Fröhlich (ed.), *Algebraic number fields*: *L-functions and Galois properties*, Academic Press, London, 1977. MR 55 #10416

[20] P. Gaudry, T. Houtmann, D. Kohel, C. Ritzenthaler, and A. Weng, *The* 2*-adic CM method for genus* 2 *curves with application to cryptography*, in Lai and Chen [25], 2006, pp. 114–129. MR 2009j:94110

[21] Eyal Z. Goren and Kristin E. Lauter, *Genus* 2 *curves with complex multiplication*, Int. Math. Res. Not. **2012** (2012), no. 5, 1068–1142. MR 2899960

[22] Helen Grundman, Jennifer Johnson-Leung, Kristin Lauter, Adriana Salerno, Bianca Viray, and Erika Wittenborn, *Igusa class polynomials*, *embeddings of quartic CM fields*, *and arithmetic intersection theory*, in Cojocaru et al. [12], 2011, pp. 35–60. MR 2777799

[23] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory*: *Proceedings of the* 9*th Biennial International Symposium* (*ANTS-IX*) *held in Nancy*, *July* 19–23, 2010, Lecture Notes in Computer Science, no. 6197, Berlin, Springer, 2010. MR 2011g:11002

[24] J. C. Lagarias and A. M. Odlyzko, *Effective versions of the Chebotarev density theorem*, in Fröhlich [19], 1977, pp. 409–464. MR 56 #5506

[25] Xuejia Lai and Kefei Chen (eds.), *Advances in cryptology—ASIACRYPT* 2006: *Proceedings of the* 12*th International Conference on the Theory and Application of Cryptology and Information Security held in Shanghai*, *December* 3–7, 2006, Lecture Notes in Computer Science, no. 4284, Berlin, Springer, 2006. MR 2009e:94091

[26] Kristin Lauter and Bianca Viray, *An arithmetic intersection formula for denominators of Igusa class polynomials*, 2012. arXiv 1210.7841 [math.NT]

[27] Kristin Lauter and Tonghai Yang, *Computing genus* 2 *curves from invariants on the Hilbert moduli space*, J. Number Theory **131** (2011), no. 5, 936–958. MR 2012e:11110

[28] Kristin E. Lauter and Damien Robert, *Improved CRT algorithm for class polynomials in genus* 2, Cryptology ePrint Archive, Report 2012/443, 2012. http://eprint.iacr.org/2012/443

[29] H. W. Lenstra, Jr., J. Pila, and Carl Pomerance, *A hyperelliptic smoothness test*, *II*, Proc. London Math. Soc. (3) **84** (2002), no. 1, 105–146. MR 2003f:11190

[30] David Lubicz and Damien Robert, *Computing isogenies between abelian varieties*, Compos. Math. **148** (2012), no. 5, 1483–1515. MR 2982438

[31] Jean-François Mestre, *Construction de courbes de genre* 2 *à partir de leurs modules*, in Mora and Traverso [32], 1991, pp. 313–334. MR 92g:14022

[32] Teo Mora and Carlo Traverso (eds.), *Effective methods in algebraic geometry*: *Papers from the symposium* (*MEGA*-90) *held in Castiglioncello*, *April* 17–21, 1990, Progress in Mathematics, no. 94, Birkhäuser, Boston, 1991. MR 91m:14003

[33] Damien Robert, *Fonctions thêta et applications à la cryptographie*, Ph.D. thesis, Université Henri Poincaré — Nancy 1, 2010. http://hal.inria.fr/tel-00528942/

[34] François Rodier and Serge Vladut (eds.), *Arithmetics*, *geometry*, *and coding theory* (*AGCT* 2005): *Papers from the conference held in Marseilles*, *September* 26–30, 2005, Séminaires et Congrès, no. 21, Société Mathématique de France, Paris, 2010. MR 2012h:14002

[35] Tanush Shaska (ed.), *Computational aspects of algebraic curves*: *Papers from the conference held at the University of Idaho*, *Moscow*, *ID*, *May* 26–28, 2005, Lecture Notes Series on Computing, no. 13, World Scientific, Hackensack, NJ, 2005. MR 2006e:14003

[36] Goro Shimura, *Abelian varieties with complex multiplication and modular functions*, Princeton Mathematical Series, no. 46, Princeton University Press, 1998. MR 99e:11076

[37] Anne-Monika Spallek, *Kurven vom Geschlecht* 2 *und ihre Anwendung in Public-Key-Kryptosystemen*, Ph.D. thesis, Universität Gesamthochschule Essen, 1994. http://www.iem.uni-due.de/zahlentheorie/AES-KG2.pdf

[38] Theodorus Cornelis Streng, *Complex multiplication of abelian surfaces*, Ph.D. thesis, Universiteit Leiden, 2010. http://www.math.leidenuniv.nl/scripties/thesisStreng.pdf

[39] Andrew V. Sutherland, *Computing Hilbert class polynomials with the Chinese remainder theorem*, Math. Comp. **80** (2011), no. 273, 501–538. MR 2011k:11177

[40] Alfred J. van der Poorten and Andreas Stein (eds.), *Algorithmic number theory*: *Proceedings of the* 8*th International Symposium* (*ANTS-VIII*) *held in Banff*, *AB*, *May* 17–22, 2008, Lecture Notes in Computer Science, no. 5011, Berlin, Springer, 2008. MR 2009h:11002

[41] Paul van Wamelen, *Examples of genus two CM curves defined over the rationals*, Math. Comp. **68** (1999), no. 225, 307–320. MR 99c:11079

[42] Annegret Weng, *Constructing hyperelliptic curves of genus* 2 *suitable for cryptography*, Math. Comp. **72** (2003), no. 241, 435–458. MR 2003i:14029

[43] Tonghai Yang, *An arithmetic intersection formula on Hilbert modular surfaces*, Amer. J. Math. **132** (2010), no. 5, 1275–1309. MR 2012a:11078

[44] ———, *Arithmetic intersection on a Hilbert modular surface and the Faltings height*, 2010. arXiv 1008.1854 [math.NT]

KRISTIN E. LAUTER: klauter@microsoft.com
Cryptography Research Group, Microsoft Research, One Microsoft Way, Redmond, WA 98052, United States

DAMIEN ROBERT: damien.robert@inria.fr
Microsoft Research, One Microsoft Way, Redmond, WA 98052, United States
Current address: *INRIA Bordeaux Sud-Ouest, 200 avenue de la Vieille Tour, 33405 Talence cedex, France*

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.463

msp

Fast computation of isomorphisms of hyperelliptic curves and explicit Galois descent

Reynald Lercier, Christophe Ritzenthaler, and Jeroen Sijsling

We show how to speed up the computation of isomorphisms of hyperelliptic curves by using covariants. We also obtain new theoretical and practical results concerning models of these curves over their field of moduli.

1. Introduction

Let X_1 and X_2 be two curves of genus $g \geq 2$ over a field k. We wish to quickly determine the (possibly empty) set of isomorphisms between them. The standard strategy mainly consists of interpolating the isomorphisms at Weierstrass or small degree places, depending on whether the characteristic of the field is zero or positive [17]. This yields algorithms of complexity at least $O(g^6)$ in general, and at least $O(g^2)$ even in very favorable cases.

In this article we restrict to hyperelliptic curves with equations $X_i : y^2 = f_i(x)$ over a field k of characteristic different from 2. The issue can then be rephrased in terms of isomorphisms of degree $2g + 2$ polynomials under the Möbius action of $\mathrm{GL}_2(k)$ (see Section 2E1). Our first contribution is to show how to compute the set of isomorphisms in a much faster way by combining two new ideas. The first one uses the factorization of the Möbius action into a diagonal matrix times a second matrix whose diagonal coefficients are equal to 1. This idea allows us to perform the computation of the isomorphisms with only univariate polynomial calculations (see Section 2B). The second idea relies on a classical generalization of invariants, called *covariants* (see Section 2C). Using covariants, we can reduce our search for an isomorphism between f_1 and f_2 to the search for an isomorphism between polynomials of lower degree. This gives us an algorithm for generic hyperelliptic curves whose complexity is quasilinear in g (see Section 2D). In

MSC2010: primary 13A50; secondary 14Q05, 14H10, 14H25.
Keywords: invariants, covariants, hyperelliptic curves, binary forms, Galois descent, isomorphism, moduli, algorithm.

the genus-2 and genus-3 cases, we analyze the small locus of curves where our strategy fails (see Section 2E2). The use of covariants was inspired by work of van Rijnswou [30], who used covariants, along with a miraculous isomorphism from representation theory, to generically reduce the isomorphism question for ternary quartics to that for binary quartics.

In a related direction, thanks to covariants, we get both theoretical and practical results on Galois descent of hyperelliptic curves over their field of moduli. As the terminology suggests, this issue is related to moduli spaces, namely as follows.

The use of invariants allows the construction of the coarse moduli space of smooth curves admitting a suitable representation (for example, hyperelliptic or planar) as a geometric quotient in the sense of Mumford [28]. Such quotients have been calculated explicitly; for instance, for genus-2 and genus-3 hyperelliptic curves, see [21; 34]. Given a field k, the k-points of these quotients correspond to curves whose field of moduli, in the sense of Definition 3.1, is equal to k (up to a possible purely inseparable extension). This statement is probably well-known, but we could not find it in the literature; therefore, we give the link between these two definitions in Section 3.

A natural question is to determine when a curve descends to its field of moduli, that is, when its field of moduli is also a field of definition (and hence the smallest possible field of definition, under inclusion). Examples of curves that do not so descend were constructed by Shimura [33] and Earle [12], among others. However, curves of genus at most 1 always descend to their field of moduli, and models over the field of moduli can be explicitly constructed. Moreover, in the genus-2 case, although an obstruction to the descent may exist, as is shown in [26] and [7], the question of explicit descent to the field of moduli is solved. One of our aims is to obtain similar results in the general hyperelliptic case.

Many theoretical results for the general case can be found in [18]. In practice, though, computing an explicit model of a given curve over its field of moduli can be a very hard task, as we explain in Section 3B1. Indeed, for a given *finite* Galois extension, Weil's criterion in [35] often leads to a computational answer; the main difficulty in our context is to work out the finite Galois extension over which a descent isomorphism is defined. As far as we know, there is no easy general way to find this extension, except when k is finite or when the geometric automorphism group of the curve is trivial. Moreover, for hyperelliptic curves there is a refinement of the descent question — namely, to ask for a descent to a model of the form $y^2 = f(x)$ — and this introduces additional difficulties.

The "magic" of the covariant method is to reduce the descent problem to lower genus, where a solution may be easier to determine (Theorem 3.8). In the genus-1 case, for example, there is always an explicit model over the field of moduli and we can quickly determine a descent isomorphism to this model, thanks to the first

part of our work. It turns out that in suitable cases, this descent induces a descent of the original hyperelliptic curve to its field of moduli.

We illustrate this descent to the field of moduli for genus-3 hyperelliptic curves with automorphism group $(\mathbb{Z}/2\mathbb{Z})^3$, a case which remained unsolved in [25]; see Section 3C1. We also look at the case of genus-3 hyperelliptic curves with automorphism group $(\mathbb{Z}/2\mathbb{Z})^2$; in this case the field of moduli is not always a field of definition, and we prove that we can always find a model over an at most quadratic extension of the field of moduli. Finally, in Section 3D we show that our method can be used to descend families of curves with the example of a 3-dimensional family of genus-5 hyperelliptic curves from [13].

We stress that we are merely beginning to exploit the full strength of these new ideas. An article on nonhyperelliptic curves is in progress. We are also developing a general version of van Rijnswou's algorithms that is much more effective over finite fields and number fields. Finally, we seek to obtain new theoretical and practical descent results by analyzing the influence of twists on covariants.

We have implemented our algorithms in Magma [3]; the resulting programs, together with other useful scripts and output that was too large to include in this paper, may be found at

http://perso.univ-rennes1.fr/christophe.ritzenthaler/programme/hyp-desc.tgz

Notation. In the following, k denotes a field of characteristic p (prime or 0) with algebraic closure K. Hyperelliptic curves are additionally assumed to be smooth, so that when a singular affine model of a curve is given, we actually consider its desingularization. Unless noted otherwise, (iso)morphisms are defined over the base field k. We use the following notation for groups: $\boldsymbol{C}_n = \mathbb{Z}/n\mathbb{Z}$; $\boldsymbol{D}_{2n}$ is the dihedral group with $2n$ elements; $\boldsymbol{U}_6$ is the group with 24 elements defined by $\langle S, T\rangle$ with $S^{12} = T^2 = 1$ and $TST = S^5$; $\boldsymbol{V}_8$ is the group with 32 elements defined by $\langle S, T\rangle$ with $S^4 = T^8 = (ST)^2 = (S^{-1}T)^2 = 1$; $\boldsymbol{S}_n$ is the symmetric group over n symbols. Finally, if f_1 and f_2 are polynomials or matrices or some other such objects over a field k, we will write $f_1 \sim f_2$ if there exists $\lambda \in k^*$ such that $f_1 = \lambda \cdot f_2$.

2. Isomorphisms between forms and hyperelliptic curves

2A. ***Isomorphisms of binary forms.*** Let $n \geq 1$ be an integer, let $V = k^2$ be the k-vector space with basis (x, z), and let $S^n(V)$ be the $(n+1)$-dimensional vector space of homogeneous forms $\sum_{i=0}^{n} a_i x^i z^{n-i}$ of degree n in (x, z). In the sequel, we call an element of $S^n(V)$ a (*binary*) *form*. When $n = 0$, we let $S^0(V) = k$. Let G be a subgroup of $\mathrm{GL}_2(k)$ and let M be an element of G. If f is a form in $S^n(V)$, we define $M.f$ by $(M.f)(x, z) = f(M^{-1}(x, z))$, where the action of a matrix on (x, z) is the standard action on ${}^t(x, z)$.

Definition 2.1. Let f_1, f_2 be forms of degree $n \geq 1$ over a field k. We denote by $\mathrm{Isom}(f_1, f_2) \subset \mathrm{PGL}_2(k)$ the set of matrices M up to scalar equivalence such that $M.f_1 \sim f_2$. Additionally, we write $\mathrm{Aut}\, f_1$ for $\mathrm{Isom}(f_1, f_1)$.

If $\mathrm{Isom}(f_1, f_2) \neq \varnothing$, this set is a principal homogeneous space over $\mathrm{Aut}\, f_1$. In particular, $\mathrm{Isom}(f_1, f_2) = M\ \mathrm{Aut}\, f_1$ for any $M \in \mathrm{Isom}(f_1, f_2)$.

Let f be a form of degree n over k. Over K, we can write $f = \prod_{i=1}^{s} (\alpha_i x - \beta_i z)^{n_i}$, where $(\alpha_i, \beta_i) \in K^2 \setminus \{(0,0)\}$ and $n_i \in \mathbb{N}$. We associate to such a form its squarefree part $\widetilde{f} = \prod_{i=1}^{s} (\alpha_i x - \beta_i z)$, which is defined up to a multiplicative constant. The action of M on f reflects the classical Möbius action of $\mathrm{PGL}_2(K)$ on the roots $(\alpha_i : \beta_i) \in \mathbb{P}^1_K$ of f. In particular, two forms of the same degree are K-isomorphic if and only if there exists an $M \in \mathrm{GL}_2(K)$ mapping the roots of the first form to the roots of the second form (counting multiplicities). Hence we have:

Lemma 2.2. *The group* $\mathrm{Aut}_K f$ *is finite if and only if* $s \geq 3$, *that is, if and only if* $\deg \widetilde{f} \geq 3$. *Moreover,* $\mathrm{Aut}_K f \subset \mathrm{Aut}_K \widetilde{f}$.

2B. ***The direct approach.*** The classical method to compute isomorphisms between two binary forms f_1, f_2 of degree n over a field k is to find a $\mathrm{PGL}_2(k)$-transformation of $\mathbb{P}^1$ which maps the roots of the first form to the root of the second form. The most time-consuming task is to compute an isomorphism between the splitting fields of f_1 and f_2. Even in the most favorable case, where k is a finite field, the fastest algorithms need at least $O(n^{2.5+o(1)})$ operations in k (see [22]).

We show here that it is actually possible to get rid of this cumbersome ring isomorphism computation, and describe an algorithm of time complexity only quasilinear in n. This algorithm takes as input binary forms $f_1 = \sum_i A_i x^i z^{n-i}$ and $f_2 = \sum_i B_i x^i z^{n-i}$ of equal degree $n \geq 3$, each having at least three distinct roots. It returns matrices representing the elements of $\mathrm{Isom}(f_1, f_2)$.

First, we suppose that the coefficient A_{n-1} is equal to zero. Note that this is typically not a big restriction, since we may apply linear transformations to f_1. A notable exception is when p divides n. We therefore assume that p is prime to n.

Second, we note that determining $\mathrm{Isom}(f_1, f_2)$ is equivalent to determining the matrices $M = (m_{i,j}) \in \mathrm{GL}_2(k)$ such that

$$f_2(m_{11}x + m_{12}z, m_{21}x + m_{22}z) = \lambda f_1(x, z) \quad \text{for some } \lambda \in k^*. \tag{1}$$

Third, because of homogeneity, we may suppose that the λ in (1) equals 1, after enlarging k by a radical extension if necessary. Note that though this radical extension is *a priori* unknown, the details of the algorithm below will show how it can be determined.

Finally, we may suppose that the M in (1) are of the form

$$M = \begin{bmatrix} 1/\alpha & \beta/\delta \\ \gamma/\alpha & 1/\delta \end{bmatrix}.$$

Of course this may not be true, because a zero may occur on the diagonal of one of these M. However, one can fix this situation by applying a suitable change of variables to f_2.

The equation $f_2(m_{11}x + m_{12}z, m_{21}x + m_{22}z) = f_1(x, z)$ now becomes

$$f_2(x + \beta z, \gamma x + z) = f_1(\alpha x, \delta z).$$

Equating the coefficients of x^n in both sides of this equation yields $A_n\alpha^n = f_2(1, \gamma)$, and we can write α^n in terms of γ. Similarly, the equality of the coefficients of $x^{n-1}z$,

$$\beta\frac{\partial f_2}{\partial x}(1, \gamma) + \frac{\partial f_2}{\partial z}(1, \gamma) = 0,$$

enables us to write β in term of γ too. More generally, equating the coefficients of $x^{n-i}z^i$ for $i = 2, \ldots, n$, where we substitute α^n and β in term of γ, yields $n-1$ equations of the form

$$A_n\left(\sum_{j=0}^{i}\binom{i}{j}\left(-\frac{\partial f_2}{\partial z}\right)^j\left(\frac{\partial f_2}{\partial x}\right)^{i-j}\frac{\partial^i f_2}{\partial x^j\,\partial z^{i-j}}\right)(1, \gamma) = i!\left(\frac{\partial f_2}{\partial x}(1, \gamma)\right)^i\left(\frac{\delta}{\alpha}\right)^i f_2(1, \gamma). \quad (2)$$

Note that the left-hand side of (2) is actually a polynomial multiple of $f_2(x, z)$, and we can divide both sides by $f_2(1, \gamma)$ — see [16, Chapter 1, §§15–16] for an elegant explanation. This yields equations of degree $i(n-2)$ in γ for the left side and of degree $i(n-1)$ in γ and degree i in δ/α on the right side.

Now, dividing the square of (2) specialized at $i = 3$ by the cube of (2) specialized at $i = 2$ allows to eliminate, up to some constant, the right-hand side of these equations, in particular the unknown δ/α. We end up with an equation of degree $6(n-2)$ in γ. Similarly, when $n > 3$, dividing (2) specialized at $i = 4$ by the square of (2) specialized at $i = 2$ yields an equation of degree $4(n-2)$ in γ. Taking the gcd, we obtain a polynomial of low degree with root γ. Generically, this gcd is of degree 1.

Under the assumptions made, the algorithm is therefore straightforward. For each possible γ, we compute α, β and δ and check whether the resulting matrix is in $\mathrm{Isom}(f_1, f_2)$.

The computations involved in this algorithm (taking gcds of polynomials of degree $O(n)$, taking n-th roots, and so forth) are all of time complexity quasilinear in n.

We have implemented the algorithm in Magma (version 2.18-2) and have timed the resulting procedure, `IsGL2EquivFast`, on a laptop (based on an Intel Core i7 M620 2.67GHz processor) for irreducible forms of increasing degree, the most

	Computations over $\mathbb{F}_{10007}$			Computations over $\mathbb{Q}$		
Genus	Old	Section 2B	Section 2D	Old	Section 2B	Section 2D
1	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.4	0.0	0.0
8	0.0	0.0	0.0	15	0.0	0.0
16	0.1	0.0	0.0	1150	0.1	0.0
32	0.2	0.0	0.0	—	0.2	0.0
64	0.9	0.1	0.0	—	0.6	0.0
128	6.5	0.6	0.0	—	3	0.2
256	39	3.7	0.1	—	30	0.6
512	242	25	0.5	—	382	3.4
1024	1560	165	2.5	—	5850	7

Table 1. Timings (in seconds) for isomorphisms between forms of degree $2g+2$, over $\mathbb{F}_{10007}$ and over $\mathbb{Q}$. The columns labeled "Old" give timings for Magma's built-in function `IsGL2Equivalent`; the columns labeled "Section 2B" give timings for the function `IsGL2EquivFast` described in Section 2B; and the columns labeled "Section 2D" give timings for the function `IsGL2EquivCovariant` described in Section 2D. Entries of "—" indicate computations that were aborted after an hour.

favorable case for the native Magma routine `IsGL2Equivalent`. We compare with `IsGL2Equivalent`, which implements the classical method, first over the finite field $\mathbb{F}_{10007}$, then over the rationals with coefficients bounded by ± 2. The results are in Table 1. (See Section 2D for the definition of `IsGL2EquivCovariant`.)

As concluding remarks, we note first of all that this algorithm is equally suitable for determining K-isomorphisms. Moreover, in the special case of binary quartics, it is just as efficient as the algorithm given in [8].

2C. ***The covariant approach.*** Let k be an infinite field of characteristic p and let $n > 1$ be an integer.

Definition 2.3. Let $r \geq 0$ be an integer. A homogeneous polynomial function $C : S^n(V) \to S^r(V)$ of degree d is a *covariant* if there exists $\omega \in \mathbb{Z}$ such that, for all $M \in G$ and all $f \in S^n(V)$, we have

$$C(M.f) = (\det M)^{-\omega} \cdot M.C(f).$$

When $r = 0$, such a C is called a (relative) *invariant* and is denoted by I.

The integer r is called the *order* of the covariant. If $nd - r$ is odd, the covariant is necessarily zero. Otherwise the integer ω is unique, and is called the *weight* of the covariant. It is equal to $(nd - r)/2$. In the sequel, we often identify C

with $C(f)$ for a general form $f \in F(a_0, \ldots, a_n)[x, z]$, where F is the prime field of k. For instance, the identity function $S^n(V) \to S^n(V)$ is a covariant of degree 1 and order n that we identify with f itself.

Remark 2.4. The determinant factor prevents the addition of covariants of different weights when $G = \mathrm{GL}_2(K)$. Hence one generally studies the graded algebra $\mathscr{C}_n$ of covariants and $\mathscr{I}_n$ of invariants under the action of $\mathrm{SL}_2(K)$. It is easy to see that the homogeneous elements of $\mathscr{C}_n$ and $\mathscr{I}_n$ are actually all the covariants or invariants under the action of $\mathrm{GL}_2(K)$. Despite this ambiguity, in the rest of the article we work with $G = \mathrm{GL}_2(K)$ instead of $\mathrm{SL}_2(K)$ because, in practice, this choice often allows us to avoid a quadratic extension of k when looking for an isomorphism M between two forms.

There is a large literature on how to generate invariants and covariants starting from f. Gordan's algorithm [15] allows to find a set of generators for the algebras $\mathscr{C}_n$ and $\mathscr{I}_n$ thanks to the use of certain differential operators, called *h-transvectants* and defined as follows. Given two covariants C_1, C_2 of degree d_1, d_2 and of order r_1, r_2, and given an integer $h \geq 1$, we can create a new covariant denoted $(C_1, C_2)_h$ and usually defined as [29, p. 88]

$$\frac{(r_1-h)!(r_2-h)!}{r_1!r_2!} \sum_{i=0}^{h} (-1)^i \binom{h}{i} \frac{\partial^h C_1}{\partial x^{h-i}\partial z^i} \frac{\partial^h C_2}{\partial x^i \partial z^{h-i}}.$$

In practice, we use the univariate counterpart. Looking at C_1, C_2 as univariate polynomials in x/z, we get [29, Theorem 5.6]

$$h!\frac{(r_1-h)!(r_2-h)!}{r_1!r_2!} \sum_{i=0}^{h} (-1)^i \binom{r_1-i}{h-i}\binom{r_2-h+i}{i} \frac{d^{h-i}C_1}{dx^{h-i}} \frac{d^i C_2}{dx^i}. \tag{3}$$

Effective methods for computing sets of generators when $K = \mathbb{C}$ have been worked out for n up to 10 (see [11; 14; 10; 2; 34; 9; 5; 4]). It has been shown that, if $\mathbb{C}$ is replaced by an algebraically closed field K of characteristic p, these computations are still valid for $g = 2$ if $p \neq 2, 3, 5$ [24] and for $g = 3$ if $p \neq 2, 3, 5, 7$ [25].

Our second idea to compute isomorphisms between forms of a given degree is to reduce the question to smaller degree by using covariants. Indeed, the following observation is a simple consequence of the definition itself.

Proposition 2.5. *Let f_1, f_2 be forms of even degree n over a field k. Let C be a covariant of order r for binary forms of degree n, defined over the prime field of k, and let $c_i = C(f_i) \in S^r(V)$. Then* $\mathrm{Isom}(f_1, f_2) \subset \mathrm{Isom}(c_1, c_2)$. □

We illustrate this idea and study its limitations with the computation of isomorphisms for forms and hyperelliptic curves in Sections 2D and 2E. As we want the covariants c_i to have the smallest degree possible and $\mathrm{Isom}(c_1, c_2)$ to be finite,

we want that $\deg \widetilde{c}_i \geq 3$. Actually, in what follows we mostly deal with forms of even degree, so nonzero covariants will be of even order, and the smallest degree meeting our restriction is then 4.

Consider a binary quartic $q = a_4x^4 + a_3x^3z + a_2x^2z^2 + a_1xz^3 + a_0z^4$ over k with $p \neq 2, 3$. We define

$$\begin{aligned} I = I(q) &= 12a_4a_0 - 3a_3a_1 + a_2^2, \\ J = J(q) &= 72a_4a_2a_0 + 9a_3a_2a_1 - 27a_4a_1^2 - 27a_0a_3^2 - 2a_2^3 \end{aligned}$$

as in [8]. The form q has distinct roots if and only if $\Delta = 4I^3 - J^2 \neq 0$. Given $I, J \in K$ such that $\Delta \neq 0$, one can easily reconstruct a form with at least three distinct roots which is K-isomorphic to q. We can take

$$q = \begin{cases} x^3z - 27(I^3/J^2)xz^3 - 27(I^3/J^2)z^4 & \text{if } J \neq 0, \\ x^3z + xz^3 & \text{otherwise.} \end{cases} \tag{4}$$

Concerning the geometric automorphisms of binary quartics, we have the following easy result, for which we could not find a reference.

Proposition 2.6. *Let q be a binary quartic form over K, with invariants I and J. Suppose that $\Delta \neq 0$. Then*

$$\operatorname{Aut} q \cong \begin{cases} \boldsymbol{A}_4 & \text{if } I = 0, \\ \boldsymbol{D}_8 & \text{if } J = 0, \\ \boldsymbol{D}_4 & \text{otherwise.} \end{cases} \tag{5}$$

Proof. Let $\Lambda \subset \mathbb{P}^1(K)$ be the set of four roots of q. Using the 3-transitivity of the action of $\mathrm{PGL}_2(K)$ on $\mathbb{P}^1(K)$, we may assume that $\Lambda = \{0, 1, \infty, \lambda\}$ for some $\lambda \in K \setminus \{0, 1\}$. Then the transformation $x \mapsto \lambda/x$ induces the permutation $(0\infty)(1\lambda)$ of Λ. By symmetry, we see that $\operatorname{Stab}\Lambda \subset \operatorname{Sym}\Lambda$ contains the Viergruppe $\boldsymbol{D}_4 \subset \operatorname{Sym}\Lambda$.

We are reduced to analyzing the case when $\operatorname{Stab}\Lambda$ properly contains $\boldsymbol{D}_4$. Since the extension $1 \to \boldsymbol{D}_4 \to \boldsymbol{S}_4 \to \boldsymbol{S}_3 \to 1$ is split and all subgroups of $\boldsymbol{S}_3$ of equal order are conjugate, this is in turn equivalent to determining when $\operatorname{Stab}\Lambda$ contains an additional given 2- or 3-cycle. These cases give rise to the exceptional groups in (5) of order 8 and 12.

First let us see for which λ the permutation (1λ) is in $\operatorname{Stab}\Lambda$. In this case, the fractional linear transformation fixes 0 and ∞ and is therefore of the form $x \mapsto cx$. This only gives a new automorphism if $c = -1$, so $\lambda = -1$ and $J = 0$.

In the case where the permutation (01λ) is in $\operatorname{Stab}\Lambda$, a slightly more involved calculation gives that $\lambda = \zeta_3 + 1$ for a primitive third root of unity ζ_3, and in that case $I = 0$. □

We will also need in the sequel the following result.

Proposition 2.7. *Let q be a binary quartic form defined over k with distinct roots, and let $\mathfrak{q}$ be the form defined by* (4)*. Assume that $I(q) \neq 0$ and $J(q) \neq 0$. Then a K-isomorphism between q and $\mathfrak{q} = z(x^3 + b_1xz^2 + b_0z^3)$ is defined over any extension of k where q has a root.*

Proof. Let k' be an extension of k where q has a root. By a change of variable defined over k', we can map this root to infinity and hence q onto $q' = zr$, where $r = x^3 + a_1xz^2 + a_0z^3 \in k'[x, z]$. Now, since

$$I(q') = -a_1/4, \quad I(\mathfrak{q}) = -b_1/4,$$
$$J(q') = -a_0/16, \quad J(\mathfrak{q}) = -b_0/16,$$

we get the relation $a_1^3/a_0^2 = b_1^3/b_0^2$. Hence if we define $\lambda \in k'$ by

$$\lambda = \frac{J(q')I(\mathfrak{q})}{J(\mathfrak{q})I(q')},$$

the k'-isomorphism $M : (x, z) \mapsto (\lambda x, z)$ maps q' onto $\mathfrak{q}$. □

2D. ***Generic forms of even degree.*** We now describe an algorithm, based on the ideas of Sections 2B and 2C, to compute the isomorphisms between two generic binary forms f_1 and f_2. Our notation is as in Section 2B.

Algorithm 2.8 (`IsGL2EquivCovariant`).

Input: Two forms f_1 and f_2 of the same degree $n \geq 3$ over k, and integer parameters $B_{\text{order}} \geq 3$, $B_{\text{degree}} \geq 2$, and $B_{\text{singular}} \geq 0$.

Output: The matrices $M = (m_{i,j})_{i,j}$ in $\mathrm{PGL}_2(k)$ such that $M.f_1 \sim f_2$.

1. *Order loop*. For o increasing from 3 to B_{order} do:
   (a) *Degree loop*. For d increasing from 2 to B_{degree} do:
      i. Compute a random covariant C of order o and degree d using transvectants.
      ii. If $\widetilde{C}(f_1)$ is of degree at least 3, then compute $\mathrm{Isom}(\widetilde{C}(f_1), \widetilde{C}(f_2))$ and return the elements which induce isomorphisms between f_1 and f_2.
      iii. Otherwise, repeat the following procedure B_{singular} times:
         – Compute a new random covariant C' of order o and degree d using transvectants, and replace C by the covariant $C + \kappa C'$ for some random κ in the field k.
         – If $\widetilde{C}(f_1)$ is of degree at least 3, compute $\mathrm{Isom}(\widetilde{C}(f_1), \widetilde{C}(f_2))$ and return the elements that induce isomorphisms between f_1 and f_2.
2. *Failure*. Return the result of `IsGL2EquivFast`(f_1, f_2).

For the purpose of computing random covariants, we follow Gordan [15]. Given an order o and a degree d, we construct recursively a covariant $C = \left(\prod C_{d',o'}, f\right)_h$ as a transvectant of some level h of the form f and a product of covariants of intermediate orders o' and degrees d', under the two constraints $d = \sum d'$ and $o = n + \sum o' - 2h$.

When n is even, the transvectant of smallest order and degree is $C_{2,4} = (f, f)_{n-2}$. The next simplest transvectant is $C_{3,4} = ((f, f)_{n/2}, f)_{n-2}$, of order 4 and degree 3. For large orders and degrees, covariants must be computed "on the fly", specialized for f_1 and f_2, since expressions are far too large to be precomputed.

To completely specify the algorithm, we have to be more precise about how to compute covariants and how to choose the loop bounds B_{order}, B_{degree} and B_{singular}. A straightforward choice for the loop bounds is $B_{\text{order}} = 4$, $B_{\text{degree}} = 2$, and $B_{\text{singular}} = 0$. With this choice, only the covariant $C_{2,4} = (f, f)_{n-2}$ is tested for n even, and when it turns out that the discriminant of this covariant vanishes, we go back to the method `IsGL2EquivFast`. First note that the covariant $(f, f)_{n-2}$ can be easily computed. Using (3), we find that we can write

$$\frac{(n!)^2}{(n-2)!}(f, f)_{n-2} = c_4 x^4 + c_3 x^3 z + c_2 x^2 z^2 + c_1 x z^3 + c_0 z^4, \tag{6}$$

where the coefficients c_i are given by

$$c_0 = \sum_{k=0}^{n-2} (-1)^k (n-k)!(k+2)!\, a_{n-2-k} a_k,$$

$$c_1 = \sum_{k=0}^{n-2} (-1)^k (n-k)!(k+2)!\big((n-1-k) a_{n-1-k} a_k + (k+1) a_{n-2-k} a_{k+1}\big),$$

$$c_2 = \frac{1}{2} \sum_{k=0}^{n-2} (-1)^k (n-k)!(k+2)!\big((k+2)(k+1) a_{k+2} a_{n-2-k} + 2(n-1-k)(k+1) a_{k+1} a_{n-1-k} + (n-k)(n-1-k) a_k a_{n-k}\big),$$

$$c_3 = \sum_{k=0}^{n-2} (-1)^k (n-k)!(k+2)!\big((n-1-k) a_{n-k} a_{k+1} + (k+1) a_{n-1-k} a_{k+2}\big),$$

$$c_4 = \sum_{k=0}^{n-2} (-1)^k (n-k)!(k+2)!\, a_{n-k} a_{k+2}.$$

Moreover, this setting is a good option for generic forms, as the following proposition shows.

Proposition 2.9. *Let $n \geq 6$ be an even integer and $p \neq 2, 3$. Let f be a generic binary form of degree n over k. Then the discriminant of $C_{2,4}(f)$ is nonzero.*

Proof. It is enough to find a single form f of degree n for which $C_{2,4}(f)$ has nonzero discriminant. First let us suppose that p is coprime to

$$n(n-2)(n-3)(n^2+3n+6).$$

We then take $f = x^n + x^{n-1}z - xz^{n-1} - z^n$. Note that this form is in fact nonsingular because $f = (x+z)(x^{n-1} - z^{n-1})$. We have that

$$-C_{2,4}(f) = \frac{4}{n}x^3z + \frac{2(n^2-n+6)}{n^2}x^2z + \frac{4}{n}xz^2.$$

This form has discriminant equal to $64(n-3)(n-2)(n^2+3n+6)/n^6$, which is nonzero by hypothesis.

One calculates similarly that for the other values of $p \neq 2, 3, 5$, one can use the form $x^n + x^{n-1}z + xz^{n-1} - z^n$ instead. Indeed, under these hypotheses on p the numerator $n^4 + 2n^3 + 5n^2 - 12n + 36$ of the resulting discriminant is coprime to the previous numerator. To finish the proof, $p = 5$ can be excluded using the form $x^n + x^{n-1}z + xz^{n-1} + 2z^n$. □

For nonrandom forms, especially forms of small degree with nontrivial automorphism group, it may be interesting to test other covariants than merely $C_{4,2}$. We then propose the following settings:

$$B_{\text{order}} = \min(8, n), \quad B_{\text{degree}} = 10, \quad \text{and} \quad B_{\text{singular}} = 10.$$

These bounds are constant in order to keep the total time complexity quasilinear in n. More precisely, the bound B_{order} is chosen to be at most 8 so as to take advantage of the classification work of [25], the bound B_{degree} is chosen to cover all the possible fundamental covariants of degree 8 and with order between 4 and 8 (see [25, Table 1, p. 607]), and the bound B_{singular} is chosen so as to increase the probability that our covariants, if singular, have distinct points of singularity (so that a linear combination may be nonsingular).

Remark 2.10. We may enter the last loop of the algorithm even if the form f has no geometric automorphisms. For example, this happens with the degree-8 form

$$x^7z + 7x^6z^2 + 7x^5z^3 + 8x^4z^4 + 2x^3z^5 + 10x^2z^6 + 9xz^7$$

over $k = \mathbb{F}_{11}$.

We have programmed Algorithm 2.8 in Magma (version 2.18-2), using the first setting of the parameters. In particular, we have implemented the covariant $C_{4,2}$ using (6), and we have measured the timings of the resulting procedure, `IsGL2EquivCovariant`, in the same experiments as in Section 2B. The results are presented in Table 1. As expected, computing isomorphisms is much faster with the help of covariants, even if the forms are split over k.

2E. *Application to isomorphisms of hyperelliptic curves.*

2E1. *Isomorphisms of forms and of hyperelliptic curves.* A curve X of genus $g \geq 1$ defined over k will be called *hyperelliptic* if X/K has a separable degree-2 map to $\mathbb{P}^1_K$. If $g > 1$, the curve X then has a unique involution ι, called the *hyperelliptic involution*, such that $Q = X/\langle\iota\rangle$ is of genus 0. This involution is in the center of $\mathrm{Aut}_K X$. We call $\overline{\mathrm{Aut}}_K X = (\mathrm{Aut}_K X)/\langle\iota\rangle$ the *reduced automorphism group* of X.

Let us assume from now on that $p \neq 2$. Then if Q has a rational point, X is birationally equivalent to an affine curve of the form $y^2 = f(x)$ for a separable polynomial f of degree $2g+1$ or $2g+2$. We say that f is a *hyperelliptic polynomial* and that X has a *hyperelliptic equation* if a curve in its isomorphism class (over k) can be written in the form above. We denote by X_f the curve associated to a hyperelliptic polynomial f. A hyperelliptic curve automatically has a hyperelliptic equation when k is algebraically closed or a finite field. However, for more general fields and curves of odd genus, this is not necessarily the case (see [25]).

By homogenizing to weighted projective coordinates of weight $(1, g+1, 1)$, we obtain an equation $y^2 = f(x, z)$. Here f is seen as a form of degree $2g+2$, taking into account a "root" at infinity when $\deg f = 2g+1$. With this convention, the roots of f are the ramification points of the cover X/Q. We will use these conventions for the roots and degree in the sequel when we speak about a hyperelliptic polynomial or the associated form.

If f_1 and f_2 are hyperelliptic polynomials of even degree $2g+2 \geq 6$, then isomorphisms between the hyperelliptic curves $y^2 = f_i(x, z)$ are represented by pairs (M, e) with

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{GL}_2(k)$$

and $e \in k^*$. To such a couple, one associates the isomorphism

$$(x, z, y) \mapsto (ax + bz, cx + dz, ey).$$

The representation is unique up to the equivalence $(M, e) \equiv (\lambda M, \lambda^{g+1} e)$ for $\lambda \in k^*$. Hence, if $M.f_1 = \mu \cdot f_2$ then the map

$$\mathrm{Isom}(f_1, f_2) \to (\mathrm{GL}_2(k) \times K^*)/\equiv, \quad M \mapsto (M, \pm\sqrt{\mu})$$

is well-defined up to the choice of a sign. It surjects onto $\mathrm{Isom}(X_{f_1}, X_{f_2})$, so knowing $\mathrm{Isom}(f_1, f_2)$ is enough to determine $\mathrm{Isom}(X_{f_1}, X_{f_2})$ "up to the hyperelliptic involution".

2E2. *Hyperelliptic curves of genus* 2 *and* 3. The covariant approach requires a covariant with at least three distinct roots, and hence it may fail in special cases, which we can specify for small genera. We give some details on the more difficult of the two cases: the genus-3 case. This problem is naturally stratified by the

$\mathrm{Aut}_K\, X_f$	$\overline{\mathrm{Aut}}_K\, X_f$	Normal models $X_f : y^2 = f$
$\boldsymbol{C}_2$	$\{1\}$	$f = x(x-1)(x^5 + ax^4 + bx^3 + cx^2 + dx + e)$
$\boldsymbol{D}_4$	$\boldsymbol{C}_2$	$f = x^8 + ax^6 + bx^4 + cx^2 + 1$ or $f = (x^2-1)(x^6 + ax^4 + bx^2 + c)$
$\boldsymbol{C}_4$	$\boldsymbol{C}_2$	$f = x(x^2-1)(x^4 + ax^2 + b)$
$\boldsymbol{C}_2^3$	$\boldsymbol{D}_4$	$f = (x^4 + ax^2 + 1)(x^4 + bx^2 + 1)$
$\boldsymbol{C}_2 \times \boldsymbol{C}_4$	$\boldsymbol{D}_4$	$f = (x^4-1)(x^4 + ax^2 + 1)$ or $f = x(x^2-1)(x^4 + ax^2 + 1)$
$\boldsymbol{D}_{12}$	$\boldsymbol{D}_6$	$f = x(x^6 + ax^3 + 1)$
$\boldsymbol{C}_2 \times \boldsymbol{D}_8$	$\boldsymbol{D}_8$	$f = x^8 + ax^4 + 1$
$\boldsymbol{C}_{14}$	$\boldsymbol{C}_7$	$f = x^7 - 1$
$\boldsymbol{U}_6$	$\boldsymbol{D}_{12}$	$f = x(x^6 - 1)$
$\boldsymbol{V}_8$	$\boldsymbol{D}_{16}$	$f = x^8 - 1$
$\boldsymbol{C}_2 \times \boldsymbol{S}_4$	$\boldsymbol{S}_4$	$f = x^8 + 14x^4 + 1$

Table 2. Automorphism groups of genus-3 hyperelliptic curves. For each automorphism group, we list the associated reduced automorphism group, together with normal model(s) for the generic hyperelliptic curve with that automorphism group. The notation for the groups is given at the end of the Introduction.

possible automorphism groups of the curve; we list these automorphism groups, together with normal models and inclusion relations between the strata, in Table 2 and Figure 1. We assume here that $p = 0$ or $p > 7$.

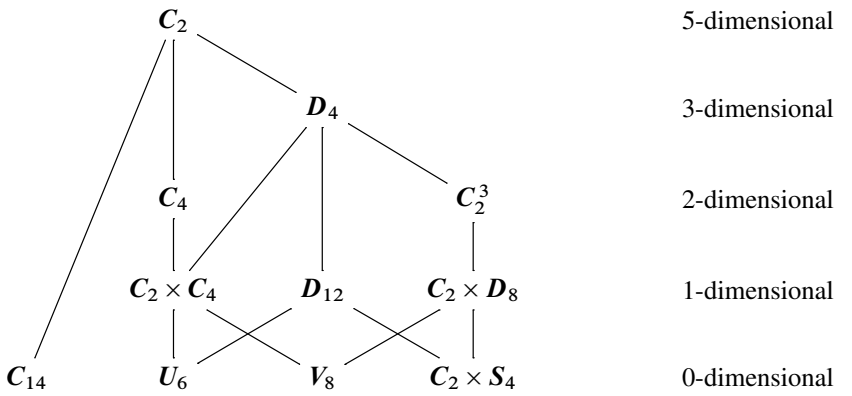


Figure 1. Dimensions and containment relationships among the moduli spaces of genus-3 hyperelliptic curves with given automorphism groups.

The moduli space of hyperelliptic curves of genus 3 is 5-dimensional, and can be explicitly described using the Shioda invariants $J_2, J_3, \ldots, J_{10}$ constructed in [34]. These invariants were used to speed up the calculations leading to the proof of the following proposition, which shows that the locus where the covariant method fails is of codimension 4 in the full moduli space. (The Magma parts of this proof, and of other proofs in this section, may be found at the URL listed in the Introduction.)

Proposition 2.11. *Let $X_f/K : y^2 = f(x)$ be a genus-3 hyperelliptic curve such that the form f cancels the discriminants of all its quartic covariants. Then* Aut X_f *contains either $\boldsymbol{D}_{12}$, $\boldsymbol{C}_2 \times \boldsymbol{D}_8$, or $\boldsymbol{C}_{14}$.*

Proof. Construct

$$C(f) \pm \kappa \cdot I(f) \cdot C'(f)$$

such that $\deg C = \deg I + \deg C'$, where C and C' run through the 14 fundamental quartic covariants given in [25, Table 1], where $I(f)$ equals either 1 or a Shioda invariant $J_i(f)$, and where κ runs through the integers between 0 and 10. We rewrite the discriminants of these covariants in terms of Shioda invariants and add to them the five Shioda relations [34, Theorem 3, p. 1042]. Using Magma, we have been able to compute a Gröbner basis of this polynomial system, over $\mathbb{Q}$, for the graded reverse lexicographical (grevlex) order $J_2 < J_3 < \cdots < J_{10}$ with weights 2, 3, ..., 10. Upon removing multiplicities, we obtain a basis with 22 polynomials, of total degree between 8 and 20. One then checks, using the stratum formulas from [25], that the irreducible components of the corresponding subscheme of the moduli space either correspond to families of forms with discriminant zero or to strata of curves X_f such that Aut X_f contain $\boldsymbol{D}_{12}$, $\boldsymbol{C}_2 \times \boldsymbol{D}_8$, or $\boldsymbol{C}_{14}$. □

We see from this that curves with automorphism group $\boldsymbol{D}_{12}$, $\boldsymbol{C}_2 \times \boldsymbol{D}_8$, or $\boldsymbol{C}_{14}$ cannot have separable quartic covariants. In these cases, using Proposition 2.5 and the normal models from Table 2, one can show:

- if Aut X is equal to $\boldsymbol{D}_{12}$ or $\boldsymbol{U}_6$ then the sextic covariant $C_{3,6} = ((f, f)_4, f)_5$ has nonzero discriminant;
- if Aut X contains $\boldsymbol{C}_2 \times \boldsymbol{D}_8$ or is equal to $\boldsymbol{C}_{14}$ then there is no order-4 or order-6 covariant with three distinct roots.

The number of covariants considered in the proof of Proposition 2.11 — namely, 1253 — is not minimal, but the redundancy helped Magma during the Gröbner basis computations. Nevertheless, similar computations show that we can easily reduce this number for curves with automorphism group larger than $\boldsymbol{C}_2$ (and also impose conditions on the automorphism groups of the covariants; see Sections 3B2

and 3C2). For example, consider the following five quartic covariants:

$$C_{2,4} = (f, f)_6, \qquad C_{4,4} = (((f, f)_4, f)_6, f)_4,$$
$$C_{3,4} = ((f, f)_4, f)_6, \quad C'_{4,4} = (((f, f)_4, f)_4, f)_6,$$
$$C_{5,4} = ((((f, f)_4, f)_6, f)_1, f)_7.$$

If X_f/K is a genus-3 hyperelliptic curve, we find that:

- If $\operatorname{Aut} X_f \cong \boldsymbol{D}_4$, one of the five covariants above has nonzero discriminant.
- If $\operatorname{Aut} X_f \cong \boldsymbol{C}_4$, one of $C_{2,4}$, $C_{3,4}$, $C_{4,4}$, and $C'_{4,4}$ has nonzero discriminant.
- If $\operatorname{Aut} X_f \cong \boldsymbol{C}_2^3$, one of $C_{2,4}$, $C_{3,4}$, and $C_{4,4}$ has nonzero discriminant.
- If $\operatorname{Aut} X_f \cong \boldsymbol{C}_2 \times \boldsymbol{C}_4$, the covariant $C_{3,4}$ has nonzero discriminant.

Remark 2.12. Similar conclusions hold for genus 2. Specifically, there is no quartic covariant with nonzero discriminant for the curves X_f/K such that $\boldsymbol{D}_{12} \subset \operatorname{Aut} X_f$ or $\operatorname{Aut} X_f \simeq \boldsymbol{C}_{10}$. Moreover, when $\operatorname{Aut} X_f \simeq \boldsymbol{D}_8$ then $(f, f)_4$ has nonzero discriminant, and when $\operatorname{Aut} X_f \simeq \boldsymbol{D}_4$ then at least one of $(f, f)_4$, $(((f, f)_2, f)_4, f)_4$, and $((((f, f)_2, f)_3, f)_2, f)_6$ has nonzero discriminant.

3. Explicit descent for hyperelliptic curves

3A. ***Field of moduli and fields of definition.*** Let X be a curve defined over K of genus $g \geq 1$, let k be a subfield of K, and let F be the prime field of K.

Definition 3.1. The *field of moduli* of X, denoted $\boldsymbol{M}_X$, is the subfield of K fixed by $\{\sigma \in \operatorname{Aut} K \mid X \simeq X^\sigma\}$.

We now restrict to hyperelliptic curves and we assume that $p \neq 2$. Let $X = X_f$ be a hyperelliptic curve over K given by a hyperelliptic polynomial f of even degree n. Our first task is to show that we can get information on $\boldsymbol{M}_X$ through the invariants.

Lemma 3.2. *Let I_1, I_2 be two invariants of the same degree for binary forms of degree n. Assume that I_1, I_2 are defined over F and that $I_2(f) \neq 0$. Then $\iota = I_1(f)/I_2(f)$ is an element of $\boldsymbol{M}_{X_f}$.*

Proof. It is enough to prove that $\iota^\sigma = \iota$ for all $\sigma \in \operatorname{Gal}(K/\boldsymbol{M}_X)$. By the definition of $\boldsymbol{M}_X$, there exists an isomorphism between X and X^σ. We have seen that such an isomorphism induces an element $M \in \operatorname{Isom}(f, f^\sigma)$. Therefore

$$\iota^\sigma = \frac{I_1(f^\sigma)}{I_2(f^\sigma)} = \frac{I_1(\lambda \cdot M.f)}{I_2(\lambda \cdot M.f)} = \iota. \qquad \square$$

It is not always practical to work with a fixed quotient of invariants as above, since $I_2(f)$ may be zero. As shown in [25], it is better to work inside a weighted projective space, for elements of which one can define a canonical representative as follows. Let $(I_1 : \cdots : I_m)$ be an m-tuple of degree-d_i invariants of degree-n binary forms, where $m \geq 2$, and suppose each I_i is defined over F. Let f be a binary form of degree n. Let d be the gcd of the degrees d_i of the invariants I_i whose values at f are nonzero. Then there exist $c_i \in \mathbb{Z}$, with $c_i = 0$ if $I_i(f) = 0$, such that $\sum c_i d_i = d$. We then define $I = \prod_i I_i^{c_i}$. The *canonical representative* of $(I_1(f) : \cdots : I_m(f))$ is

$$(\mathfrak{I}_1(f), \ldots, \mathfrak{I}_m(f)) = \left(\frac{I_1(f)}{I(f)^{d_1/d}}, \ldots, \frac{I_m(f)}{I(f)^{d_m/d}} \right) \in \boldsymbol{M}_X^m.$$

Proposition 3.3. *Let $(I_1 : \cdots : I_m)$ be a set of generators for $\mathfrak{I}_n$ defined over F. Then*

$$\boldsymbol{M}_X = F(\mathfrak{I}_1(f), \ldots, \mathfrak{I}_m(f)).$$

Proof. Let $\sigma \in \mathrm{Gal}(K/F(\mathfrak{I}_1(f), \ldots, \mathfrak{I}_m(f)))$. Since

$$(\mathfrak{I}_1(f^\sigma), \ldots, \mathfrak{I}_m(f^\sigma)) = (\mathfrak{I}_1(f), \ldots, \mathfrak{I}_m(f)),$$

and since $\mathfrak{I}_n$ separates the orbits of separable forms [28, p. 78], there exists a matrix $M \in \mathrm{GL}_2(K)$ such that $M.f \sim f^\sigma$, hence an isomorphism between X_f and X_f^σ. □

With our current knowledge of invariants, we are then able to compute $\boldsymbol{M}_{X_f}$ for $n = 6, 8, 10$. However, in the following applications to descent we will see that we often do not need a complete set of invariants.

Definition 3.4. We say that k is a *field of definition* of X if there exists a curve $\mathscr{X}/k$ such that $\mathscr{X}$ is K-isomorphic to X. The curve $\mathscr{X}/k$ is a model of X over k and we call a geometric isomorphism between the two curves a *descent isomorphism*.

A classical problem is to determine the smallest field of definition of a curve. Assuming for simplicity that every subfield of K is perfect, if $\boldsymbol{M}_X$ is a field of definition then it is the smallest possible field of definition, because it is the intersection of all the fields of definition (see [23] or [19, Theorem 1.5.8]). There might be an obstruction for $\boldsymbol{M}_X$ being a field of definition, but if there is none we will denote by $\mathscr{X}$ a model of X over $\boldsymbol{M}_X$. In the case of hyperelliptic curves of odd genus, there is a subtlety: The curve $\mathscr{X}$ does not necessarily admit a hyperelliptic equation. However, if it does, we will say that X can be *hyperelliptically defined over* $\boldsymbol{M}_X$, and we denote by $\mathfrak{f} \in \boldsymbol{M}_X[x]$ a hyperelliptic polynomial associated to this model.

One can find in the literature several sufficient conditions for a curve to be hyperelliptically defined over $\boldsymbol{M}_X$. For instance, it is always the case when K is

the algebraic closure of a finite field (see [18, Corollary 2.11]). Over an arbitrary algebraically closed field K, the work of Huggins [18] shows that if the reduced automorphism group is noncyclic then the curve can be hyperelliptically defined over its field of moduli. For $g = 2$, it has been proved that if the reduced automorphism group is nontrivial, then the curve can be hyperelliptically defined over its field of moduli [7]. This is also the case for $g = 3$, except for curves with automorphism group isomorphic to $\boldsymbol{D}_4$ (see [25] and Section 3C2).

3B. ***Explicit hyperelliptic descent.*** Now let X_f be a hyperelliptic curve over K that can be hyperelliptically defined over $\boldsymbol{M}_X$. We want to find $\mathfrak{f} \in \boldsymbol{M}_X[x]$ and $A \in \mathrm{GL}_2(K)$ such that $\mathfrak{f} \sim A.f$. The first task is of course to compute $\boldsymbol{M}_X$. As we have seen, this can be done if we have a set of generators for the invariants of the form f. However, if we do not have a full set of generators, and instead have only some invariants $(I_1, \ldots, I_m)$ over F with $m \geq 2$, we can always try to hyperelliptically descend X_f over the field k generated by $(\mathfrak{I}_1(f), \ldots, \mathfrak{I}_m(f))$. Since $k \subset \boldsymbol{M}_X$, if this can be achieved, we are done.

3B1. *The cocycle approach.* The direct approach relies on the following slightly modified version of Weil's cocycle relations (see [25]).

Lemma 3.5. *The curve X_f can be hyperelliptically defined over k if and only if there exists a finite extension k'/k such that for all $\sigma \in \mathrm{Gal}(K/k)$, there exists $M_\sigma \in \mathrm{GL}_2(k')$ such that $M_\sigma \in \mathrm{Isom}_{k'}(f, f^\sigma)$ and such that for all $\sigma, \tau \in \mathrm{Gal}(K/k)$, we have $M_{\sigma\tau} = M_\sigma^\tau M_\tau$.*

Assume that X_f can be hyperelliptically defined over k and let $\phi : X_f \to X_\mathfrak{f}$ be a descent isomorphism. It induces a matrix $\widetilde{A} \in \mathrm{Isom}_K(f, \mathfrak{f}) \subset \mathrm{PGL}_2(K)$. If we choose a representative $A \in GL_2(K)$ of $\widetilde{A}$, we can define $M_\sigma = (A^{-1})^\sigma A$ for all $\sigma \in \mathrm{Gal}(K/k)$. It is easy to check that this choice of M_σ satisfies all the hypotheses of the lemma. Moreover, if A is defined over a Galois extension L/k then $k' \subset L$, and we have $M_\sigma = \mathrm{id}$ for all $\sigma \in \mathrm{Gal}(K/k)$ such that $\sigma_{|L} = \mathrm{id}$. Conversely, the crucial step to construct such an A is to identify a Galois extension L/k satisfying this property, since in this case one can use an explicit version of Hilbert 90 as in [31, Proposition 3, p. 159]: For a general matrix $P \in \mathrm{GL}_2(k')$ the matrix

$$A = \sum_{\tau \in \mathrm{Gal}(L/k)} P^\tau M_\tau \tag{7}$$

gives a descent morphism.

Lemma 3.6. *Assume that f is defined over an extension k' of k. If $\mathrm{Aut}_K f = \{\mathrm{id}\}$ then we can take L to be the Galois closure of k'/k.*

Proof. We have to prove that A can be defined over such an L. Let A' be induced by a descent morphism. Since $A' \in \mathrm{Isom}_K(f, \mathfrak{f})$, we have

$$((A')^{-1})^\sigma A' \in \mathrm{Isom}_K(f, f^\sigma) = \mathrm{Aut}_K f$$

for all $\sigma \in \mathrm{Gal}(K/L)$; hence there exists $\lambda_\sigma \in K^*$ such that $(A')^\sigma = \lambda_\sigma \cdot A'$. One can easily check that the λ_σ satisfy a cocycle relation, so there exists $e \in K^*$ such that $\lambda_\sigma = e/e^\sigma$ for all σ. We then define $A = e \cdot A'$, and we are done. □

As far as we know, there is no easy way to determine such an L when the automorphism group is nontrivial (but see [25] for the case when k is a finite field). Naïvely, one would expect to be able to construct the cocycle over the field L_0 over which all isomorphisms between f and its conjugates are defined. Typically, what then happens is the following: Let $\sigma \in \mathrm{Gal}(L_0/k)$ be an element of order n. Then usually no M_σ exists over L_0 such that the cocycle condition $1 = M_{\sigma^n} = M^{\sigma^{n-1}} \cdots M^\sigma \cdot M$ is satisfied. We have to work with matrices of the form λM_σ, where λ belongs to a quadratic extension L of L_0. This enlarges the field and the Galois group, which may in turn give rise to more problems of the same type. Even if this problem can be resolved, the computation of (7) is time-consuming and limited to extensions of small degree (less than 50) in practice. In the next section, we present a new idea that works extremely well to get around these difficulties in certain cases.

Remark 3.7. In the odd genus case, it turns out that if we only want X_f to have a model over k, instead of a hyperelliptic model, then the cocycle condition is replaced by the condition $M_{\sigma\tau} \sim M_\sigma^\tau M_\tau$. However, even in this case we do not know a general method to address the problem effectively.

3B2. *The covariant approach.* Using covariants, we can sometimes reduce the problem of descent for X_f to a descent problem for a curve of lower genus.

Theorem 3.8. *Assume that there exists a covariant C of order $r \geq 4$ such that $c = C(f)$ is a hyperelliptic polynomial, and let $X_c : y^2 = c(x)$ be the associated curve. Then $\boldsymbol{M}_{X_c} \subset \boldsymbol{M}_{X_f}$.*

Moreover, if X_c is hyperelliptically defined over $\boldsymbol{M}_{X_c}$, then X_f is hyperelliptically defined over an extension of $\boldsymbol{M}_{X_f}$ of degree at most $[\mathrm{Aut}_K c : \mathrm{Aut}_K f]$.

In particular, if $\mathrm{Aut}_K c = \mathrm{Aut}_K f$ *and X_c is hyperelliptically defined over $\boldsymbol{M}_{X_c}$, then X_f is hyperelliptically defined over $\boldsymbol{M}_{X_f}$.*

Proof. Let σ be an element of the group $\Gamma = \mathrm{Gal}(K/\boldsymbol{M}_{X_f})$. Then there exists a K-isomorphism between X_f and X_f^σ which induces a matrix $M \in \mathrm{Isom}_K(f, f^\sigma)$. Since we have the inclusion $\mathrm{Isom}_K(f, f^\sigma) \subset \mathrm{Isom}_K(c, c^\sigma)$ by Proposition 2.5, we get a K-isomorphism between X_c and X_c^σ, so $\boldsymbol{M}_{X_c} \subset \boldsymbol{M}_{X_f}$.

Assume now that X_c can be hyperelliptically defined over $\boldsymbol{M}_{X_c}$ as $X_{\mathfrak{c}}$ for some form $\mathfrak{c} \in \boldsymbol{M}_{X_c}[x]$. There exists $A \in \mathrm{Isom}_K(c, \mathfrak{c})$. Let us consider $h = A.f$, which we can assume to be monic. We want to prove that h is defined over an extension of $\boldsymbol{M}_{X_f} = \boldsymbol{M}_{X_h}$ of degree at most

$$\ell = \#(\mathrm{Aut}_K\, c/\mathrm{Aut}_K\, f) = \#(\mathrm{Aut}_K\, \mathfrak{c}/\mathrm{Aut}_K\, h).$$

First note that $C(h) \sim A.C(f) \sim \mathfrak{c}$. Let $H \subset \Gamma$ be the subgroup consisting of the automorphisms σ such that $h \sim h^\sigma$. Since we have assumed that h is monic, we even have $h = h^\sigma$. We must show that $\#\Gamma/H \leq \ell$. To this end, we note that $\mathfrak{c}^\sigma = \mathfrak{c}$ for all $\sigma \in \Gamma$. Hence we can associate to each $\sigma \in \Gamma$ a matrix $M \in \mathrm{Isom}_K(h, h^\sigma) \subset \mathrm{Aut}_K\, \mathfrak{c}$. In fact, this association gives rise to a well-defined class of $\mathrm{Aut}_K\, \mathfrak{c}/\mathrm{Aut}_K\, h$, so we have defined a map ρ from Γ to $\mathrm{Aut}_K\, \mathfrak{c}/\mathrm{Aut}_K\, h$. If $\rho(\sigma) = \rho(\sigma')$ then we have $h^\sigma \sim h^{\sigma'}$, and hence $\sigma^{-1}\sigma' \in H$. Therefore ρ induces an injective map from Γ/H to $\mathrm{Aut}_K\, \mathfrak{c}/\mathrm{Aut}_K\, h$, and we get our result. □

To use the theorem in a constructive way, we need a covariant that has a finite automorphism group and for which we know how to find a hyperelliptic model over its field of moduli. We give some examples in Sections 3C and 3D.

Remark 3.9. The fields of moduli of X_f and X_c may be different, even when the automorphism groups of the forms are the same. For instance, let r be a root of $t^2 + 2t + 16/9 = 0$ and let f be the form

$$f = (x^4 + rx^2z^2 + z^4)(x^4 - 3rx^2z^2 + z^4);$$

then the field of moduli of f is $\mathbb{Q}(r)$, while the field of moduli of

$$c = (f, f)_6 = (16/49)x^4 + (992/441)x^2 + (16/49)$$

is $\mathbb{Q}$. Using the programs of [25], one sees that $\mathrm{Aut}_K\, f = \mathrm{Aut}_K\, c \simeq \boldsymbol{D}_4$.

3C. ***Application to genus-3 hyperelliptic curves.*** In [25], the two first authors give algorithms for reconstructing genus-3 hyperelliptic models from given invariants. These models are defined over the field of moduli, with the notable exception of the 2-dimensional stratum $\boldsymbol{C}_2^3$ and the 3-dimensional stratum $\boldsymbol{D}_4$. As an illustration of our strategy, we see how our method applies in these remaining cases.

3C1. *Descent of curves with automorphism group* $\boldsymbol{C}_2^3$. Let $X/K : y^2 = f(x)$ be a genus-3 hyperelliptic curve with automorphism group isomorphic to $\boldsymbol{C}_2^3$. Since the reduced automorphism group is not cyclic, [18] shows that X can be hyperelliptically defined over its field of moduli. In [25], we showed how to construct a hyperelliptic equation for a model over an extension of the field of moduli of degree at most 3. Using covariants, we can now give a method to get an equation over the field of moduli itself.

In Section 2E2, we checked that at least one of the quartic covariants in the list $\{C_{2,4}(f), C_{3,4}(f), C_{4,4}(f)\}$ has nonzero discriminant. Moreover, we see by Proposition 2.6 that the automorphism group of such a quartic is equal to $\boldsymbol{D}_4$ if the quartic invariants I and J are both nonzero. Using some formal computations (see the Magma scripts available at the URL listed in the Introduction), we checked that it is always the case that at least one of the three covariants has nonzero discriminant and I and J nonzero. Since $\mathrm{Aut}_K(f) \simeq \boldsymbol{D}_4$ we can use the approach of Theorem 3.8 to find a hyperelliptic equation $y^2 = \mathfrak{f}(x)$ over the field of moduli. The procedure can actually be applied to a generic element of the family, but the result is too large to be written down here; instead, we present an example.

Example 3.10. When we evaluate the parametrization formulas given in [25] for the stratum $\boldsymbol{C}_2^3$ at $t = 0$ and $u = 1$, we find the rational point

$$(j_2 : j_3 : \cdots : j_{10}) = \left(0 : 0 : -\frac{25}{98} : -\frac{25}{98} : -\frac{225}{2744} : -\frac{25}{1372} : -\frac{225}{134456} : \frac{1125}{76832} : \frac{15125}{3764768}\right)$$

in the moduli space. This gives rise to the curve $X : y^2 = f$ with

$$f = (-32\alpha^2 + 420\alpha - 2275)x^8/160 + (-12\alpha^2 + 140\alpha - 700)x^6/25 + \alpha x^4 + x^2 + (16\alpha^2 + 280\alpha - 2275)/12250$$

over $\mathbb{Q}(\alpha)$, with $\alpha^3 - (35/2)\alpha^2 + (1925/16)\alpha - (18375/64) = 0$. By Proposition 3.3, we have $\boldsymbol{M}_X = \mathbb{Q}$.

Let c be the covariant $(f, f)_6$. We find

$$c = \frac{-16\alpha^2 + 180\alpha - 875}{280}x^4 + \frac{24\alpha^2 - 630\alpha + 3150}{1225}x^2z^2 + \frac{4\alpha + 35}{490}z^4,$$

so that $I = -75/49$ and $J = -2025/343$. Then $\mathfrak{c} = x^3z + (25/9)xz^3 + (25/9)z^4$ is $\mathrm{GL}_2(\overline{\mathbb{Q}})$-equivalent to c, is defined over $\boldsymbol{M}_X = \mathbb{Q}$, and satisfies $\mathrm{Aut}_{\overline{\mathbb{Q}}}\,\mathfrak{c} \simeq \boldsymbol{D}_4$. The direct approach of Section 2B explicitly finds a $\overline{\mathbb{Q}}$-isomorphism M between c and $\mathfrak{c}$. Its inverse M^{-1} is equal to $(m_{i,j})_{i,j}$, where

$$\begin{aligned}
m_{11} &= 110250,\\
m_{12} &= (3360\alpha^2 - 58800\alpha + 147000)\beta^2 - 16800\alpha^2 + 147000\alpha - 18375,\\
m_{21} &= (-2064\alpha^2 + 24780\alpha - 60900)\beta^3 + (-3120\alpha^2 + 67200\alpha - 375375)\beta,\\
m_{22} &= (-5840\alpha^2 + 74900\alpha - 280000)\beta^3 + (16880\alpha^2 - 173600\alpha + 487375)\beta.
\end{aligned}$$

Here β satisfies

$$\beta^4 + \frac{32\alpha^2 - 280\alpha + 350}{175}\beta^2 - \frac{176\alpha^2 - 1820\alpha + 7350}{175} = 0.$$

We compute the monic form $\mathfrak{f} \sim M.f$:

$$\mathfrak{f} = x^8 + 160x^7 - 560x^6 - 2800x^5 + 64750x^4 - 91000x^3 + 3010000x^2 - 2225000x - 9696875.$$

So $y^2 = \mathfrak{f}(x)$ is a model of X over $\boldsymbol{M_X} = \mathbb{Q}$.

3C2. *Descent of curves with automorphism group* $\boldsymbol{D_4}$. It is proved in [19, Chapter 5] that there may be an obstruction for a genus-3 hyperelliptic curve over K with automorphism group isomorphic to $\boldsymbol{D_4}$ to have a model over its field of moduli. In [25], we were able to construct a model of such curves over an extension of the field of moduli of degree at most 8. Using Theorem 3.8, we find:

Proposition 3.11. *Let X_f be a genus-3 hyperelliptic curve over K with automorphism group isomorphic to $\boldsymbol{D_4}$. Then there exists an explicit model of X over an at most quadratic extension of $\boldsymbol{M_X}$.*

Proof. Applying the methods of Proposition 2.11 to the stratum $\boldsymbol{D_4}$ shows that at least one of the five binary covariants $C_{2,4}(f)$, $C_{3,4}(f)$, $C_{4,4}(f)$, $C'_{4,4}(f)$, $C_{5,4}(f)$ has not only a discriminant different from 0, but also $I(f) \neq 0$, $J(f) \neq 0$. (The computations can be found in the Magma scripts available at the URL given in the Introduction.) One then combines Proposition 2.6 and Theorem 3.8. □

We plan to investigate how to apply the theory of twists to the binary quartics used in the application of Theorem 3.8 to give a precise characterization of the obstruction to the descent on the field of moduli.

3D. ***Application to a family of Fuertes-González-Diez in genus 5.*** Let k be the degree-3 Galois extension of $\mathbb{Q}$ defined by the irreducible polynomial $t^3 - 3t + 1$. Let r_1, r_2, r_3 be the roots of this polynomial in k. Then, as in [13], we can consider the family

$$y^2 = \prod_{i=4}^{6} \left(x^4 - 2\left(1 - 2\,\frac{r_3 - r_1}{r_3 - r_2}\,\frac{q_i - r_2}{q_4 - r_1}\right)x^2 + 1 \right) \tag{8}$$

of genus-5 hyperelliptic curves, with q_4, q_5, q_6 in $\mathbb{Q}$. It was proved in [13] that the members of this family have field of moduli equal to $\mathbb{Q}$ and automorphism group isomorphic to $\boldsymbol{C_2^3}$. Moreover, it was claimed in [13] that these curves cannot be hyperelliptically defined over $\mathbb{Q}$, in contradiction with [18]. However, the proof turns out to contain a subtle error. Still, the explicit descent of any of the member of the family was extremely hard.

As in Example 3.10, we can use Theorem 3.8 to construct an explicit descent for the curves in this family. For this particular family, the descent can even be performed uniformly to yield a general expression in q_4, q_5, q_6. Let $F = k(q_4, q_5, q_6)$

be the rational function field over k in three indeterminates, and define the binary quartic form $f \in F[x, z]$ as the homogenization of the right-hand side of (8). Let c be the transvectant $(f, f)_{10}$. Then c is a covariant of order 4 with nonzero discriminant and nonzero $I(c)$ and $J(c)$, and hence has automorphism group $\boldsymbol{D}_4$. The field of moduli of X_c is contained in the field of moduli of X_f, which is a subfield of $\mathbb{Q}(q_4, q_5, q_6)$; therefore the quartic $\mathfrak{c}$ as in (4) is defined over $\mathbb{Q}(q_4, q_5, q_6)$ and is $\mathrm{GL}_2(\bar{F})$-equivalent to c.

Now let L be the degree-4 extension of F defined by the dehomogenization of c. From Proposition 2.7, we can explicitly construct an L-isomorphism between c and $\mathfrak{c}$. This transformation gives a descent of the curve corresponding to c, which by Theorem 3.8 also yields a descent of the curve corresponding to f. The resulting expression, though indeed defined over the rationals, is huge and impossible to give here. (The computations above, their final result, and the program to compute the descent of any given specialization are available at the URL listed in the Introduction.) However, we can give an example for a specialization.

Example 3.12. Take $q_4 = 1$, $q_5 = 2$, $q_6 = 3$. The hyperelliptic equation over $\mathbb{Q}$ is

$$\begin{aligned} y^2 = {} & 199950247575x^{12} - 296949924611352x^{11} - 66659816245812750x^{10} \\ & - 1542197549550736065 6x^9 + 20056355190424553708745x^8 \\ & + 13079208886477241946120 0x^7 + 44148454149188354317253820x^6 \\ & - 97188470839086936498039591 36x^5 \\ & + 93749472927036312839424054441x^4 \\ & + 86331359417888600607650948443656x^3 \\ & - 74239120806631825130459382051613 26x^2 \\ & + 24951119764116840493951094604151 5184x \\ & - 30066561438584723177639735809842 60681. \end{aligned}$$

Acknowledgment

The authors acknowledge support by grant ANR-09-BLAN-0020-01.

References

[1] Lars V. Ahlfors, Lipman Bers, Hershel M. Farkas, Robert C. Gunning, Irwin Kra, and Harry E. Rauch (eds.), *Advances in the theory of Riemann surfaces*: *Proceedings of the* 1969 *Stony Brook Conference*, Annals of Mathematics Studies, no. 66, Princeton University Press, 1971. MR 43 #5023

[2] Leonid Bedratyuk, *On complete system of invariants for the binary form of degree 7*, J. Symbolic Comput. **42** (2007), no. 10, 935–947. MR 2008h:13009

[3] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system, I: The user language*, J. Symbolic Comput. **24** (1997), no. 3–4, 235–265. MR 1484478

[4] Andries E. Brouwer and Mihaela Popoviciu, *The invariants of the binary decimic*, J. Symbolic Comput. **45** (2010), no. 8, 837–843. MR 2011f:13007

[5] ———, *The invariants of the binary nonic*, J. Symbolic Comput. **45** (2010), no. 6, 709–720. MR 2011e:13012

[6] Duncan Buell (ed.), *Algorithmic number theory: Proceedings of the 6th International Symposium (ANTS-VI) held at the University of Vermont, Burlington, VT, June 13–18, 2004*, Lecture Notes in Computer Science, no. 3076, Berlin, Springer, 2004. MR 2005m:11002

[7] Gabriel Cardona and Jordi Quer, *Field of moduli and field of definition for curves of genus 2*, in Shaska [32], 2005, pp. 71–83. MR 2006h:14036

[8] J. E. Cremona and T. A. Fisher, *On the equivalence of binary quartics*, J. Symbolic Comput. **44** (2009), no. 6, 673–682. MR 2010c:11049

[9] H. Cröni, *Zur Berechnung von Kovarianten von Quantiken*, Ph.D. thesis, Universität des Saarlandes, 2002.

[10] J. Dixmier and D. Lazard, *Le nombre minimum d'invariants fondamentaux pour les formes binaires de degré* 7, Portugal. Math. **43** (1985/86), no. 3, 377–392. MR 88f:15045

[11] Jacques Dixmier, *Quelques aspects de la théorie des invariants*, Gaz. Math. (1990), no. 43, 39–64. MR 90m:15047

[12] Clifford J. Earle, *On the moduli of closed Riemann surfaces with symmetries*, in Ahlfors et al. [1], 1971, pp. 119–130. MR 45 #5343

[13] Y. Fuertes and G. González-Diez, *Fields of moduli and definition of hyperelliptic covers*, Arch. Math. (Basel) **86** (2006), no. 5, 398–408. MR 2007f:14028

[14] August Freiherr von Gall, *Das vollständige Formensystem der binären Form* 7^{ter} *Ordnung*, Math. Ann. **31** (1888), no. 3, 318–336. MR 1510486

[15] Paul Gordan, *Beweis, dass jede Covariante und Invatiante einer binären Form eine ganze Function mit numerischen Coefficienten einer endlichen Anzahl solcher Formen ist*, J. Reine Angew. Math. **69** (1868), 323–354.

[16] J. H. Grace and A. Young, *The algebra of invariants*, Cambridge University Press, 1903.

[17] F. Hess, *An algorithm for computing isomorphisms of algebraic function fields*, in Buell [6], 2004, pp. 263–271. MR 2006d:11141

[18] Bonnie Huggins, *Fields of moduli of hyperelliptic curves*, Math. Res. Lett. **14** (2007), no. 2, 249–262. MR 2009a:14040

[19] Bonnie Sakura Huggins, *Fields of moduli and fields of definition of curves*, Ph.D. thesis, University of California, Berkeley, 2005. arXiv math/0610247 [math.NT]

[20] IEEE (ed.), *49th IEEE Symposium on Foundations of Computer Science (FOCS 2008), held in Philadelphia, October 25–28, 2008*, Los Alamitos, CA, Institute of Electrical and Electronics Engineers, IEEE Computer Society, 2008.

[21] Jun-ichi Igusa, *Arithmetic variety of moduli for genus two*, Ann. of Math. (2) **72** (1960), 612–649. MR 22 #5637

[22] Kiran S. Kedlaya and Christopher Umans, *Fast modular composition in any characteristic*, in IEEE [20], 2008, pp. 146–155.

[23] Shoji Koizumi, *The fields of moduli for polarized abelian varieties and for curves*, Nagoya Math. J. **48** (1972), 37–55. MR 50 #4582

[24] Reynald Lercier and Christophe Ritzenthaler, *Invariants and reconstructions for genus* 2 *curves in any characteristic*, available in MAGMA 2.15 [3] and later.

[25] ———, *Hyperelliptic curves and their invariants: Geometric, arithmetic and algorithmic aspects*, J. Algebra **372** (2012), 595–636. MR 2990029

[26] Jean-François Mestre, *Construction de courbes de genre* 2 *à partir de leurs modules*, in Mora and Traverso [27], 1991, pp. 313–334. MR 92g:14022

[27] Teo Mora and Carlo Traverso (eds.), *Effective methods in algebraic geometry: Papers from the symposium (MEGA-90) held in Castiglioncello, April 17–21, 1990*, Progress in Mathematics, no. 94, Birkhäuser, Boston, 1991. MR 91m:14003

[28] David Mumford and John Fogarty, *Geometric invariant theory*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete, no. 34, Springer, Berlin, 1982. MR 86a:14006

[29] Peter J. Olver, *Classical invariant theory*, London Mathematical Society Student Texts, no. 44, Cambridge University Press, 1999. MR 2001g:13009

[30] Sander Matthijs van Rijnswou, *Testing the equivalence of planar curves*, Ph.D. thesis, Technische Universiteit Eindhoven, 2001. http://repository.tue.nl/543172

[31] Jean-Pierre Serre, *Corps locaux*, 2nd ed., Publications de l'Université de Nancago, no. VIII, Hermann, Paris, 1968. MR 50 #7096

[32] Tanush Shaska (ed.), *Computational aspects of algebraic curves*: *Papers from the conference held at the University of Idaho*, *Moscow*, *ID*, *May* 26–28, 2005, Lecture Notes Series on Computing, no. 13, World Sci. Publ., Hackensack, NJ, 2005. MR 2006e:14003

[33] Goro Shimura, *On the field of rationality for an abelian variety*, Nagoya Math. J. **45** (1972), 167–178. MR 46 #5342

[34] Tetsuji Shioda, *On the graded ring of invariants of binary octavics*, Amer. J. Math. **89** (1967), 1022–1046. MR 36 #3790

[35] André Weil, *The field of definition of a variety*, Amer. J. Math. **78** (1956), 509–524. MR 18,601a

REYNALD LERCIER: reynald.lercier@m4x.org
DGA MI, La Roche Marguerite, 35174 Bruz, France

and

Laboratoire IRMAR, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes, France

CHRISTOPHE RITZENTHALER: christophe.ritzenthaler@univ-rennes1.fr
Institut de Mathématiques de Luminy, UMR 6206 du CNRS, Luminy, Case 907, 13288 Marseille, France
Current address: *Laboratoire IRMAR, UMR CNRS 6625, Campus de Beaulieu, 35042 Rennes, France*

JEROEN SIJSLING: sijsling@gmail.com
Laboratoire IRMAR, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes, France
Current address: *Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, United Kingdom*

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.487

msp

Elliptic factors in Jacobians of hyperelliptic curves with certain automorphism groups

Jennifer Paulhus

We decompose the Jacobian varieties of hyperelliptic curves up to genus 20, defined over an algebraically closed field of characteristic zero, with reduced automorphism group A_4, S_4, or A_5. Among these curves is a genus-4 curve with Jacobian variety isogenous to $E_1^2 \times E_2^2$ and a genus-5 curve with Jacobian variety isogenous to E^5, for E and E_i elliptic curves. These types of results have some interesting consequences for questions of ranks of elliptic curves and ranks of their twists.

1. Introduction

Curves with Jacobian varieties that have many elliptic curve factors in their decompositions up to isogeny have been studied in many different contexts. Ekedahl and Serre found examples of curves whose Jacobians split completely into elliptic curves (not necessarily isogenous to one another) [13] (see also [27], [14, §5]). In genus 2, Cardona showed connections between curves whose Jacobians have two isogenous elliptic curve factors and $\mathbb{Q}$-curves of degree 2 and 3 [5]. There are applications of such curves to ranks of twists of elliptic curves [24], results on torsion [19], and cryptography [12].

Let J_X denote the Jacobian variety of a curve X and let $\sim$ represent an isogeny between abelian varieties. Consider the following question.

Question 1. For a fixed genus g, what is the largest positive integer t such that $J_X \sim E^t \times A$ for some genus-g curve X over the algebraic closure of $\mathbb{Q}$, where E is an elliptic curve and A an abelian variety?

MSC2010: primary 14H40; secondary 11G30, 14H37.
Keywords: Jacobian varieties, hyperelliptic curves, automorphism groups of Riemann surfaces.

In [22] the author developed a method for decomposing the Jacobian variety of a curve X with automorphism group G, based on idempotent relations in the group ring $\mathbb{Q}[G]$. This technique yielded thitherto unknown examples of curves of genus 4 through 6 where t is as large as is possible — that is, t is equal to the genus g. For genus 7 through 10, examples of curves whose Jacobians have many isogenous elliptic curves in their decompositions were also found. All these examples are nonhyperelliptic curves.

In this paper we apply the methods of [22] to hyperelliptic curves with certain automorphism groups. Let X be a hyperelliptic curve defined over a field of characteristic 0, with hyperelliptic involution ω. The automorphism group of the curve X modulo the subgroup $\langle\omega\rangle$ is called the *reduced automorphism group* and must be one of the groups C_n, D_n, A_4, S_4, or A_5; here C_n represents the cyclic group of order n and D_n is the dihedral group of order $2n$. This follows from a result of Dickson on transformations of binary forms [7].

We study hyperelliptic curves with reduced automorphism group one of A_4, S_4, or A_5. These reduced automorphism groups were chosen for two reasons. First, results from genus 2 and 3 suggest that these families may yield curves with many isogenous elliptic curve factors in higher genus. Second, for any genus, the list of full automorphism groups with reduced automorphism group one of A_4, S_4, or A_5 is manageable.

Section 3 reviews the method from [22], and Section 4 gives proofs of results for genus up to 20. This bound of genus 20 is somewhat arbitrary. The technique will work for any genus, but the computations become more complicated as the genus increases. Section 5 discusses some computational obstructions to producing results in higher genus. In that section we also work with families of curves with three particular automorphism groups. These groups have special properties that allow us to prove results about the decomposition of the curves' Jacobians for arbitrary genus.

A brief word on fields of definition: Unless specifically stated otherwise, curves in this paper are defined over an algebraically closed field of characteristic zero. The method of decomposition works generally for curves over any field; however, a particular field must be specified in order to determine the automorphism group of the curve. In each individual case, the decomposition results will hold for the Jacobian of the curve defined over any field over which every geometric automorphism of the curve is defined. Partial answers to Question 1 are known for curves over fields of characteristic p; see, for example, [28; 17; 9].

2. Overview of results

The decompositions of Jacobian varieties of hyperelliptic curves with reduced automorphism group A_4, S_4, or A_5 up to genus 20 are summarized in Theorem 5.

Jacobian varieties with several isogenous elliptic curve factors are also found, and many are improvements on the best known results for t [22]. Two results of particular interest are:

Theorem 1. *The hyperelliptic curve of genus* 4 *with affine model*

$$X: \quad y^2 = x(x^4 - 1)(x^4 + 2\sqrt{-3}\,x^2 + 1)$$

has a Jacobian variety that decomposes as $E_1^2 \times E_2^2$ *for two elliptic curves* E_i.

Theorem 2. *The genus-*5 *hyperelliptic curve with affine model*

$$X: \quad y^2 = x(x^{10} + 11x^5 - 1)$$

has $J_X \sim E^5$ *for the elliptic curve* E *with equation* $y^2 = x(x^2 + 11x - 1)$.

The first theorem is an improvement from best decompositions of genus-4 hyperelliptic curves from [23]. The second theorem is, to the author's knowledge, the first example in the literature of a hyperelliptic curve with a Jacobian variety that decomposes into five isogenous elliptic curves over a number field. Proofs of these results may be found in Section 4.

3. Review of technique

Fix an algebraically closed field k of characteristic 0. Throughout the paper the word *curve* will mean a smooth projective variety of dimension 1. For simplicity, models are affine, when given. Any parameters in the affine model (labeled as "a_i") are elements of k. Also, ζ_n will denote a primitive n-th root of unity.

Given a curve X of genus g over k, the *automorphism group* of X is the automorphism group of the field extension $k(X)$ over k, where $k(X)$ is the function field of X. This group will always be finite for $g \geq 2$. Throughout, G will denote the automorphism group of a curve X. In the case of hyperelliptic curves over algebraically closed fields of characteristic zero, all possible automorphism groups are known for a given genus [2; 4; 25].

Kani and Rosen [20] proved a result connecting certain idempotent relations in the endomorphism algebra $\mathrm{End}^0 J_X = (\mathrm{End}\, J_X) \otimes_{\mathbb{Z}} \mathbb{Q}$ to isogenies among images of J_X under endomorphisms. If α_1 and α_2 are elements of $\mathrm{End}^0 J_X$, we write $\alpha_1 \sim \alpha_2$ if $\chi(\alpha_1) = \chi(\alpha_2)$ for all $\mathbb{Q}$-characters χ of $\mathrm{End}^0 J_X$.

Theorem 3 [20, Theorem A]. *Let* $\varepsilon_1, \ldots, \varepsilon_n, \varepsilon'_1, \ldots, \varepsilon'_m \in \mathrm{End}^0 J_X$ *be idempotents. Then the idempotent relation*

$$\varepsilon_1 + \cdots + \varepsilon_n \sim \varepsilon'_1 + \cdots + \varepsilon'_m$$

holds in $\mathrm{End}^0 J_X$ *if and only if there is the isogeny relation*

$$\varepsilon_1(J_X) \times \cdots \times \varepsilon_n(J_X) \sim \varepsilon'_1(J_X) \times \cdots \times \varepsilon'_m(J_X).$$

There is a natural $\mathbb{Q}$-algebra homomorphism from $\mathbb{Q}[G]$ to $\mathrm{End}^0 J_X$, which we will denote by e. It is a well-known result of Wedderburn [11, §18.2] that any group ring of the form $\mathbb{Q}[G]$ has a decomposition into a direct sum of matrix rings over division rings Δ_i:

$$\mathbb{Q}[G] \cong \bigoplus_i M_{n_i}(\Delta_i). \tag{1}$$

Define $\pi_{i,j}$ to be the idempotent in $\mathbb{Q}[G]$ which is the zero matrix for all components except the i-th component where it is the matrix with a 1 in the (j, j) position and zeros elsewhere. The following equation is an idempotent relation in $\mathbb{Q}[G]$:

$$1_{\mathbb{Q}[G]} = \sum_{i,j} \pi_{i,j}.$$

Applying the map e to this relation and using Theorem 3, we find

$$J_X \sim \bigoplus_{i,j} e(\pi_{i,j}) J_X. \tag{2}$$

Recall that our primary goal is to study isogenous elliptic curves that appear in the decomposition above. In order to identify which summands in (2) have dimension 1, we use results from [15, §5.2] to compute the dimensions of these factors. This requires a certain representation of G.

Definition. The *Hurwitz representation* V of a group G is defined by the action of G on $H_1(X, \mathbb{Z}) \otimes \mathbb{Q}$.

The character of this representation is computed as follows. Let $\rho: X \to Y = X/G$ be the natural map from X to its quotient by G. Suppose ρ is branched at s points, with monodromy $g_1, \ldots, g_s \in G$ (unique up to conjugation). Let χ_{triv} be the trivial character of G, and for each i let $\chi_{\langle g_i \rangle}$ denote the character of G induced from the trivial character of the subgroup $\langle g_i \rangle$ of G; observe that $\chi_{\langle 1_G \rangle}$ is the character of the regular representation. If we let g_Y denote the genus of Y, then the character of the Hurwitz representation V is defined as

$$\chi_V = 2\chi_{\text{triv}} + 2(g_Y - 1)\chi_{\langle 1_G \rangle} + \sum_i \big(\chi_{\langle 1_G \rangle} - \chi_{\langle g_i \rangle}\big). \tag{3}$$

Note that for a hyperelliptic curve X, we have $X/G \cong \mathbb{P}^1$ (since G contains the hyperelliptic involution) and so $g_Y = 0$. Also, $\chi_{\langle g_i \rangle} = \chi_{\langle g_j \rangle}$ if $\langle g_i \rangle$ and $\langle g_j \rangle$ are conjugate subgroups.

Via the regular representation, each element g_i can be written as an element of the symmetric group S_n, where $n = \#G$. The monodromy type of a cover will be written as an ordered tuple $(t_1^{(a_1)}, \ldots, t_s^{(a_s)})$ where $t_i^{(a_i)}$ corresponds to g_i and denotes a permutation consisting of a_i t_i-tuples. If χ_i is the irreducible $\mathbb{Q}$-character

associated to the i-th component from (1), then the dimensions of the summands in (2) are

$$\dim e(\pi_{i,j})J_X = \tfrac{1}{2}\dim_{\mathbb{Q}} \pi_{i,j}V = \tfrac{1}{2}\langle \chi_i, \chi_V \rangle. \tag{4}$$

See [15, §5.2] for more information on the dimension computations.

Hence, given the automorphism group G of a curve X and monodromy for the cover X over Y, to compute these dimensions we first determine the degrees of the irreducible $\mathbb{Q}$-characters of G, which will be the n_i values in (1). Next we identify elements of the automorphism group that satisfy the monodromy conditions. We compute the Hurwitz character for this group and covering using (3), and finally compute the inner product of the irreducible $\mathbb{Q}$-characters with the Hurwitz character.

Again, our particular interest is in factors that are *isogenous* to one another. The following proposition gives a condition for the factors to be isogenous.

Proposition 4 [23]. *With notation as above*, $e(\pi_{i,j_1})J_X \sim e(\pi_{i,j_2})J_X$.

Suppose a curve of genus g has automorphism group with group ring decomposition as in (1) with at least one matrix ring of degree close to g; that is, one n_i value close to g — call it n_j. If the computations of dimensions of abelian variety factors outlined above lead to a dimension-1 variety in the place corresponding to that matrix ring (the j-th place), Proposition 4 implies that the Jacobian variety decomposition consists of n_j isogenous elliptic curves. Our goal then is to apply the steps above to hyperelliptic curves of genus up to 20 and with reduced automorphism group isomorphic to A_4, S_4, or A_5.

4. Results

For hyperelliptic curves over an algebraically closed field of characteristic zero, the existence of curves of a fixed genus with reduced automorphism group isomorphic to one of A_4, S_4, or A_5 is completely determined by whether the genus is in certain residue classes modulo 6, 12, and 30, respectively [25].

For each reduced automorphism group there are several possible full automorphism groups. Table 1 lists all groups and the modular conditions for their existence in a certain genus, as well as monodromy type, listed using the notation described in the previous section. The data from this table is taken from [25, Table 1, p. 250]. Explanations of how this data was produced may be found in [25], along with affine models for all of the corresponding families. The groups

$$W_2 = \langle u, v \mid u^4, v^3, vu^2v^{-1}u^2, (uv)^4 \rangle,$$
$$W_3 = \langle u, v \mid u^4, v^3, u^2(uv)^4, (uv)^8 \rangle$$

mentioned in the table are both of order 48.

Automorphism group			
Reduced	Full	Genus restrictions	Monodromy
A_4	$A_4 \times C_2$	$5 \bmod 6$	$(3^{(8)}, 3^{(8)}, 2^{(12)}, \ldots, 2^{(12)})$
	$A_4 \times C_2$	$1 \bmod 6$	$(3^{(8)}, 6^{(4)}, 2^{(12)}, \ldots, 2^{(12)})$
	$A_4 \times C_2$	$3 \bmod 6, g > 3$	$(6^{(4)}, 6^{(4)}, 2^{(12)}, \ldots, 2^{(12)})$
	$\mathrm{SL}_2(3)$	$2 \bmod 6, g > 2$	$(4^{(6)}, 3^{(8)}, 3^{(8)}, 2^{(12)}, \ldots, 2^{(12)})$
	$\mathrm{SL}_2(3)$	$4 \bmod 6$	$(4^{(6)}, 3^{(8)}, 6^{(4)}, 2^{(12)}, \ldots, 2^{(12)})$
	$\mathrm{SL}_2(3)$	$0 \bmod 6, g > 6$	$(4^{(6)}, 6^{(4)}, 6^{(4)}, 2^{(12)}, \ldots, 2^{(12)})$
S_4	$S_4 \times C_2$	$11 \bmod 12$	$(3^{(16)}, 4^{(12)}, 2^{(24)}, \ldots, 2^{(24)})$
	$S_4 \times C_2$	$3 \bmod 12$	$(6^{(8)}, 4^{(12)}, 2^{(24)}, \ldots, 2^{(24)})$
	$\mathrm{GL}_2(3)$	$2 \bmod 12$	$(3^{(16)}, 8^{(6)}, 2^{(24)}, \ldots, 2^{(24)})$
	$\mathrm{GL}_2(3)$	$6 \bmod 12$	$(6^{(8)}, 8^{(6)}, 2^{(24)}, \ldots, 2^{(24)})$
	W_2	$5 \bmod 12$	$(4^{(12)}, 4^{(12)}, 3^{(16)}, 2^{(24)}, \ldots, 2^{(24)})$
	W_2	$9 \bmod 12$	$(4^{(12)}, 4^{(12)}, 6^{(8)}, 2^{(24)}, \ldots, 2^{(24)})$
	W_3	$8 \bmod 12$	$(4^{(12)}, 3^{(16)}, 8^{(6)}, 2^{(24)}, \ldots, 2^{(24)})$
	W_3	$0 \bmod 12$	$(4^{(12)}, 6^{(8)}, 8^{(6)}, 2^{(24)}, \ldots, 2^{(24)})$
A_5	$A_5 \times C_2$	$29 \bmod 30$	$(3^{(40)}, 5^{(24)}, 2^{(60)}, \ldots, 2^{(60)})$
	$A_5 \times C_2$	$5 \bmod 30$	$(3^{(40)}, 10^{(12)}, 2^{(60)}, \ldots, 2^{(60)})$
	$A_5 \times C_2$	$15 \bmod 30$	$(6^{(20)}, 10^{(12)}, 2^{(60)}, \ldots, 2^{(60)})$
	$A_5 \times C_2$	$9 \bmod 30$	$(6^{(20)}, 5^{(24)}, 2^{(60)}, \ldots, 2^{(60)})$
	$\mathrm{SL}_2(5)$	$14 \bmod 30$	$(4^{(30)}, 3^{(40)}, 5^{(24)}, 2^{(60)}, \ldots, 2^{(60)})$
	$\mathrm{SL}_2(5)$	$20 \bmod 30$	$(4^{(30)}, 3^{(40)}, 10^{(12)}, 2^{(60)}, \ldots, 2^{(60)})$
	$\mathrm{SL}_2(5)$	$24 \bmod 30$	$(4^{(30)}, 6^{(20)}, 5^{(24)}, 2^{(60)}, \ldots, 2^{(60)})$
	$\mathrm{SL}_2(5)$	$0 \bmod 30$	$(4^{(30)}, 6^{(20)}, 10^{(12)}, 2^{(60)}, \ldots, 2^{(60)})$

Table 1. Full automorphism groups of hyperelliptic curves with certain reduced automorphism groups. For each group $\tilde{G}$ in the first column, we list the possible automorphism groups G occurring for hyperelliptic curves with reduced automorphism group $\tilde{G}$. The third column lists restrictions on the genus g of hyperelliptic curves with the given automorphism group, and the fourth column lists the monodromy of such curves.

Applying the technique in Section 3 to hyperelliptic curves of genus 3 through 20 produces results that are summarized in the following theorem.

Theorem 5. *For hyperelliptic curves up to genus* 20 *defined over an algebraically closed field of characteristic zero with reduced automorphism group A_4, S_4, or A_5, Table 2 gives a decomposition of the Jacobian of these curves up to isogeny. In the table E_i represents an elliptic curve and $A_{i,j}$ is an abelian variety of dimension $i > 1$, indexed if necessary by j. The dimension of the family with each automorphism group in the moduli space is also included.*

Genus	Automorphism Group	Dimension	Jacobian decomposition
3	$S_4 \times C_2$	0	E^3
4	$\mathrm{SL}_2(3)$	0	$E_1^2 \times E_2^2$
5	$A_4 \times C_2$	1	$E^3 \times A_2$
	W_2	0	$E_1^2 \times E_2^3$
	$A_5 \times C_2$	0	E^5
6	$\mathrm{GL}_2(3)$	0	$E_1^2 \times E_2^4$
7	$A_4 \times C_2$	1	$E_1 \times E_2^3 \times E_3^3$
8	$\mathrm{SL}_2(3)$	1	$A_{2,1}^2 \times A_{2,2}^2$
	W_3	0	$E^4 \times A_2^2$
9	$A_4 \times C_2$	1	$E^3 \times A_2^3$
	W_2	0	$E_1 \times E_2^2 \times A_2^3$
	$A_5 \times C_2$	0	$E_1^4 \times E_2^5$
10	$\mathrm{SL}_2(3)$	1	$A_2^2 \times A_3^2$
11	$A_4 \times C_2$	2	$A_2 \times A_3^3$
	$S_4 \times C_2$	1	$E^3 \times A_{2,1} \times A_{2,2}^3$
12	$\mathrm{SL}_2(3)$	1	$A_2^2 \times A_4^2$
	W_3	0	$A_{2,1}^2 \times A_{2,2}^4$
13	$A_4 \times C_2$	2	$E \times A_{3,1} \times A_{3,2}^3$
14	$\mathrm{SL}_2(3)$	2	$A_3^2 \times A_4^2$
	$\mathrm{GL}_2(3)$	1	$A_2^4 \times A_3^2$
	$\mathrm{SL}_2(5)$	0	$E_1^4 \times E_2^6 \times A_2^2$
15	$A_4 \times C_2$	2	$A_2^3 \times A_3^3$
	$S_4 \times C_2$	1	$E_1 \times E_2^2 \times A_{2,1}^3 \times A_{2,2}^3$
	$A_5 \times C_2$	0	$E_1^4 \times E_2^5 \times A_2^3$
16	$\mathrm{SL}_2(3)$	2	$A_3^2 \times A_5^2$
17	$A_4 \times C_2$	3	$E \times A_{4,1} \times A_{4,2}^3$
	W_2	1	$E \times A_2^2 \times A_4^3$
18	$\mathrm{SL}_2(3)$	2	$A_3^2 \times A_6^2$
	$\mathrm{GL}_2(3)$	1	$A_{3,1}^2 \times A_{3,2}^4$
19	$A_4 \times C_2$	3	$E \times A_2^3 \times A_4^3$
20	$\mathrm{SL}_2(3)$	3	$A_4^2 \times A_6^2$
	W_3	1	$A_3^4 \times A_4^2$
	$\mathrm{SL}_2(5)$	0	$E^4 \times A_{2,1}^2 \times A_{2,2}^6$

Table 2. Jacobian variety decompositions. For each genus g and automorphism group G, we list the dimension of the moduli space of genus-g hyperelliptic curves with automorphism group G, along with a decomposition of the Jacobian of these curves. The notation is explained in Theorem 5.

The technique described in the previous section does not necessarily guarantee the finest decomposition of the Jacobian varieties. We have not ruled out the pos-

sibility that some of the abelian varieties $e(\pi_{i,j})J_X$ from (2) decompose further. In fact, in many cases there will be subfamilies where the decomposition is finer. However, for those curves in Table 2 which have affine models defined over $\mathbb{Q}$, we found a finite field where the factorization of the zeta function of that curve is no better than what our Jacobian decompositions predict. Hence, in those cases, the decomposition cannot be any finer, at least over $\mathbb{Q}$. Using ideas similar to those employed by Stoll [26, §2] one could show that, in fact, many of these decompositions cannot be refined even over the algebraic closure of $\mathbb{Q}$.

4.1. ***Finding monodromy and $\mathbb{Q}$-characters.*** The list of possible automorphism groups for hyperelliptic curves is well known, and most of these groups have easily identifiable character tables; thus, for hyperelliptic curves the most computationally difficult part of the technique summarized in Section 3 is finding the branching data. Breuer [3] developed an algorithm to generate a database of automorphism groups of Riemann surfaces, and he implemented this algorithm, up to genus 48, in the computer algebra package GAP [16]. Breuer's algorithm relies on the classifications of small groups in GAP. While the algorithm itself computes branching data, specific information about the monodromy was not recorded when Breuer originally ran the program.

We have now implemented in Magma [1] a version of Breuer's algorithm which *does* output the monodromy data. In cases below where the monodromy may not be obvious (for instance, if there is more than one conjugacy class of elements of a certain order for a particular automorphism group), our program provides the monodromy data.

We use Magma to compute the Hurwitz character χ_V and the inner product of χ_V with the irreducible $\mathbb{Q}$-characters. The $\mathbb{Q}$-character tables for the groups considered in this paper are well known in the literature so, alternatively, the computations could be done by hand.

4.2. ***Reduced automorphism group A_4.*** If a hyperelliptic curve has reduced automorphism group isomorphic to A_4, its full automorphism group is isomorphic to $\mathrm{SL}_2(3)$ or $A_4 \times C_2$. For $3 \leq g \leq 20$ the former group occurs in genus 4 and in all even genera greater than or equal to 8, while the latter group occurs in odd genera at least 5.

The group $\mathrm{SL}_2(3)$ has seven conjugacy classes. The identity, the unique element of order 2, and all the order-4 elements form three distinct conjugacy classes. The order-3 and order-6 elements each split into two conjugacy classes. The group ring $\mathbb{Q}[G]$ has Wedderburn decomposition

$$\mathbb{Q}[\mathrm{SL}_2(3)] \cong \mathbb{Q} \oplus \mathbb{Q}(\zeta_3) \oplus M_2(\mathbb{Q}) \oplus M_2(\mathbb{Q}(\zeta_3)) \oplus M_3(\mathbb{Q}).$$

	Conjugacy class order						
Character	1	2	3	3	4	6	6
χ_1	1	1	1	1	1	1	1
χ_2	2	2	−1	−1	2	−1	−1
χ_3	2	−2	−1	−1	0	1	1
χ_4	4	−4	1	1	0	−1	−1
χ_5	3	3	0	0	−1	0	0

Table 3. $\mathbb{Q}$-character table for $\mathrm{SL}_2(3)$.

So $\mathrm{SL}_2(3)$ has two $\mathbb{Q}$-characters of degree 1 (which we denote by χ_1 and χ_2), two of degree 2 (which we denote by χ_3 and χ_4), and one of degree 3 (which we denote by χ_5). The values of these characters on the conjugacy classes of $\mathrm{SL}_2(3)$ are well known [10, §38] and given in Table 3.

Recall from Section 2:

Theorem 1. *The hyperelliptic curve of genus* 4 *with affine model*

$$X: \quad y^2 = x(x^4 - 1)(x^4 + 2\sqrt{-3}\,x^2 + 1)$$

has a Jacobian variety that decomposes as $E_1^2 \times E_2^2$ *for two elliptic curves* E_i.

Everett Howe used an order-3 automorphism of X to compute that one of the factors of J_X (up to isogeny), say E_1, is given by E_1 with equation $y^2 = x^3 - 21x^2 + 12x + 8$.

Proof. Shaska [25, Tables 1 and 2, pp. 250, 252] shows that the curve X has automorphism group $\mathrm{SL}_2(3)$ and monodromy type $(4^{(6)}, 3^{(8)}, 6^{(4)})$. Thus the monodromy consists of elements g_1, g_2, and $g_3 \in \mathrm{SL}_2(3)$ of order 4, 3, and 6, respectively. As noted above, the six elements of order 4 are all in the same conjugacy class. Thus $\chi_{\langle g\rangle}$ (the induced character of the trivial character of the subgroup generated by $g \in G$) will be the same for all g of order 4, and likewise for the elements of order 3 and the elements of order 6, since all order-3 elements generate conjugate subgroups, as do the order-6 elements. Computing the Hurwitz character yields

$$\begin{aligned}\chi_V &= 2\chi_{\mathrm{triv}} - 2\chi_{\langle 1_G\rangle} + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_1\rangle}) + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_2\rangle}) + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_3\rangle}) \\ &= 2\chi_{\mathrm{triv}} + \chi_{\langle 1_G\rangle} - \chi_{\langle g_1\rangle} - \chi_{\langle g_2\rangle} - \chi_{\langle g_3\rangle}.\end{aligned}$$

The value of χ_V on conjugacy classes (listed in the same order as in Table 3) is the 7-tuple $(8, -8, -1, -1, 0, 1, 1)$. Computing the inner product of the irreducible $\mathbb{Q}$-characters with χ_V yields a value of 2 for each of the degree-2 characters and 0 for all the other characters. Applying (4) and Proposition 4 gives $J_X \sim E_1^2 \times E_2^2$. □

Similar results may be found for $g \geq 8$. See Section 5 for the generalization to arbitrary even genus.

The group $A_4 \times C_2$ has four irreducible $\mathbb{Q}$-characters of degree 1 and two of degree 3. For genus 5, the family of curves with affine model

$$X: \quad y^2 = x^{12} - ax^{10} - 33x^8 + 2ax^6 - 33x^4 - ax^2 + 1$$

has automorphism group $A_4 \times C_2$ and monodromy type $(3^{(8)}, 3^{(8)}, 2^{(12)}, 2^{(12)})$; see Shaska [25, Tables 1 and 2, pp. 250, 252]. We compute the Hurwitz character using the monodromy found through Breuer's algorithm, and then compute the inner products of the irreducible $\mathbb{Q}$-characters and the Hurwitz character. The inner product is 4 for one of the degree-1 characters and 2 for one of the degree-3 characters. By (4), the Jacobian variety of X decomposes into a 2-dimensional variety and three 1-dimensional varieties. Proposition 4 asserts that the three elliptic curves in this decomposition are isogenous to one another, so $J_X \sim A_2 \times E^3$ for some abelian surface A_2 and elliptic curve E.

Computations similar to those in the genus-5 case give the decompositions for higher odd genus described in Table 2.

4.3. ***Reduced automorphism group S_4.*** When a hyperelliptic curve has reduced automorphism group S_4, there are four options for its full automorphism group: $S_4 \times C_2$, $\mathrm{GL}_2(3)$, and the groups W_2 and W_3 defined at the beginning of this section. (The notation for the latter two groups of order 48 is as in [25].)

In genus 3, 11, and 15 there are curves with full automorphism group $S_4 \times C_2$. In [23], the Jacobian variety of the genus-3 curve was decomposed into the product of three isogenous elliptic curves. This result also appears in the literature using other techniques [21].

The decompositions of the families of genus-11 and genus-15 curves may be found using monodromy computed with Breuer's algorithm. The group $S_4 \times C_2$ has three irreducible $\mathbb{Q}$-characters of degree 1, two of degree 2, and three of degree 3. Combining this information with the technique in Section 3 yields the decompositions listed in Table 2.

As determined in [25], there is one genus-6 curve, up to isomorphism, with automorphism group $\mathrm{GL}_2(3)$:

$$X: \quad y^2 = x(x^4 - 1)(x^8 + 14x^4 + 1).$$

Additionally, there are 1-dimensional families of curves of genus 14 and 18 with this automorphism group.

The group $\mathrm{GL}_2(3)$ has two irreducible $\mathbb{Q}$-characters each of degrees 1, 2, and 3, as well as one of degree 4. In genus 6, the inner products of the irreducible $\mathbb{Q}$-characters with the Hurwitz character give values of 2 for one of the degree-2

characters and for the degree-4 character; from this we may conclude that $J_X \sim E_1^2 \times E_2^4$. Similar computations yield $J_X \sim A_3^2 \times A_4^2$ for the genus-14 curves and $J_X \sim A_{3,1}^2 \times A_{3,2}^4$ for the genus-18 curves.

For genus 5 and 9 there is one curve with automorphism group W_2, and in genus 17 there is a 1-dimensional family of curves with this automorphism group. In genus 5 the curve has an affine model

$$X: \ y^2 = x^{12} - 33x^8 - 33x^4 + 1,$$

in genus 9 a model is

$$X: \ y^2 = (x^8 + 14x^4 + 1)(x^{12} - 33x^8 - 33x^4 + 1),$$

and in genus 17 a model is

$$X: \ y^2 = (x^{12} - 33x^8 - 33x^4 + 1)\big(x^{24} + ax^{20} + (759 - 4a)x^{16} + 2(3a + 1288)x^{12} + (759 - 4a)x^8 + ax^4 + 1\big).$$

This group has eight irreducible $\mathbb{Q}$-characters: three of degree 1, two of degree 2, and three of degree 3. Computations with the genus-5 curve yield $J_X \sim E_1^2 \times E_2^3$, while for genus 9 we have $J_X \sim E_1 \times E_2^2 \times A_2^3$ and for genus 17, $J_X \sim E \times A_2^2 \times A_4^3$.

In genus 8 the curve with model

$$X: \ y^2 = x(x^4 - 1)(x^{12} - 33x^8 - 33x^4 + 1)$$

has automorphism group W_3 and monodromy type $(4^{(12)}, 3^{(16)}, 8^{(6)})$. The irreducible $\mathbb{Q}$-characters consist of two each of degrees 1, 2, and 3, as well as one of degree 4. Computations show the Jacobian of this curve decomposes as $A_2^2 \times E^4$. For higher-genus curves with this automorphism group, see the general results in Section 5.3.

In [22], in the course of considering different families of curves up to genus 10 we found a genus-8 curve with Jacobian decomposition $A_4 \times E_1^2 \times E_2^2$, so the result above is an improvement on our previous results on the bound on t from Question 1 in the introduction.

4.4. ***Reduced automorphism group A_5.*** As we see from Table 1, if a hyperelliptic curve has reduced automorphism group isomorphic to A_5, its full automorphism group is isomorphic to $A_5 \times C_2$ or $\mathrm{SL}_2(5)$. In genus 14 and 20 there is a hyperelliptic curve with automorphism group isomorphic to $\mathrm{SL}_2(5)$. This group has special properties that allow us to prove results about the decomposition of Jacobians generally for any genus. In Section 5.2 we discuss the general results.

Up to isomorphism, there is one curve of genus 5 with automorphism group $A_5 \times C_2$, one of genus 9, and one of genus 15. Here we prove the following result, which was mentioned in Section 2.

	Conjugacy class order									
Character	1	2	2	2	3	5	5	6	10	10
χ_1	1	1	1	1	1	1	1	1	1	1
χ_2	1	−1	1	−1	1	1	1	−1	−1	−1
χ_3	6	−6	−2	2	0	1	1	0	−1	−1
χ_4	6	6	−2	−2	0	1	1	0	1	1
χ_5	4	4	0	0	1	−1	−1	1	−1	1
χ_6	4	−4	0	0	1	−1	−1	−1	1	1
χ_7	5	5	1	1	−1	0	0	−1	0	0
χ_8	5	−5	1	−1	−1	0	0	1	0	0

Table 4. $\mathbb{Q}$-character table for $A_5 \times C_2$.

Theorem 2. *The genus-5 hyperelliptic curve with affine model*

$$X: \ y^2 = x(x^{10} + 11x^5 - 1)$$

has $J_X \sim E^5$ for the elliptic curve E with equation $y^2 = x(x^2 + 11x - 1)$.

Proof. We see from [25, §4.5] that the curve X has automorphism group $A_5 \times C_2$ and monodromy type $(3^{(40)}, 10^{(12)}, 2^{(60)})$ — although note that the coefficient 11 in the model given for X was misprinted in [25]. The irreducible $\mathbb{Q}$-characters of this group consist of two characters each of degrees 1, 3, 4, and 5. The monodromy consists of elements g_1, g_2, and $g_3 \in G$ of order 3, 10, and 2 respectively; this may be computed using Breuer's algorithm [3]. Table 4 gives the values of the irreducible $\mathbb{Q}$-characters on the conjugacy classes of $A_5 \times C_2$.

The Hurwitz character is

$$\begin{aligned}\chi_V &= 2\chi_{\text{triv}} - 2\chi_{\langle 1_G\rangle} + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_1\rangle}) + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_2\rangle}) + (\chi_{\langle 1_G\rangle} - \chi_{\langle g_3\rangle}) \\ &= 2\chi_{\text{triv}} + \chi_{\langle 1_G\rangle} - \chi_{\langle g_1\rangle} - \chi_{\langle g_2\rangle} - \chi_{\langle g_3\rangle}\end{aligned}$$

and its value on conjugacy classes (in the same order as Table 4) is given by the 10-tuple $(10, -10, 2, -2, -2, 0, 0, 2, 0, 0)$. The inner product of each of the irreducible $\mathbb{Q}$-characters with χ_V results in a value of 0 for all except one of the degree-5 characters, where the inner product is 2. By (4) and Proposition 4 this gives the desired decomposition. □

Applying this same idea to the genus-9 curve with affine model

$$X: \ y^2 = x^{20} - 228x^{15} + 494x^{10} - 228x^5 + 1$$

yields inner products with a value of 0 for all irreducible $\mathbb{Q}$-characters except for one degree-4 and one degree-5 character, where the inner product is 2. Again,

by (4) and Proposition 4, we find that J_X is isogenous to $E_1^4 \times E_2^5$, for elliptic curves E_i.

Similar computations in genus 15 for a curve with model

$$X: \quad y^2 = x(x^{10} + 11x - 1)(x^{20} - 228x^{15} + 494x^{10} - 228x^5 + 1)$$

yield the decomposition $J_X \sim E_1^4 \times E_2^5 \times A_2^3$.

5. General results

One obstacle to extending these results to higher genus is the computation of the monodromy for the cover $X \to X/G$. Beyond genus 48, Breuer's algorithm cannot currently compute the monodromy in many cases.

The groups $\mathrm{SL}_2(3)$, $\mathrm{SL}_2(5)$, and W_3 all share the following property: If X is a curve with automorphism group isomorphic to one of these groups, and if m is the order of any element of the monodromy of the cover X over X/G, then $\chi_{\langle g_i \rangle} = \chi_{\langle g_j \rangle}$ whenever $|g_i| = |g_j| = m$. We will denote this common character by $\chi_{(m)}$. Note that this property allows us to compute the Hurwitz character for X just by knowing the monodromy type. We then apply the technique from Section 3 to produce general decompositions for arbitrary genus.

Keep in mind that our technique does not necessarily guarantee the finest decomposition of the Jacobian variety. It is possible that for specific genera below the Jacobian decomposes further.

5.1. ***The group*** $\mathbf{SL_2(3)}$**.** Every even genus $g > 2$, except genus 6, has a hyperelliptic curve over k with automorphism group $\mathrm{SL}_2(3)$. For a given g, let $d = \lfloor (g-1)/6 \rfloor$, and let

$$G(x) = \prod_{i=1}^{d} (x^{12} - a_i x^{10} - 33x^8 + 2a_i x^6 - 33x^4 - a_i x^2 + 1),$$

where the a_i are distinct elements of k. Table 5 gives affine models and monodromy for curves of each even genus. These results may be found in [25]. Also recall the Wedderburn decomposition of $\mathbb{Q}[\mathrm{SL}_2(3)]$ and the irreducible characters of $\mathrm{SL}_2(3)$ from Section 4.2.

Computing the Hurwitz character given by (3) requires computing $\chi_{\langle g_i \rangle}$, the trivial character of $\langle g_i \rangle$ induced to $\mathrm{SL}_2(3)$, for each branched point g_i. The monodromy types listed in Table 5 give us the order of each branch point. As mentioned above, for this particular group, the order of the element is sufficient to compute the induced character. Table 6 lists the values of these induced characters on each conjugacy class.

Suppose X is a curve of genus g with automorphism group $\mathrm{SL}_2(3)$. Let $d = \lfloor (g-1)/6 \rfloor$ be as above. The computation of χ_V depends on the value of g mod 6.

g mod 6	Affine model	Monodromy
0	$y^2 = x(x^4-1)(x^8+14x^4+1)G(x)$	$(4^{(6)}, 6^{(4)}, 6^{(4)}, \underbrace{2^{(12)}, \ldots, 2^{(12)}}_{d})$
2	$y^2 = x(x^4-1)G(x)$	$(4^{(6)}, 3^{(8)}, 3^{(8)}, \underbrace{2^{(12)}, \ldots, 2^{(12)}}_{d})$
4	$y^2 = x(x^4-1)(x^4+2sx^2+1)G(x)$	$(4^{(6)}, 3^{(8)}, 6^{(4)}, \underbrace{2^{(12)}, \ldots, 2^{(12)}}_{d})$

Table 5. Hyperelliptic curves with automorphism group $\mathrm{SL}_2(3)$. For each even genus $g > 2$, we give a model for the generic hyperelliptic curve of genus g with automorphism group $\mathrm{SL}_2(3)$, together with its monodromy. Here $d = \lfloor (g-1)/6 \rfloor$, $s^2 = -3$, and $G(x)$ is as defined at the beginning of Section 5.1.

- Suppose $g \equiv 2 \bmod 6$. Applying the monodromy information given in Table 5 to (3) yields

$$\chi_V = 2\chi_{\mathrm{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - 2\chi_{(3)} - d\chi_{(2)}.$$

  Computing the inner product of each irreducible $\mathbb{Q}$-character (see Table 3) with χ_V gives $J_X \sim A_{d+1}^2 \times A_{2d}^2$.
- Suppose $g \equiv 4 \bmod 6$. Applying the monodromy information from Table 5, we find that

$$\chi_V = 2\chi_{\mathrm{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(6)} - \chi_{(3)} - d\chi_{(2)}.$$

  This gives $J_X \sim A_{d+1}^2 \times A_{2d+1}^2$.
- Finally, suppose $g \equiv 0 \bmod 6$. Using Table 5, we compute that

$$\chi_V = 2\chi_{\mathrm{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - 2\chi_{(6)} - d\chi_{(2)}.$$

  This gives $J_X \sim A_{d+1}^2 \times A_{2(d+1)}^2$.

5.2. ***The group*** $\mathbf{SL_2(5)}$**.** If g is congruent to 0, 14, 20, or 24 modulo 30 there is a hyperelliptic curve of genus g with automorphism group $\mathrm{SL}_2(5)$. Letting

	Conjugacy class order						
Character	1	2	3	3	4	6	6
$\chi_{(2)}$	12	12	0	0	0	0	0
$\chi_{(3)}$	8	0	2	2	0	0	0
$\chi_{(4)}$	6	6	0	0	2	0	0
$\chi_{(6)}$	4	4	1	1	0	1	1

Table 6. Induced characters for $\mathrm{SL}_2(3)$.

$d = \lfloor (g-1)/30 \rfloor$, the moduli space of such hyperelliptic curves has dimension d, and can be described as follows (see [25, §4.5]): Given d elements $a_1, \ldots, a_d$ of k, set

$$\begin{aligned} G_i(x) = (a_i - 1)x^{60} &- 36(19a_i + 29)x^{55} + 6(26239a_i - 42079)x^{50} \\ &- 540(23199a_i - 19343)x^{45} + 105(737719a_i - 953143)x^{40} \\ &- 72(1815127a_i - 145087)x^{35} - 4(8302981a_i + 49913771)x^{30} \\ &+ 72(1815127a_i - 145087)x^{25} + 105(737719a_i - 953143)x^{20} \\ &+ 540(23199a_i - 19343)x^{15} + 6(26239a_i - 42079)x^{10} \\ &+ 36(19a_i + 29)x^5 + (a_i - 1) \end{aligned}$$

and

$$\begin{aligned} G(x) &= \prod_{i=1}^{d} G_i(x) \\ F(x) &= x^{30} + 522x^{25} - 10005x^{20} - 10005x^{10} - 522x^5 + 1 \\ H(x) &= x^{20} - 228x^{15} + 494x^{10} + 228x^5 + 1 \\ K(x) &= x(x^{10} + 11x^5 - 1). \end{aligned}$$

Then Table 7 lists models and monodromy for the genus-g hyperelliptic curves with automorphism group $\mathrm{SL}_2(5)$, depending on the congruence class of the genus modulo 30.

Again, the induced characters depend only upon the order of the element generating the subgroup. The values for these induced characters on the conjugacy

g mod 30	Affine model	Monodromy
0	$y^2 = K(x)H(x)F(x)G(x)$	$(4^{(30)}, 6^{(20)}, 10^{(12)}, \underbrace{2^{(60)}, \ldots, 2^{(60)}}_{d})$
14	$y^2 = F(x)G(x)$	$(4^{(30)}, 3^{(40)}, 5^{(24)}, \underbrace{2^{(60)}, \ldots, 2^{(60)}}_{d})$
20	$y^2 = K(x)F(x)G(x)$	$(4^{(30)}, 3^{(40)}, 10^{(12)}, \underbrace{2^{(60)}, \ldots, 2^{(60)}}_{d})$
24	$y^2 = H(x)F(x)G(x)$	$(4^{(30)}, 6^{(20)}, 5^{(24)}, \underbrace{2^{(60)}, \ldots, 2^{(60)}}_{d})$

Table 7. Hyperelliptic curves with automorphism group $\mathrm{SL}_2(5)$. For each genus g congruent to 0, 14, 20, or 24 modulo 30, we give a model for the generic hyperelliptic curve of genus g with automorphism group $\mathrm{SL}_2(5)$, together with its monodromy. Here $d = \lfloor (g-1)/30 \rfloor$, and the polynomials $F(x)$, $G(x)$, $H(x)$, and $K(x)$ are as defined at the beginning of Section 5.2.

	Conjugacy class order								
Character	1	2	3	4	5	5	6	10	10
$\chi_{(2)}$	60	60	0	0	0	0	0	0	0
$\chi_{(3)}$	40	0	4	0	0	0	0	0	0
$\chi_{(4)}$	30	30	0	2	0	0	0	0	0
$\chi_{(5)}$	24	0	0	0	4	4	0	0	0
$\chi_{(6)}$	20	20	2	0	0	0	2	0	0
$\chi_{(10)}$	12	12	0	0	2	2	0	2	2

Table 8. Induced characters for $\mathrm{SL}_2(5)$.

classes are listed in Table 8. The group ring for this group is

$$\mathbb{Q}[\mathrm{SL}_2(5)] \cong \mathbb{Q} \oplus M_2(\mathbb{Q}(\sqrt{5})) \oplus M_3(\mathbb{Q}(\sqrt{5})) \oplus M_4(\mathbb{Q}) \oplus M_4(\mathbb{Q}) \oplus M_5(\mathbb{Q}) \oplus M_6(\mathbb{Q}).$$

Computing the inner products of the irreducible $\mathbb{Q}$-characters (which are well known [10, §38]) with χ_V (listed below for the four congruence classes of g) produces decompositions of the form $A_{2(d+1)}^2 \times A_j^4 \times A_k^6$, where d, j, and k are determined by the congruence class of g modulo 30, and where $d = \lfloor (g-1)/30 \rfloor$ is the dimension of the family of curves with this automorphism group.

- Suppose $g \equiv 14 \bmod 30$. Then the Hurwitz character is
$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(3)} - \chi_{(5)} - d\chi_{(2)},$$
and we have $j = 2d+1$ and $k = 3d+1$.
- Suppose $g \equiv 20 \bmod 30$. Then the Hurwitz character is
$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(3)} - \chi_{(10)} - d\chi_{(2)},$$
and we have $j = 2d+1$ and $k = 3d+2$.
- Suppose $g \equiv 24 \bmod 30$. Then the Hurwitz character is
$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(6)} - \chi_{(5)} - d\chi_{(2)},$$
and we have $j = 2(d+1)$ and $k = 3d+2$.
- Finally, suppose $g \equiv 0 \bmod 30$. Then the Hurwitz character is
$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(6)} - \chi_{(10)} - d\chi_{(2)},$$
so $j = 2(d+1)$ and $k = 3(d+1)$.

g mod 12	Affine model	Monodromy
0	$y^2 = (x^8 + 14x^4 + 1)H(x)G(x)$	$(4^{(12)}, 6^{(8)}, 8^{(6)}, \underbrace{2^{(24)}, \ldots, 2^{(24)}}_{d})$
8	$y^2 = H(x)G(x)$	$(4^{(12)}, 3^{(16)}, 8^{(6)}, \underbrace{2^{(24)}, \ldots, 2^{(24)}}_{d})$

Table 9. Hyperelliptic curves with automorphism group W_3. For each genus g congruent to 0 or 8 modulo 12, we give a model for the generic hyperelliptic curve of genus g with automorphism group W_3, together with its monodromy. Here $d = \lfloor (g-1)/12 \rfloor$, and the polynomials $G(x)$ and $H(x)$ are as defined at the beginning of Section 5.3.

5.3. ***The group W_3.*** When g is congruent to 0 or 8 modulo 12, there is a curve of genus g with automorphism group W_3. Models for these curves and their monodromy are listed in Table 9, where we use the notation $d = \lfloor (g-1)/12 \rfloor$,

$$G(x) = \prod_{i=1}^{d} \big(x^{24} + a_i x^{20} + (759 - 4a_i)x^{16} + 2(3a_i + 1288)x^{12} + (759 - 4a_i)x^8 + a_i x^4 + 1\big),$$

and $H(x) = x(x^4 - 1)(x^{12} - 33x^8 - 33x^4 + 1)$. Again, explanations of these models and monodromy can be found in [25].

The group W_3 has seven irreducible $\mathbb{Q}$-characters: two each of degrees 1, 2, and 3, and one of degree 4. The group ring decomposes as follows:

$$\mathbb{Q}[W_3] \cong \mathbb{Q} \oplus \mathbb{Q} \oplus M_2(\mathbb{Q}) \oplus M_2(\mathbb{Q}(\sqrt{2})) \oplus M_3(\mathbb{Q}) \oplus M_3(\mathbb{Q}) \oplus M_4(\mathbb{Q}).$$

As in the previous two cases, there is only one possible value for the induced character, except for the characters induced from subgroups generated by order-4 elements. However, only certain order-4 elements show up in the monodromy and they all have the same induced character. The values for these induced characters on the conjugacy classes are listed in Table 10.

	Conjugacy class order							
Character	1	2	3	4	4	6	8	8
$\chi_{(2)}$	24	24	0	0	0	0	0	0
$\chi_{(3)}$	16	0	4	0	0	0	0	0
$\chi_{(4)}$	12	12	0	2	0	0	0	0
$\chi_{(6)}$	8	8	2	0	0	2	0	0
$\chi_{(8)}$	6	6	0	2	0	0	2	2

Table 10. Induced characters for W_3.

We compute the decomposition of the Jacobian in the two cases as follows:

- When $g \equiv 8 \bmod 12$, the Hurwitz character is

$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(3)} - \chi_{(8)} - d\chi_{(2)}$$

  and $J_X \sim A^2_{2(d+1)} \times A^4_{2d+1}$.
- When $g \equiv 0 \bmod 12$, the Hurwitz character is

$$\chi_V = 2\chi_{\text{triv}} + (d+1)\chi_{(1)} - \chi_{(4)} - \chi_{(6)} - \chi_{(8)} - d\chi_{(2)}$$

  and $J_x = A^2_{2(d+1),1} \times A^4_{2(d+1),2}$.

Acknowledgments

The author would like to thank the anonymous referees for their helpful suggestions, including pointing out a hitherto unknown word usage issue, and Jordan Ellenberg and Everett Howe for helpful discussions related to this work. The author also appreciates useful comments during the ANTS X conference from Nils Bruin, Noam Elkies, Kiran Kedlaya, and John Voight.

References

[1] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system, I: The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[2] Rolf Brandt and Henning Stichtenoth, *Die Automorphismengruppen hyperelliptischer Kurven*, Manuscripta Math. **55** (1986), no. 1, 83–92. MR 87m:14033

[3] Thomas Breuer, *Characters and automorphism groups of compact Riemann surfaces*, London Mathematical Society Lecture Note Series, no. 280, Cambridge University Press, 2000. MR 2002i:14034

[4] E. Bujalance, J. M. Gamboa, and G. Gromadzki, *The full automorphism groups of hyperelliptic Riemann surfaces*, Manuscripta Math. **79** (1993), no. 3-4, 267–282. MR 94f:20093

[5] Gabriel Cardona, *$\mathbb{Q}$-curves and abelian varieties of GL_2-type from dihedral genus 2 curves*, in Cremona et al. [6], 2004, pp. 45–52. MR 2005b:11075

[6] John Cremona, Joan-Carles Lario, Jordi Quer, and Kenneth Ribet (eds.), *Modular curves and abelian varieties: Papers from the conference held in Bellaterra, July* 15–18, 2002, Progress in Mathematics, no. 224, Birkhäuser, Basel, 2004. MR 2004k:11004

[7] Leonard Eugene Dickson, *Invariants of binary forms under modular transformations*, Trans. Amer. Math. Soc. **8** (1907), no. 2, 205–232, errata: [8]. MR 1500782

[8] ———, *Errata: "Invariants of binary forms under modular transformations" [Trans. Amer. Math. Soc.* 8 (1907), *no.* 2, 205–232; 1500782*]*, Trans. Amer. Math. Soc. **8** (1907), no. 4, 535. MR 1500482

[9] Claus Diem and Jasper Scholten, *Ordinary elliptic curves of high rank over $\overline{\mathbb{F}}_p(x)$ with constant j-invariant, II*, J. Number Theory **124** (2007), no. 1, 31–41. MR 2008b:11063

[10] Larry Dornhoff, *Group representation theory, part A: Ordinary representation theory*, Pure and Applied Mathematics, no. 7, Marcel Dekker, New York, 1971. MR 50 #458a

[11] David S. Dummit and Richard M. Foote, *Abstract algebra*, Prentice Hall, Upper Saddle River, NJ, 1999.

[12] Iwan Duursma and Negar Kiyavash, *The vector decomposition problem for elliptic and hyperelliptic curves*, J. Ramanujan Math. Soc. **20** (2005), no. 1, 59–76. MR 2006b:14038

[13] Torsten Ekedahl and Jean-Pierre Serre, *Exemples de courbes algébriques à jacobienne complètement décomposable*, C. R. Acad. Sci. Paris Sér. I Math. **317** (1993), no. 5, 509–513. MR 94j:14029

[14] Noam D. Elkies, Everett W. Howe, and Christophe Ritzenthaler, *Genus bounds for curves with fixed Frobenius eigenvalues*, 2010. arXiv 1006.0822 [math.NT]

[15] Jordan S. Ellenberg, *Endomorphism algebras of Jacobians*, Adv. Math. **162** (2001), no. 2, 243–271. MR 2003c:11061

[16] The GAP Group, *GAP — Groups, Algorithms, and Programming* (*version* 4.4), 2006. http://www.gap-system.org

[17] Josep González, *Fermat Jacobians of prime degree over finite fields*, Canad. Math. Bull. **42** (1999), no. 1, 78–86. MR 2000h:11065

[18] Hoon Hong (ed.), *ISSAC* 2003—*Proceedings of the* 2003 *International Symposium on Symbolic and Algebraic Computation held in Philadelphia, PA, August* 3–6, 2003, Association for Computing Machinery (ACM), New York, 2003. MR 2004j:68018

[19] Everett W. Howe, Franck Leprévost, and Bjorn Poonen, *Large torsion subgroups of split Jacobians of curves of genus two or three*, Forum Math. **12** (2000), no. 3, 315–364. MR 2001e:11071

[20] E. Kani and M. Rosen, *Idempotent relations and factors of Jacobians*, Math. Ann. **284** (1989), no. 2, 307–327. MR 90h:14057

[21] Masato Kuwata, *Quadratic twists of an elliptic curve and maps from a hyperelliptic curve*, Math. J. Okayama Univ. **47** (2005), 85–97. MR 2006i:11061

[22] Jennifer Paulhus, *Decomposing Jacobians of curves with extra automorphisms*, Acta Arith. **132** (2008), no. 3, 231–244. MR 2009c:14049

[23] Jennifer R. Paulhus, *Elliptic factors in Jacobians of low genus curves*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2007. http://search.proquest.com/docview/304849148

[24] Karl Rubin and Alice Silverberg, *Rank frequencies for quadratic twists of elliptic curves*, Experiment. Math. **10** (2001), no. 4, 559–569. MR 2002k:11081

[25] Tanush Shaska, *Determining the automorphism group of a hyperelliptic curve*, in Hong [18], 2003, pp. 248–254. MR 2005c:14037

[26] Michael Stoll, *Two simple* 2*-dimensional abelian varieties defined over* $\mathbf{Q}$ *with Mordell-Weil group of rank at least* 19, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 10, 1341–1345. MR 96j:11084

[27] Takuya Yamauchi, *On* $\mathbb{Q}$*-simple factors of Jacobian varieties of modular curves*, Yokohama Math. J. **53** (2007), no. 2, 149–160. MR 2008k:11062

[28] Noriko Yui, *On the Jacobian variety of the Fermat curve*, J. Algebra **65** (1980), no. 1, 1–35. MR 82m:14016

JENNIFER PAULHUS: paulhusj@grinnell.edu
Department of Mathematics and Statistics, Grinnell College, Grinnell, IA 50112, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.507

msp

Isogeny volcanoes

Andrew V. Sutherland

The remarkable structure and computationally explicit form of isogeny graphs of elliptic curves over a finite field have made these graphs an important tool for computational number theorists and practitioners of elliptic curve cryptography. This expository paper recounts the theory behind isogeny graphs and examines several recently developed algorithms that realize substantial (and often dramatic) performance gains by exploiting this theory.

1. Introduction

A *volcano* is a certain type of graph, one whose shape reminds us of the geological formation of the same name. A typical volcano consists of a cycle with isomorphic balanced trees rooted at each vertex.

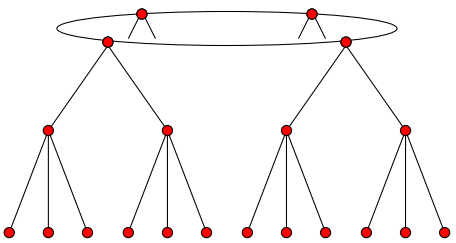


Figure 1. A volcano.

More formally, let ℓ be a prime. We define an *ℓ-volcano* as follows.

Definition 1. An *ℓ-volcano* V is a connected undirected graph whose vertices are partitioned into one or more *levels* $V_0, \ldots, V_d$ such that the following hold:

(1) The subgraph on V_0 (the *surface*) is a regular graph of degree at most 2.

MSC2010: primary 11G07, 11Y16; secondary 11G15, 11G20.
Keywords: elliptic curves, isogeny graphs.

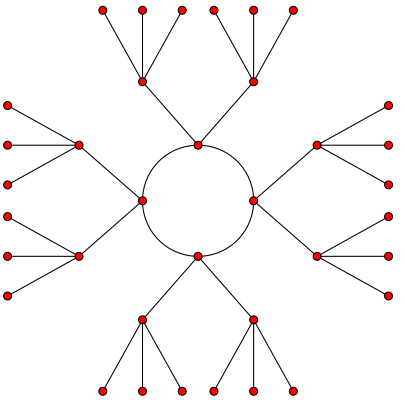


Figure 2. A 3-volcano of depth 2.

(2) For $i > 0$, each vertex in V_i has exactly one neighbor in level V_{i-1}, and this accounts for every edge not on the surface.

(3) For $i < d$, each vertex in V_i has degree $\ell + 1$.

Self-loops and multi-edges are permitted in an ℓ-volcano, but it follows from condition (2) that these can only occur on the surface. The integer d is the *depth* of the volcano (some authors use the term *height*). When $d = 0$ only condition (1) applies, and in this case V is a connected regular graph of degree at most 2. Such a graph is either a single vertex with up to two self-loops, two vertices connected by one or two edges, or a simple cycle on three or more vertices (the general case). Figure 2 gives an overhead view of the volcano depicted in Figure 1, a 3-volcano of depth 2.

We have defined volcanoes in purely graph-theoretic terms, but we are specifically interested in volcanoes that arise as components of graphs of isogenies between elliptic curves. Our first objective is to understand how and why volcanoes arise in such graphs. The definitive work in this area was done by David Kohel, whose thesis explicates the structure of isogeny graphs of elliptic curves over finite fields [30]. The term "volcano" came later, in work by Fouquet and Morain [16; 17] that popularized Kohel's work and gave one of the first examples of how isogeny volcanoes could be exploited by algorithms that work with elliptic curves.

This leads to our second objective: to show how isogeny volcanoes can be used to develop better algorithms. We illustrate this with four examples of algorithms that use isogeny volcanoes to solve some standard computational problems related to elliptic curves over finite fields. In each case, the isogeny volcano approach yields a substantial practical and asymptotic improvement over the best previous results.

2. Isogeny graphs of elliptic curves

We begin by recalling some basic facts about elliptic curves and isogenies, all of which can be found in standard references such as [31; 41; 42].

2.1. ***Elliptic curves.*** Let k be a field. An *elliptic curve* E/k is a smooth projective curve of genus 1 over k, together with a distinguished k-rational point 0. If k'/k is any field extension, the set $E(k')$ of k'-rational points of E forms an abelian group with 0 as its identity element. For convenience we assume that the characteristic of k is neither 2 nor 3, in which case every elliptic curve E/k can be written as the projective closure of a short Weierstrass equation of the form

$$Y^2 = X^3 + aX + b,$$

where the coefficients $a, b \in k$ satisfy $4a^3+27b^2 \neq 0$; here the distinguished point 0 is taken to be the "point at infinity" on the projective closure. Distinct Weierstrass equations may define isomorphic curves: The curves defined by $Y^2 = X^3+a_1X+b_1$ and $Y^2 = X^3+a_2X+b_2$ are isomorphic to one another over the algebraic closure $\bar{k}$ of k if and only if $a_2 = u^4a_1$ and $b_2 = u^6b_1$ for some $u \in \bar{k}$; the isomorphism is then defined over the field $k(u)$. It follows that the quantity

$$j(a,b) = 1728\frac{4a^3}{4a^3+27b^2}$$

depends only on the $\bar{k}$-isomorphism class of E, so we may define the *j-invariant* $j(E)$ of E to be $j(a,b)$ for any model $Y^2 = X^3 + aX + b$ of E. Note that while $j(E)$ lies in k, it only determines the isomorphism class of E over the algebraic closure $\bar{k}$. Elliptic curves with the same j-invariant need not be isomorphic to one another over k; such curves are said to be *twists* of each other.

Every $j \in k$ arises as the j-invariant of an elliptic curve E/k: We have $0 = j(0,b)$ and $1728 = j(a,0)$, while if $j \neq 0, 1728$ we can take

$$a = 3j(1728-j) \quad \text{and} \quad b = 2j(1728-j)^2,$$

and we find that $j = j(a,b)$. There is thus a one-to-one correspondence between the field k and the set of $\bar{k}$-isomorphism classes of elliptic curves over k. This is the vertex set of the isogeny graphs that we wish to define.

An *automorphism* of an elliptic curve E is an automorphism of E as a curve that fixes the identity element 0. Most elliptic curves have automorphism groups of order 2, with the nontrivial automorphism being the map $(X,Y) \mapsto (X,-Y)$; the only exceptions are the elliptic curves with j-invariants equal to 0 and 1728, which may have extra automorphisms. To simplify matters we will occasionally exclude these special cases from consideration.

2.2. ***Isogenies.*** Let E_1 and E_2 be elliptic curves over a field k. An *isogeny* $\varphi : E_1 \to E_2$ is a nonzero morphism of elliptic curves, that is, a nonconstant rational map that takes the identity of E_1 to the identity of E_2. (We do not require that the morphism be defined over k; we allow maps defined over the algebraic closure.) The *degree* of an isogeny is its degree as a rational map. We call an isogeny of degree n an *n-isogeny*. Elliptic curves related by an isogeny of degree n are said to be *n-isogenous*. We say that two elements j_1, j_2 of k are *n-isogenous* over k if there are n-isogenous elliptic curves E_1, E_2 over k with $j(E_1) = j_1$ and $j(E_2) = j_2$. For a given E/k, if one thinks of $j(E)$ as representing the set of twists of E, then saying that $j(E_1)$ and $j(E_2)$ are n-isogenous means that one can choose twists of E_1 and E_2 that are n-isogenous. Over an algebraically closed field, the set of twists is trivial, so the choice of twist is easy; but even over non-algebraically closed fields, it is easy in practice to find compatible twists.

Every isogeny $\varphi : E_1 \to E_2$ induces a surjective group homomorphism from $E_1(\bar{k})$ to $E_2(\bar{k})$ that has a finite kernel; in this paper, when we speak of the *kernel* of an isogeny, we will always mean the set of points in the kernel over $\bar{k}$. The kernel of an n-isogeny typically has cardinality n (in which case the isogeny is said to be *separable*), and this is always the case when n is not divisible by the characteristic of k. We are primarily interested in isogenies of prime degree $\ell \neq \operatorname{char} k$, and we shall only distinguish isogenies up to isomorphism, regarding isogenies ϕ and φ as equivalent if $\phi = \iota \circ \varphi \circ \iota'$ for some isomorphisms ι and ι'.

There are two important facts about isogenies that we need. The first is that every finite subgroup of $E_1(\bar{k})$ is the kernel of a separable isogeny over $\bar{k}$ that is uniquely determined (up to isomorphism) [41, Proposition III.4.12], and this isogeny can be explicitly computed using Vélu's algorithm [48]. The second is that every n-isogeny $\varphi : E_1 \to E_2$ has a unique *dual isogeny* $\widehat{\varphi} : E_2 \to E_1$ that satisfies

$$\varphi \circ \widehat{\varphi} = \widehat{\varphi} \circ \varphi = [n],$$

where $[n]$ is the *multiplication-by-n map* that sends $P \in E_1(\bar{k})$ to $nP = P + \cdots + P$; see [41, Theorem III.6.1]. The dual isogeny $\widehat{\varphi}$ has degree n, and $[n]$ has degree n^2.

The kernel of the multiplication-by-n map is the *n-torsion subgroup*

$$E[n] = \{P \in E(\bar{k}) : nP = 0\},$$

and for n not divisible by the characteristic of k we have

$$E[n] \simeq \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

For primes $\ell \neq \operatorname{char} k$, there are $\ell + 1$ cyclic subgroups in $E[\ell]$ of order ℓ, each of which is the kernel of a separable ℓ-isogeny (over $\bar{k}$). Every ℓ-isogeny φ from E arises in this way, since any point in the kernel of φ also lies in the kernel of $\widehat{\varphi} \circ \varphi = [\ell]$.

Not every cyclic subgroup of $E[\ell]$ is the kernel of an isogeny defined over k; this occurs precisely when the subgroup is invariant under the action of the Galois group $G = \mathrm{Gal}(k(E[\ell])/k)$. The Galois group acts linearly on $E[\ell] \simeq \mathbb{Z}/\ell\mathbb{Z} \times \mathbb{Z}/\ell\mathbb{Z}$, which we may view as an $\mathbb{F}_\ell$-vector space of dimension two in which the order-ℓ subgroups of $E[\ell]$ are linear subspaces. If G fixes more than two linear subspaces of a two-dimensional vector space then it must fix all of them. This yields the following lemma.

Lemma 2. *Let E/k be an elliptic curve with j-invariant not equal to* 0 *or* 1728, *and let $\ell \neq \mathrm{char}\, k$ be a prime. Up to isomorphism, the number of k-rational ℓ-isogenies from E is* $0, 1, 2,$ *or* $\ell + 1$.

2.3. ***The modular equation.*** Let $j(\tau)$ be the classical modular function defined on the upper half plane $\mathbb{H}$. For any $\tau \in \mathbb{H}$, the complex numbers $j(\tau)$ and $j(N\tau)$ are the j-invariants of elliptic curves defined over $\mathbb{C}$ that are related by an isogeny whose kernel is a cyclic group of order N. The minimal polynomial $\Phi_N(Y)$ of the function $j(Nz)$ over the field $\mathbb{C}(j(z))$ has coefficients that are integer polynomials in $j(z)$. If we replace $j(z)$ with X we obtain the *modular polynomial* $\Phi_N \in \mathbb{Z}[X, Y]$, which is symmetric in X and Y and has degree $\ell + 1$ in both variables. It parametrizes pairs of elliptic curves over $\mathbb{C}$ related by a cyclic N-isogeny. The *modular equation* $\Phi_N(X, Y) = 0$ is a canonical equation for the modular curve $Y_0(N) = \Gamma_0(N)\backslash\mathbb{H}$.

When N is a prime ℓ, every N-isogeny is cyclic, and we have

$$\Phi_\ell\big(j(E_1), j(E_2)\big) = 0 \quad \Longleftrightarrow \quad j(E_1) \text{ and } j(E_2) \text{ are } \ell\text{-isogenous.}$$

This moduli interpretation remains valid over every field, even those of positive characteristic.

2.4. ***The graph of ℓ-isogenies.*** We now use the modular equation to define the graph of ℓ-isogenies over a field k of characteristic different from ℓ.

Definition 3. The *ℓ-isogeny graph* $G_\ell(k)$ is the directed graph with vertex set k and edges (j_1, j_2) present with multiplicity equal to the multiplicity of j_2 as a root of $\Phi_\ell(j_1, Y)$.

The vertices of $G_\ell(k)$ are j-invariants, and its edges correspond to (isomorphism classes of) ℓ-isogenies. Every edge (j_1, j_2) that is not incident to 0 or 1728 occurs with the same multiplicity as (j_2, j_1). Thus the subgraph of $G_\ell(k)$ on $k\backslash\{0, 1728\}$ is bidirected, and we may view it as an undirected graph. For any fixed k, the graphs $G_\ell(k)$ all have the same vertex set, but different edge sets, depending on ℓ. Given an elliptic curve E/k, we may view $j(E)$ as a vertex in any of these graphs, a fact that has many applications.

2.5. ***Supersingular and ordinary components.*** Over a field of positive characteristic p, an elliptic curve is *supersingular* if its p-torsion subgroup $E[p]$ is trivial; otherwise it is *ordinary*. If E is supersingular, then so is any elliptic curve isogenous to E; therefore $G_\ell(k)$ is composed of ordinary and supersingular components.

Every supersingular curve over k can be defined over a quadratic extension of the prime field of k; thus every supersingular j-invariant in $\bar{k}$ lies in $\mathbb{F}_{p^2}$ [41, Theorem V.3.1]. It follows that if E is supersingular, then the roots of $\Phi_\ell(j(E), Y)$ all lie in $\mathbb{F}_{p^2}$. Thus every vertex in a supersingular component of $G_\ell(\mathbb{F}_{p^2})$ has out-degree $\ell + 1$. (Every vertex other than those equal to or adjacent to 0 or 1728 also has in-degree $\ell + 1$.)

Remark 4. *Ramanujan graphs.* In fact, $G_\ell(\mathbb{F}_{p^2})$ has just one supersingular component [30, Corollary 78], and when $p \equiv 1 \bmod 12$ it is a *Ramanujan graph* [35], an expander graph with an essentially optimal expansion factor. This fact has cryptographic applications [10].

We are primarily interested in the ordinary components of $G_\ell(k)$, since this is where we will find isogeny volcanoes. First we need to recall some facts from the theory of complex multiplication.

2.6. ***Complex multiplication.*** A morphism from an elliptic curve E/k to itself is called an *endomorphism*; an endomorphism of E is either the zero map or an isogeny from E to itself. (We do not require that endomorphisms be defined over the base field k.) The endomorphisms of an elliptic curve E form a ring $\mathrm{End}(E)$ in which addition and multiplication are defined via the formulas

$$(\phi + \varphi)(P) = \phi(P) + \varphi(P) \quad \text{and} \quad (\phi\varphi)(P) = \phi(\varphi(P)) \quad \text{for all } P \in E(\bar{k}).$$

For every positive integer n, the multiplication-by-n map $[n]$ lies in $\mathrm{End}(E)$, and we have $[n]\phi = \phi + \cdots + \phi = n\phi$ for all $\phi \in \mathrm{End}(E)$. Since $[n]$ is never the zero endomorphism, it follows that $\mathrm{End}(E)$ contains a subring isomorphic to $\mathbb{Z}$, which we shall identify with $\mathbb{Z}$.

When $\mathrm{End}(E)$ is larger than $\mathbb{Z}$ we say that E has *complex multiplication* (CM), a term that arises from the fact that over the complex numbers, endomorphisms that do not lie in $\mathbb{Z}$ may be viewed as "multiplication-by-α" maps for some algebraic integers α. Over a finite field $\mathbb{F}_q$, every elliptic curve has complex multiplication; for ordinary elliptic curves over $\mathbb{F}_q$, the Frobenius endomorphism that sends the point (X, Y) to (X^q, Y^q) is an example of an endomorphism that does not lie in $\mathbb{Z}$.

When E has complex multiplication there are two possibilities:

$$\mathrm{End}(E) \simeq \begin{cases} \text{an order } \mathbb{O} \text{ in an imaginary quadratic field, or} \\ \text{an order } \mathbb{O} \text{ in a definite quaternion algebra,} \end{cases}$$

and in either case we say that E has CM by $\mathcal{O}$. The second case occurs if and only if E is supersingular, which is possible only in positive characteristic; we are primarily interested in the first case. It will be convenient to fix an isomorphism $\mathcal{O} \xrightarrow{\sim} \mathrm{End}(E)$ so that we may regard elements of $\mathcal{O}$ as elements of $\mathrm{End}(E)$ and vice versa.

The *endomorphism algebra* $\mathrm{End}^0(E) = \mathrm{End}(E) \otimes \mathbb{Q}$ is preserved by isogenies. Thus if E has complex multiplication, then so does every elliptic curve isogenous to E, but not necessarily by the same order $\mathcal{O}$.

2.7. ***Horizontal and vertical isogenies.*** Let $\varphi : E_1 \to E_2$ be an ℓ-isogeny of elliptic curves with CM by imaginary quadratic orders $\mathcal{O}_1$ and $\mathcal{O}_2$, respectively. Then $\mathcal{O}_1 = \mathbb{Z} + \tau_1\mathbb{Z}$ and $\mathcal{O}_2 = \mathbb{Z} + \tau_2\mathbb{Z}$, for some $\tau_1, \tau_2 \in \mathbb{H}$. The isogeny $\widehat{\varphi} \circ \tau_2 \circ \varphi$ lies in $\mathrm{End}(E_1)$, and this implies that $\ell\tau_2 \in \mathcal{O}_1$; similarly, $\ell\tau_1 \in \mathcal{O}_2$. There are thus three possibilities:

(1) $\mathcal{O}_1 = \mathcal{O}_2$, in which case we say that φ is *horizontal*.

(2) $[\mathcal{O}_1 : \mathcal{O}_2] = \ell$, in which case we say that φ is *descending*.

(3) $[\mathcal{O}_2 : \mathcal{O}_1] = \ell$, in which case we say that φ is *ascending*.

In the last two cases we say that φ is a *vertical* ℓ-isogeny. The orders $\mathcal{O}_1$ and $\mathcal{O}_2$ necessarily have the same fraction field $K = \mathrm{End}^0(E_1) = \mathrm{End}^0(E_2)$, and both lie in the maximal order $\mathcal{O}_K$, the ring of integers of K.

2.8. ***The CM torsor.*** Let E/k be an elliptic curve with CM by an imaginary quadratic order $\mathcal{O}$, and let $\mathfrak{a}$ be an invertible $\mathcal{O}$-ideal. The $\mathfrak{a}$-*torsion subgroup*

$$E[\mathfrak{a}] = \{P \in E(\bar{k}) : \alpha(P) = 0 \text{ for all } \alpha \in \mathfrak{a}\}$$

is the kernel of a separable isogeny $\varphi_\mathfrak{a} : E \to E'$. Provided that $\mathfrak{a}$ has norm not divisible by the characteristic of k, we have $\deg \varphi_\mathfrak{a} = N(\mathfrak{a}) = [\mathcal{O} : \mathfrak{a}]$. Using the fact that $\mathfrak{a}$ is invertible, one can show that $\mathrm{End}(E) \simeq \mathrm{End}(E')$; thus $\varphi_\mathfrak{a}$ is a horizontal isogeny.

If $\mathfrak{a}$ and $\mathfrak{b}$ are two invertible $\mathcal{O}$-ideals then $\varphi_{\mathfrak{a}\mathfrak{b}} = \varphi_\mathfrak{a}\varphi_\mathfrak{b}$. Thus the group of invertible $\mathcal{O}$-ideals acts on the set of elliptic curves with endomorphism ring $\mathcal{O}$. When $\mathfrak{a}$ is a principal ideal we have $E \simeq E'$; hence there is an induced action of the ideal class group $\mathrm{Cl}(\mathcal{O})$ on the set

$$\mathrm{Ell}_\mathcal{O}(k) = \{j(E) : E/k \text{ with } \mathrm{End}(E) \simeq \mathcal{O}\}.$$

This action is faithful (only principal ideals act trivially) and transitive (see [42, Proposition II.1.2] for a proof in the case that $k = \mathbb{C}$ and $\mathcal{O} = \mathcal{O}_K$, which may be generalized via [31, Chapters 10, 13]). Provided it is nonempty, the set $\mathrm{Ell}_\mathcal{O}(k)$ is thus a principal homogeneous space, a *torsor*, for the group $\mathrm{Cl}(\mathcal{O})$. The cardinality

of $\mathrm{Ell}_{\mathcal{O}}(k)$ is either 0 or h, where $h = h(\mathcal{O}) = \#\mathrm{Cl}(\mathcal{O})$ is the *class number*. Thus either every curve $E/\bar{k}$ with CM by $\mathcal{O}$ can be defined over k, or none of them can.

Remark 5. *Decomposing isogenies.* The CM action allows us to express horizontal isogenies $\varphi_{\mathfrak{a}}$ of large degree as the composition of a sequence of isogenies of smaller degree. Even if $\mathfrak{a}$ has prime norm, we may find that $[\mathfrak{a}] = [\mathfrak{p}_1 \cdots \mathfrak{p}_s]$ in $\mathrm{Cl}(\mathcal{O})$, where the $\mathfrak{p}_i$ are prime ideals with norms smaller than $\mathfrak{a}$. Under the generalized Riemann hypothesis (GRH), we can find, in probabilistic subexponential time, an equivalence $[\mathfrak{a}] = [\mathfrak{p}_1 \cdots \mathfrak{p}_s]$ in which the $\mathfrak{p}_i$ have norms that are polylogarithmic in the class number h and $s = O(\log h)$; see [11, Theorem 2.1]. This makes horizontal isogenies asymptotically easier to compute than vertical isogenies (this holds even without the GRH), which has implications for cryptography; see [6; 18; 19; 20; 27; 28].

2.9. ***Horizontal isogenies.*** Every horizontal ℓ-isogeny φ arises from the action of an invertible $\mathcal{O}$-ideal $\mathfrak{l}$ of norm ℓ, namely, the ideal of endomorphisms $\alpha \in \mathcal{O}$ whose kernels contain the kernel of φ. If ℓ divides the index of $\mathcal{O}$ in the maximal order $\mathcal{O}_K$ of its fraction field K, then no such ideals exist. Otherwise we say that $\mathcal{O}$ is *maximal at* ℓ, and in this case the number of invertible $\mathcal{O}$-ideals of norm ℓ is equal to

$$1 + \left(\frac{\mathrm{disc}(K)}{\ell}\right) = \begin{cases} 0 & \text{if } \ell \text{ is inert in } K, \\ 1 & \text{if } \ell \text{ is ramified in } K, \\ 2 & \text{if } \ell \text{ splits in } K. \end{cases}$$

Each such $\mathcal{O}$-ideal gives rise to a horizontal ℓ-isogeny. In the split case we have $(\ell) = \mathfrak{l} \cdot \bar{\mathfrak{l}}$, and the $\mathfrak{l}$-orbits partition $\mathrm{Ell}_{\mathcal{O}}(k)$ into cycles corresponding to the cosets of $\langle [\mathfrak{l}] \rangle$ in $\mathrm{Cl}(\mathcal{O})$. When $\mathfrak{l}$ is principal the ideal class $[\mathfrak{l}]$ is trivial, which leads to self-loops in $G_\ell(k)$. We can also have $[\mathfrak{l}] = [\bar{\mathfrak{l}}]$ even though $\mathfrak{l} \neq \bar{\mathfrak{l}}$, which gives rise to double edges in $G_\ell(k)$.

2.10. ***Vertical isogenies.*** Let $\mathcal{O}$ be an imaginary quadratic order with discriminant D, and let $\mathcal{O}' = \mathbb{Z} + \ell\mathcal{O}$ be the order of index ℓ in $\mathcal{O}$. To simplify matters, let us assume that $\mathcal{O}$ and $\mathcal{O}'$ have the same group of units $\{\pm 1\}$; this holds whenever $D < -4$, and excludes only the cases $\mathcal{O} = \mathbb{Z}[i]$ and $\mathcal{O} = \mathbb{Z}[\zeta_3]$, which correspond to the special j-invariants 1728 and 0, respectively.

The map that sends each invertible $\mathcal{O}'$-ideal $\mathfrak{a}$ to the invertible $\mathcal{O}$-ideal $\mathfrak{a}\mathcal{O}$ preserves norms and induces a surjective homomorphism

$$\rho : \mathrm{Cl}(\mathcal{O}') \to \mathrm{Cl}(\mathcal{O}).$$

See [12, Proposition 7.20] for a proof in the case that $\mathcal{O}$ is the maximal order; the general case is proved similarly (see [4, Lemma 3] and [7, §3]). Under a suitable identification of the class groups $\mathrm{Cl}(\mathcal{O}')$ and $\mathrm{Cl}(\mathcal{O})$ with their torsors $\mathrm{Ell}_{\mathcal{O}'}(k)$ and

$\mathrm{Ell}_{\mathcal{O}}(k)$, the vertical isogenies from $\mathrm{Ell}_{\mathcal{O}'}(k)$ to $\mathrm{Ell}_{\mathcal{O}}(k)$ correspond to the map from $\mathrm{Cl}(\mathcal{O}')$ to $\mathrm{Cl}(\mathcal{O})$ given by ρ. To show this, let us prove the following lemma.

Lemma 6. *Let E'/k be an elliptic curve with CM by $\mathcal{O}'$. Then there is a unique ascending ℓ-isogeny from E' to an elliptic curve E/k with CM by $\mathcal{O}$.*

Proof. The existence of E'/k implies that $\mathrm{Ell}_{\mathcal{O}'}(k)$ is nonempty, and since $\mathcal{O}$ contains $\mathcal{O}'$, it follows that $\mathrm{Ell}_{\mathcal{O}}(k)$ is also nonempty.[1]

Let us suppose that there exists an ascending ℓ-isogeny $\phi_1 : E_1' \to E_1$, for some elliptic curve E_1' with CM by $\mathcal{O}'$. Twisting E_1 if necessary, we may choose an invertible $\mathcal{O}'$-ideal $\mathfrak{a}'$ so that the horizontal isogeny $\varphi_{\mathfrak{a}'}$ maps E_1' to E'. If we now set $\mathfrak{a} = \rho(\mathfrak{a}')$ and let E be the image of $\varphi_{\mathfrak{a}} \circ \phi_1$, then E has CM by $\mathcal{O}$, and there is a unique isogeny $\phi : E' \to E$ such that $\phi \circ \varphi_{\mathfrak{a}'} = \varphi_{\mathfrak{a}} \circ \phi_1$, by [41, Corollary 4.11]. We have $\deg \phi = \deg \varphi_{\mathfrak{a}} \deg \phi_1 / \deg \varphi_{\mathfrak{a}'} = \ell$, thus ϕ is an ascending ℓ-isogeny. It follows that if any elliptic curve E_1'/k with CM by $\mathcal{O}'$ admits an ascending ℓ-isogeny, then so does every such elliptic curve.

We now proceed by induction on $d = v_\ell([\mathcal{O}_K : \mathcal{O}])$. Let $D_K = \mathrm{disc}(K)$. For $d = 0$, every elliptic curve E/k with CM by $\mathcal{O}$ admits $\ell + 1$ k-rational ℓ-isogenies, of which $1 + \left(\frac{D_K}{\ell}\right)$ are horizontal. The remaining $\ell - \left(\frac{D_K}{\ell}\right) > 0$ must be descending, and their duals are ascending ℓ-isogenies from elliptic curves with CM by $\mathcal{O}'$. It follows that there are a total of $\left(\ell - \left(\frac{D_K}{\ell}\right)\right) h(\mathcal{O})$ ascending ℓ-isogenies from $\mathrm{Ell}_{\mathcal{O}'}(k)$ to $\mathrm{Ell}_{\mathcal{O}}(k)$. By [12, Theorem 7.24], this is equal to the cardinality $h(\mathcal{O}')$ of $\mathrm{Ell}_{\mathcal{O}'}(k)$. Since there is at least one ascending ℓ-isogeny from each elliptic curve E'/k with CM by $\mathcal{O}'$, there must be exactly one in each case.

The argument for $d > 0$ is similar. By the inductive hypothesis, every elliptic curve E/k with CM by $\mathcal{O}$ admits exactly one ascending ℓ-isogeny, and since ℓ now divides $[\mathcal{O}_K : \mathcal{O}]$, there are no horizontal isogenies from E, and all ℓ of the remaining ℓ-isogenies from E must by descending. There are thus a total of $\ell h(\mathcal{O})$ ascending ℓ-isogenies from $\mathrm{Ell}_{\mathcal{O}'}(k)$, which equals the cardinality $h(\mathcal{O}')$ of $\mathrm{Ell}_{\mathcal{O}'}(k)$. □

It follows from the proof of Lemma 5 that there is a one-to-one correspondence between the graph of the function ρ and the edges of $G_\ell(k)$ that lead from $\mathrm{Ell}_{\mathcal{O}'}(k)$ to $\mathrm{Ell}_{\mathcal{O}}(k)$. Indeed, let us pick a vertex $j_1' \in \mathrm{Ell}_{\mathcal{O}'}(k)$ and let j_1 be its unique neighbor in $\mathrm{Ell}_{\mathcal{O}}(k)$ given by Lemma 6. If we identify the edge (j_1', j_1) in $G_\ell(k)$ with the edge $(1_{\mathrm{Cl}(\mathcal{O}')}, 1_{\mathrm{Cl}(\mathcal{O})})$ in the graph of ρ, then every other edge in the correspondence is determined in a way that is compatible with the actions of $\mathrm{Cl}(O')$ and $\mathrm{Cl}(O)$ on the torsors $\mathrm{Ell}_{\mathcal{O}'}(k)$ and $\mathrm{Ell}_{\mathcal{O}}(k)$. Under this correspondence, the vertices in $\mathrm{Ell}_{\mathcal{O}'}(k)$ that are connected to a given vertex v in $\mathrm{Ell}_{\mathcal{O}}(k)$ (the *children* of v) correspond to

[1] One way to see this is to note that k contains all the roots of the Hilbert class polynomial for $\mathcal{O}'$, hence it must contain all the roots of the Hilbert class polynomial for $\mathcal{O}$, since the ring class field of $\mathcal{O}'$ contains the ring class field of $\mathcal{O}$; see Section 3.4.

a coset of the kernel of ρ, a cyclic group of order $\ell - \left(\frac{D_K}{\ell}\right)$ generated by the class of an invertible $\mathcal{O}'$-ideal of norm ℓ^2; see [7, Lemma 3.2].

2.11. ***Ordinary elliptic curves over finite fields.*** We now assume that k is a finite field $\mathbb{F}_q$. Let $E/\mathbb{F}_q$ be an ordinary elliptic curve and let π_E denote the Frobenius endomorphism $(X, Y) \mapsto (X^q, Y^q)$. The *trace of Frobenius* is given by

$$t = \operatorname{Tr} \pi_E = q + 1 - \#E(\mathbb{F}_q),$$

and π_E satisfies the characteristic equation $\pi_E^2 - t\pi_E + q = 0$. As an element of the imaginary quadratic order $\mathcal{O} \simeq \operatorname{End}(E)$, the Frobenius endomorphism corresponds to an algebraic integer with trace t and norm q. Thus we have the *norm equation*

$$4q = t^2 - v^2 D_K,$$

in which D_K is the discriminant of the field $K = \mathbb{Q}(\sqrt{t^2 - 4q})$ containing $\mathcal{O}$, and $v = [\mathcal{O}_K : \mathbb{Z}[\pi_E]]$. We have

$$\mathbb{Z}[\pi_E] \subseteq \mathcal{O} \subseteq \mathcal{O}_K,$$

thus $[\mathcal{O}_K : \mathcal{O}]$ divides v, and the same is true for any elliptic curve $E/\mathbb{F}_q$ with Frobenius trace t.

Let us now define

$$\operatorname{Ell}_t(\mathbb{F}_q) = \{j(E) : E/\mathbb{F}_q \text{ satisfies } \operatorname{Tr} \pi_E = t\},$$

the set of $\overline{\mathbb{F}}_p$-isomorphism classes of elliptic curves over $\mathbb{F}_p$ with a given Frobenius trace t. By a theorem of Tate [47], $\operatorname{Ell}_t(\mathbb{F}_q)$ corresponds to an isogeny class, but note that $\operatorname{Ell}_t(\mathbb{F}_q) = \operatorname{Ell}_{-t}(\mathbb{F}_q)$. For any ordinary elliptic curve $E/\mathbb{F}_q$ with Frobenius trace $t = \operatorname{Tr} \pi_E$, we may write $\operatorname{Ell}_t(\mathbb{F}_q)$ as the disjoint union

$$\operatorname{Ell}_t(\mathbb{F}_q) = \bigsqcup_{\mathbb{Z}[\pi_E] \subseteq \mathcal{O} \subseteq \mathcal{O}_K} \operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_q),$$

of cardinality equal to the Kronecker class number $H(t^2 - 4q)$; see [40, Definition 2.1].

2.12. ***The main theorem.*** We now arrive at our main theorem, which states that the ordinary components of $G_\ell(\mathbb{F}_q)$ (other than the components of 0 and 1728) are ℓ-volcanoes, and characterizes the structure of these components. The proof follows easily from the material we have presented, as the reader may wish to verify.

Theorem 7 (Kohel). *Let V be an ordinary component of $G_\ell(\mathbb{F}_q)$ that does not contain* 0 *or* 1728. *Then V is an ℓ-volcano for which the following hold*:

(1) *The vertices in level V_i all have the same endomorphism ring $\mathcal{O}_i$.*

(2) *The subgraph on* V_0 *has degree* $1 + \left(\frac{D_0}{\ell}\right)$, *where* $D_0 = \operatorname{disc}(\mathbb{O}_0)$.

(3) *If* $\left(\frac{D_0}{\ell}\right) \geq 0$, *then* $|V_0|$ *is the order of* $[\mathfrak{l}]$ *in* $\operatorname{Cl}(\mathbb{O}_0)$; *otherwise* $|V_0| = 1$.

(4) *The depth of* V *is* $d = v_\ell\big((t^2 - 4q)/D_0\big)/2$, *where* $t^2 = (\operatorname{Tr}\pi_E)^2$ *for any* E *with* $j(E) \in V$.

(5) *We have* $\ell \nmid [\mathbb{O}_K : \mathbb{O}_0]$ *and* $[\mathbb{O}_i : \mathbb{O}_{i+1}] = \ell$ *for* $0 \leq i < d$.

Remark 8. *Special cases.* Theorem 7 is easily extended to the case where V contains 0 or 1728. Parts (1)–(5) still hold; the only necessary modification is the claim that V is an ℓ-volcano. When V contains 0, if V_1 is nonempty then it contains $\frac{1}{3}\big(\ell - \left(\frac{-3}{\ell}\right)\big)$ vertices, and each vertex in V_1 has three incoming edges from 0 but only one outgoing edge to 0. When V contains 1728, if V_1 is nonempty then it contains $\frac{1}{2}\big(\ell - \left(\frac{-1}{\ell}\right)\big)$ vertices, and each vertex in V_1 has two incoming edges from 1728 but only one outgoing edge to 1728. This 3-to-1 (respectively, 2-to-1) discrepancy arises from the action of $\operatorname{Aut}(E)$ on the cyclic subgroups of $E[\ell]$ when $j(E) = 0$ (respectively, $j(E) = 1728$). Otherwise, V satisfies all the requirements of an ℓ-volcano, and most of the algorithms we present in the next section are equally applicable to V.

Example 9. Let $p = 411751$ and $\ell = 3$. The graph $G_3(\mathbb{F}_p)$ has a total of 206254 components, of which 205911 are ordinary and 343 are supersingular. The supersingular components all lie in the same isogeny class (which is connected in $G_3(\mathbb{F}_{p^2})$), while the ordinary components lie in 1283 distinct isogeny classes.

Let us consider the isogeny class $\operatorname{Ell}_t(\mathbb{F}_p)$ for $t = 52$. We then have $4p = t^2 - v^2 D$ with $v = 2 \cdot 3^2 \cdot 5$ and $D = -203$. The subgraph $G_{\ell,t}(\mathbb{F}_p)$ of $G_\ell(\mathbb{F}_p)$ on $\operatorname{Ell}_t(\mathbb{F}_p)$ (known as a *cordillera* [33]) consists of ten 3-volcanoes, all of which have depth $d = v_\ell(v) = 2$. It contains a total of 1008 vertices distributed as follows:

- 648 vertices lie in six 3-volcanoes with $[\mathbb{O}_K : \mathbb{O}_0] = 10$ and $|V_0| = 12$.
- 216 vertices lie in two 3-volcanoes with $[\mathbb{O}_K : \mathbb{O}_0] = 5$ and $|V_0| = 12$.
- 108 vertices lie in a 3-volcano with $[\mathbb{O}_K : \mathbb{O}_0] = 2$ and $|V_0| = 12$.
- 36 vertices lie in a 3-volcano with $[\mathbb{O}_K : \mathbb{O}_0] = 1$ and $|V_0| = 4$.

For comparison:

- $G_{2,52}(\mathbb{F}_p)$ consists of 252 2-volcanoes of depth 1 with $|V_0| = 1$.
- $G_{5,52}(\mathbb{F}_p)$ consists of 144 5-volcanoes of depth 1 with $|V_0| = 1$.
- $G_{7,52}(\mathbb{F}_p)$ consists of 504 7-volcanoes with two vertices and one edge.
- $G_{11,52}(\mathbb{F}_p)$ consists of 1008 11-volcanoes that are all isolated vertices.

3. Applications

We now consider several applications of isogeny volcanoes, starting with one that is very simple, but nevertheless instructive.

3.1. ***Finding the floor.*** Let $E/\mathbb{F}_q$ be an ordinary elliptic curve. Then $j(E)$ lies in an ordinary component V of $G_\ell(\mathbb{F}_q)$. We wish to find a vertex on the *floor* of V, that is, a vertex v in level V_d, where d is the depth of V. Such vertices v are easily distinguished by their (out-)degree, which is the number of roots of $\Phi_\ell(v, Y)$ that lie in $\mathbb{F}_q$ (counted with multiplicity).

Proposition 10. *Let v be a vertex in an ordinary component V of depth d in $G_\ell(\mathbb{F}_q)$. Either* $\deg v \le 2$ *and* $v \in V_d$, *or* $\deg v = \ell + 1$ *and* $v \notin V_d$.

Proof. If $d = 0$ then $V = V_0 = V_d$ is a regular graph of degree at most 2 and $v \in V_d$. Otherwise, either $v \in V_d$ and v has degree 1, or $v \notin V_d$ and v has degree $\ell + 1$. □

We note that if $j(E)$ is on the floor then $E[\ell](\mathbb{F}_q)$ is necessarily cyclic (otherwise there would be another level below the floor). This is useful, for example, when using the CM method to construct Edwards curves [34], and shows that every ordinary elliptic curve $E/\mathbb{F}_q$ is isogenous to some $E'/\mathbb{F}_q$ with $E'(\mathbb{F}_q)$ cyclic.

Our strategy for finding the floor is simple: If $v_0 = j(E)$ is not already on the floor then we will construct a random path from v_0 to a vertex v_s on the floor. By a *path*, we mean a sequence of vertices $v_0, v_1, \ldots, v_s$ such that each pair (v_{i-1}, v_i) is an edge and $v_i \neq v_{i-2}$ (so backtracking is prohibited).

Algorithm (FINDFLOOR).

Input: An ordinary vertex $v_0 \in G_\ell(\mathbb{F}_q)$.

Output: A vertex on the floor of the component of v_0.

1. If $\deg v_0 \le 2$ then output v_0 and terminate.
2. Pick a random neighbor v_1 of v_0 and set $s \leftarrow 1$.
3. While $\deg v_s > 1$: Pick a random neighbor $v_{s+1} \neq v_{s-1}$ of v_s and increment s.
4. Output v_s.

The complexity of FINDFLOOR is given by the following proposition, in which $\mathsf{M}(n)$ denotes the time to multiply two n-bit integers. It is worth noting that for large ℓ the complexity is dominated by the time to substitute v into $\Phi_\ell(X, Y)$, not by root-finding (a fact that is occasionally overlooked).

Proposition 11. *Given $\Phi_\ell \in \mathbb{F}_q[X, Y]$, each step of* FINDFLOOR *can be accomplished in $O(\ell^2 \mathsf{M}(n) + \mathsf{M}(\ell n) n)$ expected time, where $n = \log q$. The expected number of steps s is $\delta + O(1)$, where δ is the distance from v_0 to the floor.*

Proof. Computing $\phi(Y) = \Phi_\ell(v, Y)$ involves $O(\ell^2)$ $\mathbb{F}_q$-operations, or $O(\ell^2 \mathsf{M}(n))$ bit operations. The neighbors of v are the distinct roots of $\phi(Y)$ that lie in $\mathbb{F}_q$, which are precisely the roots of $f(Y) = \gcd(Y^q - Y, \phi(Y))$. Computing $Y^q \bmod \phi$ involves $O(n)$ multiplications in the ring $\mathbb{F}_q[Y]/(\phi)$, each of which can be accomplished using $O(\mathsf{M}(\ell n))$ bit operations, via Kronecker substitution [22], yielding an $O(\mathsf{M}(\ell n)n)$ bound. With the fast Euclidean algorithm the gcd of two polynomials of degree $O(\ell)$ can be computed using $O(\mathsf{M}(\ell n) \log \ell)$ bit operations. If $\log \ell < n$ then this is bounded by $O(\mathsf{M}(\ell n)n)$, and otherwise it is bounded by $O(\ell^2 \mathsf{M}(n))$. Thus the total time to compute $f(Y)$ for any particular v is $O(\ell^2 \mathsf{M}(n) + \mathsf{M}(\ell n)n)$.

The degree of $f(Y)$ is the number of distinct roots of $\Phi_\ell(Y, v)$ in $\mathbb{F}_q$. For $\ell > 3$, this is less than or equal to 2 if and only if v is on the floor. For $\ell \leq 3$ we can count roots with multiplicity by taking gcds with derivatives of ϕ, within the same time bound. To find a random root of $f(Y)$ we use the probabilistic splitting algorithm of [37]; since we need only one root, this takes $O(\mathsf{M}(\ell n)n)$ expected time.

For every vertex v in a level V_i above the floor, at least 1/3 of v's neighbors lie in level V_{i+1}, thus within $O(1)$ expected steps the path will be extended along a descending edge. Once this occurs, every subsequent edge in the path must be descending, since we are not allowed to backtrack along the single ascending edge, and we will reach the floor within $\delta + O(1)$ steps. □

Remark 12. *Removing known roots.* As a minor optimization, rather than picking v_{s+1} as a root of $\phi(Y) = \Phi_\ell(v_s, Y)$ in step 3 of the FINDFLOOR algorithm, we may use $\phi(Y)/(Y - v_{s-1})^e$, where e is the multiplicity of v_{s-1} as a root of $\phi(Y)$. This is slightly faster and eliminates the need to check that $v_{s+1} \neq v_{s-1}$.

The FINDFLOOR algorithm finds a path of expected length $\delta + O(1)$ from v_0 to the floor. With a bit more effort we can find a path of exactly length δ, using a simplified version of an algorithm from [17].

Algorithm (FINDSHORTESTPATHTOFLOOR).

Input: An ordinary $v_0 \in G_\ell(\mathbb{F}_q)$.

Output: A shortest path to the floor of the component of v_0.

1. Let $v_0 = j(E)$. If $\deg v_0 \leq 2$ then output v_0 and terminate.
2. Pick three neighbors of v_0 and extend paths from each of these neighbors in parallel, stopping as soon as any of them reaches the floor. (If v_0 does not have three distinct neighbors then just pick all of them.)
3. Output a path that reached the floor.

The correctness of the algorithm follows from the fact that at most two of v_0's neighbors do not lie along descending edges, so one of the three paths must begin with a descending edge. This path must then consist entirely of descending edges,

yielding a shortest path to the floor. The algorithm takes at most 3δ steps, each of which has complexity bounded as in Proposition 11.

The main virtue of FINDSHORTESTPATHTOFLOOR is that it allows us to compute δ, which tells us the level $V_{d-\delta}$ of $j(E)$ relative to the floor V_d. It effectively gives us an "altimeter" $\delta(v)$ that may be used to navigate V. We can determine whether a given edge (v_1, v_2) is horizontal, ascending, or descending, by comparing $\delta(v_1)$ to $\delta(v_2)$, and we can determine the exact level of any vertex; see [43, §4.1] for algorithms and further details. We should also mention that an alternative approach based on pairings has recently been developed by Ionica and Joux [25; 26], which is more efficient when d is large.

3.2. ***Identifying supersingular curves.*** Both algorithms in the previous section assume that their input is the j-invariant of an ordinary elliptic curve. But what if this is not the case? If we attempt to "find the floor" on the supersingular component of $G_\ell(\mathbb{F}_{p^2})$ we will never succeed, since every vertex has out-degree $\ell + 1$. On the other hand, from part (4) of Theorem 7 (and Remark 8), we know that every ordinary component of $G_\ell(\mathbb{F}_{p^2})$ has depth less than $\log_\ell 2p$, so we can bound the length of the shortest path to the floor from any ordinary vertex.

This suggests that, with minor modifications, the algorithm FINDSHORTESTPATHTOFLOOR can be used to determine whether a given elliptic curve $E/\mathbb{F}_q$ is ordinary or supersingular. If $j(E) \notin \mathbb{F}_{p^2}$ then E must be ordinary, so we may assume $v_0 = j(E) \in \mathbb{F}_{p^2}$ (even if E is defined over $\mathbb{F}_p$, we want to work in $\mathbb{F}_{p^2}$). We modify step 2 of the algorithm so that if none of the three paths reaches the floor within $\log_\ell 2p$ steps, it reports that its input is supersingular and terminates. Otherwise, the algorithm succeeds and can report that its input is ordinary. This works for any prime ℓ, but using $\ell = 2$ gives the best running time.

This yields a Las Vegas algorithm to determine whether a given elliptic curve is ordinary or supersingular in $\widetilde{O}(n^3)$ expected time, where $n = \log q$. For comparison, the best previously known Las Vegas algorithm has an expected running time of $\widetilde{O}(n^4)$, and the best known deterministic algorithm runs in $\widetilde{O}(n^5)$ time. Remarkably, the average time for a random input is only $\widetilde{O}(n^2)$. This matches the complexity of the best known Monte Carlo algorithm for this problem, with better constant factors; see [45] for further details.

3.3. ***Computing endomorphism rings.*** We now turn to a more difficult problem: determining the endomorphism ring of an ordinary elliptic curve $E/\mathbb{F}_q$. We assume that the trace of Frobenius $t = \operatorname{Tr} \pi_E$ is known; this can be computed in polynomial time using Schoof's algorithm [39]. By factoring $4q - t^2$, we can compute the positive integer v and fundamental discriminant D satisfying the norm equation $4q = t^2 - v^2 D$. We then know that $\mathbb{Z}[\pi_E]$ has index v in the maximal order $\mathbb{O}_K$,

where $K = \mathbb{Q}(\sqrt{D})$. The order $\mathcal{O} \simeq \mathrm{End}(E)$ is uniquely determined by its index u in $\mathcal{O}_K$, and u must be a divisor of v. Let us assume $D < -4$.

We can determine u by determining the level of $j(E)$ in its component of $G_\ell(\mathbb{F}_q)$ for each of the primes ℓ dividing v. If $v = \ell_1^{e_1} \cdots \ell_w^{e_w}$ is the prime factorization of v, then $u = \ell_1^{d_1} \cdots \ell_w^{d_w}$, where $\delta_i = e_i - d_i$ is the distance from $j(E)$ to the floor of its ℓ_i-volcano. But it may not be practical to compute δ_i using FINDSHORTESTPATHTOFLOOR when ℓ_i is large: Its complexity is quasiquadratic in ℓ_i, which may be exponential in $\log q$ (and computing Φ_{ℓ_i} is even harder). More generally, we do not know any algorithm for computing a vertical ℓ-isogeny whose complexity is not at least linear in ℓ (in general, quadratic in ℓ). This would seem to imply that we cannot avoid a running time that is exponential in $\log q$.

However, as noted in Remark 5, computing horizontal isogenies is easier than computing vertical isogenies. We now sketch an approach to computing $\mathrm{End}(E)$ that uses horizontal isogenies to handle large primes dividing v, based on the algorithm in [4]. To simplify the presentation, we assume that v is squarefree; the generalization to arbitrary v is straightforward.

Let $\mathcal{L}$ be the lattice of orders in $\mathcal{O}_K$ that contain $\mathbb{Z}[\pi_E]$. Our strategy is to determine whether u is divisible by a given prime divisor ℓ of v using a smooth relation that holds in an order $\mathcal{O} \in \mathcal{L}$ if and only if $\mathcal{O}$ is maximal at ℓ. This relation will hold in $\mathrm{End}(E)$ if and only if u is not divisible by ℓ.

A *smooth relation* R is a multiset $\{\mathfrak{p}_1^{r_1}, \ldots, \mathfrak{p}_s^{r_s}\}$ in which the $\mathfrak{p}_i$ are invertible $\mathbb{Z}[\pi_E]$-ideals with prime norms p_i occurring with multiplicity r_i, such that p_i and r_i satisfy bounds that are subexponential in $\log q$. We say that R *holds* in $\mathcal{O} \in \mathcal{L}$ if the $\mathcal{O}$-ideal $R_\mathcal{O} = (\mathfrak{p}_1\mathcal{O})^{r_1} \cdots (\mathfrak{p}_s\mathcal{O})^{r_s}$ is principal. If $\mathcal{O}' \subset \mathcal{O}$, the existence of the norm-preserving homomorphism $\rho : \mathrm{Cl}(\mathcal{O}') \to \mathrm{Cl}(\mathcal{O})$ defined as in Section 2.10 implies that if R holds in $\mathcal{O}'$, then it holds in $\mathcal{O}$. It thus suffices to find a relation that holds in the order of index v/ℓ in $\mathcal{O}_K$, but not in the order of index ℓ in $\mathcal{O}_K$. Under the GRH, for $\ell > 3$ we can find such an R in probabilistic subexponential time [3].

To determine whether R holds in $\mathcal{O} \simeq \mathrm{End}(E)$, we compute the CM action of $[R_\mathcal{O}] \in \mathrm{Cl}(\mathcal{O})$ on $j(E) \in \mathrm{Ell}_O(\mathbb{F}_q)$. This involves walking r_i steps along the surface of a p_i-volcano for each of the $\mathfrak{p}_i$ appearing in R and then checking whether we wind up back at our starting point $j(E)$. None of the p_i divide v, so these p_i-volcanoes all have depth 0 and consist of either a single edge or a cycle. We must choose a direction to walk along each cycle (one corresponds to the action of $\mathfrak{p}_i$, the other to $\overline{\mathfrak{p}}_i$). There are methods to determine the correct choice, but in practice we can make s small enough so that it is easy to simply try every combination of choices and count how many work; see [4] for details.

Under the GRH, this algorithm has a subexponential expected running time of $L[1/2, \sqrt{3}/2]$ plus the cost of factoring $4q - t^2$ (the latter is heuristically negligible, using the number field sieve, and provably bounded by $L[1/2, 1]$ in [32]). Bisson [3] has recently improved this to $L[1/2, \sqrt{2}/2]$ plus the cost of factoring $4q - t^2$.

Example 13. Let $q = 2^{320} + 261$ and suppose that $E/\mathbb{F}_q$ has Frobenius trace

$$t = 230641434457621363389123643439267139273704045955 8.$$

Then $4q = t^2 - v^2 D$, where $D = -147759$ and $v = 2^2 p_1 p_2$, with

$$p_1 = 16447689059735824784039,$$
$$p_2 = 71003976975490059472571.$$

We can easily determine the level of $j(E)$ in its 2-volcano by finding a shortest path to the floor. For p_1 and p_2 we instead use smooth relations R_1 and R_2.

Let $\mathcal{O}_1$ be the order of index p_1 in $\mathcal{O}_K$, and $\mathcal{O}_1'$ the order of index v/p_1 in $\mathcal{O}_K$. The relation

$$R_1 = \{\mathfrak{p}_5, \mathfrak{p}_{19}^2, \bar{\mathfrak{p}}_{23}^{210}, \mathfrak{p}_{29}, \mathfrak{p}_{31}, \bar{\mathfrak{p}}_{41}^{145}, \mathfrak{p}_{139}, \bar{\mathfrak{p}}_{149}, \mathfrak{p}_{167}, \bar{\mathfrak{p}}_{191}, \bar{\mathfrak{p}}_{251}^{6}, \mathfrak{p}_{269}, \bar{\mathfrak{p}}_{587}^{7}, \bar{\mathfrak{p}}_{643}\}$$

holds in $\mathcal{O}_1$ but not in $\mathcal{O}_1'$ (here $\mathfrak{p}_\ell$ denotes the ideal of norm ℓ corresponding to the reduced binary quadratic form $\ell x^2 + bxy + cy^2$ with $b \geq 0$). If we now let $\mathcal{O}_2$ be the order of index p_2 in $\mathcal{O}_K$ and $\mathcal{O}_2'$ the order of index v/p_2 in $\mathcal{O}_K$, then

$$R_2 = \{\mathfrak{p}_{11}, \bar{\mathfrak{p}}_{13}^{576}, \mathfrak{p}_{23}^2, \bar{\mathfrak{p}}_{41}, \bar{\mathfrak{p}}_{47}, \mathfrak{p}_{83}, \mathfrak{p}_{101}, \bar{\mathfrak{p}}_{197}^{28}, \bar{\mathfrak{p}}_{307}^{3}, \mathfrak{p}_{317}, \bar{\mathfrak{p}}_{419}, \mathfrak{p}_{911}\}$$

holds in $\mathcal{O}_2$ but not in $\mathcal{O}_2'$.

Including the time to compute the required modular polynomials and the time to find the relations R_1 and R_2, the total time to compute $\mathrm{End}(E)$ in this example is less than half an hour. In contrast, it would be completely infeasible to directly compute a vertical isogeny of degree p_1 or p_2; writing down even a single element of the kernel of such an isogeny would require more than 2^{80} bits.

3.4. ***Computing Hilbert class polynomials.*** Let $\mathcal{O}$ be an imaginary quadratic order with discriminant D. The *Hilbert class polynomial* H_D is defined by

$$H_D(X) = \prod_{j \in \mathrm{Ell}_{\mathcal{O}}(\mathbb{C})} (X - j).$$

Equivalently, $H_D(X)$ is the minimal polynomial of the j-invariant of the lattice $\mathcal{O}$ over the field $K = \mathbb{Q}(\sqrt{D})$. Remarkably, its coefficients lie in $\mathbb{Z}$.

The field $K_{\mathcal{O}} = K(j(\mathcal{O}))$ is the *ring class field* of $\mathcal{O}$. If a prime q splits completely in $K_{\mathcal{O}}$, then $H_D(X)$ splits completely in $\mathbb{F}_q[X]$ and its roots form the set $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_q)$. Each root is then the j-invariant of an elliptic curve $E/\mathbb{F}_q$ with $\mathrm{End}(E) \simeq \mathcal{O}$. We

must have $\#E(\mathbb{F}_q) = q + 1 - t$, where the norm equation $4q = t^2 - v^2 D$ uniquely determines the integers t and v up to sign, for $D < -4$. We can thus use a root of $H_D(X)$ in $\mathbb{F}_q$ to construct an elliptic curve $E/\mathbb{F}_q$ with exactly $q + 1 - t$ rational points; under some reasonable heuristic assumptions about the distribution of prime numbers, we can achieve any desired cardinality for $E(\mathbb{F}_q)$ by choosing q and D appropriately [8]. This is known as the *CM method*, which is commonly used in elliptic curve cryptography and elliptic curve primality proving.

We now outline an algorithm to compute $H_D(X)$ using the CRT approach described in [1; 43]. Under the GRH it runs in $O(|D|(\log|D|)^{5+o(1)})$ expected time, which is quasilinear in the $O(|D|\log|D|)$ size of $H_D(X)$. The same approach can be used to compute many other types of class polynomials; see [14].

Algorithm (COMPUTEHILBERTCLASSPOLYNOMIAL).

Input: An imaginary quadratic discriminant D.

Output: The Hilbert class polynomial $H_D(X)$.

1. Select a sufficiently large set of primes p that satisfy $4p = t^2 - v^2 D$.
2. For each prime p, compute $H_D(X) \bmod p$ as follows:
   (a) Generate random elliptic curves $E/\mathbb{F}_p$ until $\#E(\mathbb{F}_p) = p + 1 - t$.
   (b) Use volcano climbing to find E' isogenous to E with $\mathrm{End}(E') \simeq \mathbb{O}$.
   (c) Enumerate $\mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)$ by applying the $\mathrm{Cl}(\mathbb{O})$-action to $j(E')$.
   (d) Compute $H_D(X) = \prod_{j \in \mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)}(X - j) \bmod p$.
3. Use the CRT to recover $H_D(X)$ over $\mathbb{Z}$ (or over $\mathbb{F}_q$ via the explicit CRT).

Isogeny volcanoes play a key role in the efficient implementation of this algorithm, not only in step 2(b), but also in step 2(c), which is the most critical step and merits further discussion. Given any sequence of generators $\alpha_1, \ldots, \alpha_k$ for a finite abelian group G, if we let $G_i = \langle \alpha_1, \ldots, \alpha_i \rangle$ and define $r_i = [G_i : G_{i-1}]$, then every element β of G can be uniquely represented in the form $\beta = \alpha_1^{e_1} \cdots \alpha_k^{e_k}$, with $0 \le e_i < r_i$. This is a special case of a *polycyclic presentation*. We can use a polycyclic presentation of $\mathrm{Cl}(\mathbb{O})$ to enumerate the torsor $\mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)$ by enumerating the list of exponent vectors $(e_1, \ldots, e_k)$ in reverse lexicographic order. At each step we apply the action of the generator α_i that transforms the current exponent vector to the next in the list (usually $i = 1$, since e_1 varies most frequently).

Using generators of the form $\alpha_i = [\mathfrak{l}_i]$, where $\mathfrak{l}_i$ is an invertible $\mathbb{O}$-ideal of prime norm ℓ_i, this amounts to walking along the surfaces of various ℓ-volcanoes. To make this process as efficient as possible, it is crucial to minimize the size of the primes ℓ_i. This is achieved by choosing $\mathfrak{l}_1$ to minimize ℓ_1 and then minimizing each ℓ_i subject to $[\mathfrak{l}_i] \notin \langle [\mathfrak{l}_1], \ldots, [\mathfrak{l}_{i-1}] \rangle$; this is called an *optimal presentation* [43, §5.1]. This will often cause us to use a set of generators that is larger than strictly needed.

As an example, for $D = -79947$ the class group $\mathrm{Cl}(\mathcal{O})$ is cyclic of order 100, generated by the class of an ideal with norm 19. But the optimal presentation for $\mathrm{Cl}(\mathcal{O})$ uses ideals $\mathfrak{l}_1$ and $\mathfrak{l}_2$ with norms 2 and 13, respectively. The classes of these ideals are not independent, we have $[\mathfrak{l}_2]^5 = [\mathfrak{l}_1]^{18}$, but they do form a polycyclic presentation with $r_1 = 20$ and $r_2 = 5$. Using this presentation to enumerate $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ is more than 100 times faster than using any single generator of $\mathrm{Cl}(\mathcal{O})$. One can construct examples where the optimal presentation is exponentially faster than any presentation that minimizes the number of generators; see [43, §5.3].

Enumerating $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ using a polycyclic presentation involves walking along the surfaces of various ℓ-volcanoes, as in the previous section when testing relations. But using an optimal presentation will often mean that some of the primes ℓ_i divide v. This always happens, for example, when $D \equiv 1 \bmod 8$, since in this case $\ell_1 = 2$ divides v. Thus we must be prepared to walk along the surface of an ℓ-volcano of nonzero depth. We now give a simple algorithm to do this.

Algorithm (WALKSURFACEPATH).

Input: A vertex v_0 on the surface V_0 of an ℓ-volcano of depth d and a positive integer $n < \#V_0$.

Output: A path $v_0, \ldots, v_n$ in V_0.

1. If v_0 has a single neighbor v_1, then return the path v_0, v_1. Otherwise, walk a path $v_0, \ldots, v_d$ and set $i \leftarrow 0$.
2. While $\deg v_{i+d} = 1$: Replace $v_{i+1}, \ldots, v_{i+d}$ by extending the path $v_0, \ldots, v_i$ by d steps, starting from an unvisited neighbor v'_{i+1} of v_i.
3. Extend the path $v_0, \ldots, v_{i+d}$ to $v_0, \ldots, v_{i+d+1}$ and increment i.
4. If $i = n$ then return $v_0, \ldots, v_n$; otherwise, go to step 2.

Algorithm WALKSURFACEPATH requires us to know the depth d of the ℓ-volcano, which we may determine from the norm equation. It works by walking an arbitrary path to the floor and then backing up d steps to a vertex that must be on the surface (whenever we leave the surface we must descend to the floor in exactly d steps). When d or ℓ is large, this algorithm is not very inefficient and the pairing-based approach of [25] may be faster. But in the context of computing Hilbert class polynomials, both d and ℓ are typically quite small.

Remark 14. *Walking the surface with gcds.* An alternative approach to walking the surface using gcds is given in [14]. Suppose we have already enumerated $v_0, \ldots, v_n$ along the surface of an ℓ-volcano, and have also taken a single step from v_0 to an adjacent vertex v'_0 on the surface of an ℓ'-volcano. We can then compute a path $v'_0, \ldots, v'_n$ along the surface of the ℓ-volcano containing v'_0 by computing each v'_{i+1} as the unique root of $f(Y) = \gcd(\Phi_\ell(v'_i, Y), \Phi_{\ell'}(v_{i+1}, Y))$.

The vertex v'_{i+1} is guaranteed to be on the surface, and the root-finding operation is trivial, since $f(Y)$ has degree 1. This approach is generally much faster than using either WALKSURFACEPATH or the algorithm in [25], and in practice most of the vertices in $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ can be enumerated this way; see [14] for further details.

Remark 15. *Space complexity.* A key virtue of the CRT approach is that by using the *explicit CRT* [2, Theorem 3.2], it is possible to directly compute the coefficients of $H_D(X)$ modulo an integer m (the characteristic of $\mathbb{F}_q$, for example), without first computing the coefficients over $\mathbb{Z}$. This means we can compute $H_D(X)$ over $\mathbb{F}_q$ with a space complexity that is quasilinear in $h(D)\log q$, which may be much smaller than $|D|\log|D|$. When $h(D)$ is sufficiently composite (often the case), we can use a decomposition of the ring class field to find a root of $H_D(X)$ in $\mathbb{F}_q$ with a space complexity quasilinear in $h(D)^{1/2}\log q$; see [44]. The low space complexity of the CRT approach has greatly increased the range of feasible discriminants for the CM method: Examples with $|D| \approx 10^{16}$ can now be handled [44], whereas $|D| \approx 10^{10}$ was previously regarded as a practical upper limit [13].

3.5. ***Computing modular polynomials.*** All of the algorithms we have discussed depend on modular polynomials $\Phi_\ell(X, Y)$; we even used them to define the graph of ℓ-isogenies. We now outline an algorithm to compute Φ_ℓ, using the CRT approach described in [7]. Under the GRH, it runs in $O(\ell^3(\log \ell)^{3+o(1)})$ expected time, which makes it the fastest method known for computing $\Phi_\ell(X, Y)$.

Algorithm (COMPUTEMODULARPOLYNOMIAL).

Input: An odd prime ℓ.

Output: The modular polynomial $\Phi_\ell(X, Y)$.

1. Pick an order $\mathcal{O}$ with $h(\mathcal{O}) > \ell + 1$ and let $D = \mathrm{disc}(\mathcal{O})$.
2. Select a sufficiently large set of primes p that satisfy $4p = t^2 - \ell^2 v^2 D$, with $\ell \nmid v$ and $p \equiv 1 \bmod \ell$.
3. For each prime p, compute $\Phi_\ell(X, Y) \bmod p$ as follows:
   (a) Enumerate $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ starting from a root v_0 of $H_D(X) \bmod p$.
   (b) Use Vélu's algorithm to compute a descending ℓ-isogeny from v_0 to v'_0.
   (c) Enumerate $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ using v'_0 as a starting point, where $[\mathcal{O} : \mathcal{O}'] = \ell$.
   (d) Map the ℓ-volcanoes that make up $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p) \cup \mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$.
   (e) Interpolate $\Phi_\ell(X, Y) \bmod p$.
4. Use the CRT to recover $\Phi_\ell(X, Y)$ over $\mathbb{Z}$ (or over $\mathbb{F}_q$ via the explicit CRT).

The restrictions on p ensure that each element of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ lies on the surface of an ℓ-volcano of depth 1 whose floor consists of elements of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$. An example with $\ell = 5$ and $D = -151$ is shown in Figure 3.

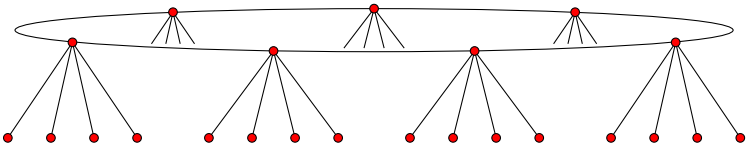


Figure 3. A volcano with $\ell = 5$ and $D = -151$.

When we enumerate $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ in step 3(a), we use a polycyclic presentation $\boldsymbol{\alpha}$ for $\mathrm{Cl}(\mathcal{O})$ derived from prime ideals whose norms are all less than ℓ (for $\ell > 2$ this is always possible). By expressing the class γ of an invertible $\mathcal{O}$-ideal of norm ℓ in terms of $\boldsymbol{\alpha}$, we can then determine all of the horizontal ℓ-isogenies between elements of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ without knowing Φ_ℓ. In our example with $D = -151$, the presentation $\boldsymbol{\alpha}$ consists of a single generator α corresponding to an ideal of norm 2, with $\gamma = \alpha^3$. Thus our enumeration of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ yields a cycle of 2-isogenies that we can convert to a cycle of 5-isogenies by simply picking out every third element.

The application of Vélu's algorithm in step 3(b) involves picking a random point P of order ℓ and computing the ℓ-isogeny φ with $\langle P \rangle$ as its kernel. This process is greatly facilitated by our choice of p, which ensures that P has coordinates in $\mathbb{F}_p$, rather than an extension field. We may find that φ is a horizontal ℓ-isogeny, but we can easily detect this and try again with a different P.

As in step 3(a), when we enumerate $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ in step 3(c) we use a polycyclic presentation $\boldsymbol{\beta}$ for $\mathrm{Cl}(\mathcal{O}')$ derived from prime ideals whose norms are all less than ℓ. There are no horizontal ℓ-isogenies between elements of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$, but we need to connect each element of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ to its ℓ-isogenous parent in $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$. This is done by identifying one child v' of each parent and then identifying that child's siblings, which are precisely the elements of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ related to v' by a cyclic isogeny of degree ℓ^2. By expressing the class γ' of an invertible $\mathcal{O}'$ ideal of norm ℓ^2 in terms of $\boldsymbol{\beta}$, we can identify the ℓ^2-isogeny cycles of siblings in $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$; these are precisely the cosets of the homomorphism $\rho : \mathrm{Cl}(\mathcal{O}') \to \mathrm{Cl}(\mathcal{O})$ in Section 2.10.

After identifying the horizontal isogenies among the vertices v in $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ and the children of each v, we can completely determine the subgraph of $G_\ell(\mathbb{F}_p)$ on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p) \cup \mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$; this is what it means to "map" the ℓ-volcanoes in step 3(d). In our example with $D = -151$ there is just one ℓ-volcano; Figure 4 depicts the result of mapping this ℓ-volcano when $p = 4451$.

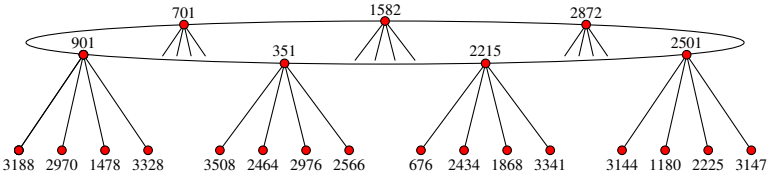


Figure 4. The fully labeled example.

In step 3(e) we compute, for each of $\ell+2$ vertices $v_i \in \mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$, the polynomial $\phi_i(Y) = \Phi_\ell(v_i, Y) = \prod_j (Y - v_{ij})$, where v_{ij} ranges over the $\ell+1$ neighbors of v_i in $G_\ell(\mathbb{F}_p)$. We can then interpolate the coefficients of $\Phi_\ell(X,Y) = \sum_{i,j} c_{ij} X^i Y^j$ as follows: If $\psi_j(X)$ is the unique polynomial of degree at most $\ell+1$ for which $\psi_j(v_i)$ is the coefficient of Y^j in $\phi_i(Y)$, then c_{ij} is the coefficient of X^i in $\psi_j(X)$.

Remark 16. *Weber modular polynomials.* This algorithm can compute modular polynomials for many modular functions besides the j-function; see [7, §7]. This includes the Weber f-function that satisfies $(f(z)^{24} - 16)^3 = f(z)^{24} j(z)$. The modular polynomials $\Phi_\ell^f(X,Y)$ for $f(z)$ are sparser than $\Phi_\ell(X,Y)$ by a factor of 24, and have coefficients whose binary representation is smaller by a factor of approximately 72. Thus the total size of Φ_ℓ^f is roughly 1728 times smaller than Φ_ℓ, and it can be computed nearly 1728 times faster.

Remark 17. *Modular polynomials of composite level.* A generalization of this approach that efficiently computes modular polynomials $\Phi_N(X,Y)$ for composite values of N can be found in [9].

Remark 18. *Evaluating modular polynomials.* Most applications that use $\Phi_\ell(X,Y)$, including all the algorithms we have considered here, only require the instantiated polynomial $\phi(Y) = \Phi_\ell(j(E), Y)$. A space-efficient algorithm for directly computing $\phi(Y)$ without using $\Phi_\ell(X,Y)$ appears elsewhere in this volume [46].

The isogeny volcano algorithm for computing $\Phi_\ell(X,Y)$ has substantially increased the feasible range of ℓ: It is now possible to compute Φ_ℓ with $\ell \approx 10{,}000$, and for Φ_ℓ^f we can handle $\ell \approx 60{,}000$. It has also greatly reduced the time required for these computations, as may be seen in the tables of [7, §8].

Acknowledgements

I am grateful to Gaetan Bisson for his feedback on an early draft of this article, and to the editors for their careful review.

The author was supported by NSF grant DMS-1115455.

References

[1] Juliana Belding, Reinier Bröker, Andreas Enge, and Kristin Lauter, *Computing Hilbert class polynomials*, in van der Poorten and Stein [36], 2008, pp. 282–295. MR 2009j:11200

[2] Daniel J. Bernstein and Jonathan P. Sorenson, *Modular exponentiation via the explicit Chinese remainder theorem*, Math. Comp. **76** (2007), no. 257, 443–454. MR 2007f:11142

[3] Gaetan Bisson, *Computing endomorphism rings of elliptic curves under the GRH*, J. Math. Cryptol. **5** (2011), no. 2, 101–113. MR 2012k:11201

[4] Gaetan Bisson and Andrew V. Sutherland, *Computing the endomorphism ring of an ordinary elliptic curve over a finite field*, J. Number Theory **131** (2011), no. 5, 815–831. MR 2012a:11080

[5] Ljiljana Brankovic, Paul D. Coddington, John F. Roddick, Chris Steketee, James R. Warren, and Andrew L. Wendelborn (eds.), *ACSW Frontiers* 2007: *Proceedings of the Fifth Australasian Symposium on Grid Computing and e-Research* (*AusGrid* 2007), *the Fifth Australasian Information Security Workshop* (*Privacy Enhancing Technologies*) (*AISW* 2007), *and the Australasian Workshop on Health Knowledge Management and Discovery* (*HKMD* 2007), *Ballarat*, *Australia*, *January* 2007, Conferences in Research and Practice in Information Technology, no. 68, Sydney, Australian Computer Society, 2007.

[6] Reinier Bröker, Denis Charles, and Kristin Lauter, *Evaluating large degree isogenies and applications to pairing based cryptography*, in Galbraith and Paterson [21], 2008, pp. 100–112. MR 2012i:94143

[7] Reinier Bröker, Kristin Lauter, and Andrew V. Sutherland, *Modular polynomials via isogeny volcanoes*, Math. Comp. **81** (2012), no. 278, 1201–1231. MR 2012m:11180

[8] Reinier Bröker and Peter Stevenhagen, *Efficient CM-constructions of elliptic curves over finite fields*, Math. Comp. **76** (2007), no. 260, 2161–2179. MR 2008i:11077

[9] Jan Hendrik Bruinier, Ken Ono, and Andrew V. Sutherland, *Class polynomials for nonholomorphic modular functions*, 2013. arXiv 1301.5672 [math.NT]

[10] Denis X. Charles, Kristin E. Lauter, and Eyal Z. Goren, *Cryptographic hash functions from expander graphs*, J. Cryptology **22** (2009), no. 1, 93–113. MR 2010d:94074

[11] Andrew M. Childs, David Jao, and Vladimir Soukharev, *Constructing elliptic curve isogenies in quantum subexponential time*, 2012. arXiv 1012.4019v2 [quant-ph]

[12] David A. Cox, *Primes of the form* $x^2 + ny^2$: *Fermat*, *class field theory and complex multiplication*, John Wiley & Sons, New York, 1989. MR 90m:11016

[13] Andreas Enge, *The complexity of class polynomial computation via floating point approximations*, Math. Comp. **78** (2009), no. 266, 1089–1107. MR 2010h:11097

[14] Andreas Enge and Andrew V. Sutherland, *Class invariants by the CRT method*, in Hanrot et al. [23], 2010, pp. 142–156. MR 2012d:11246

[15] Claus Fieker and David R. Kohel (eds.), *Algorithmic number theory*: *Proceedings of the* 5*th International Symposium* (*ANTS-V*) *held at the University of Sydney*, *July* 7–12, 2002, Lecture Notes in Computer Science, no. 2369, Berlin, Springer, 2002. MR 2004j:11002

[16] Mireille Fouquet, *Anneau d'endomorphismes et cardinalité des courbes elliptiques*: *aspects algorithmiques*, Ph.D. thesis, École polytechnique, 2001. http://www.math.jussieu.fr/~fouquet/Manuscrit.ps.gz

[17] Mireille Fouquet and François Morain, *Isogeny volcanoes and the SEA algorithm*, in Fieker and Kohel [15], 2002, pp. 276–291. MR 2005c:11077

[18] Steven Galbraith and Anton Stolbunov, *Improved algorithm for the isogeny problem for ordinary elliptic curves*, Appl. Algebra Engrg. Comm. Comput. **24** (2013), no. 2, 107–131. MR 3063894

[19] Steven D. Galbraith, *Constructing isogenies between elliptic curves over finite fields*, LMS J. Comput. Math. **2** (1999), 118–138. MR 2001k:11113

[20] Steven D. Galbraith, Florian Hess, and Nigel P. Smart, *Extending the GHS Weil descent attack*, in Knudsen [29], 2002, pp. 29–44. MR 2004f:94060

[21] Steven D. Galbraith and Kenneth G. Paterson (eds.), *Pairing-based cryptography—Pairing* 2008: *Proceedings of the* 2*nd International Conference held at Royal Holloway*, *University of London*, *Egham*, *September* 1–3, 2008, Lecture Notes in Computer Science, no. 5209, Berlin, Springer, 2008. MR 2011j:94001

[22] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, 2nd ed., Cambridge University Press, 2003. MR 2004g:68202

[23] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory*: *Proceedings of the* 9*th Biennial International Symposium* (*ANTS-IX*) *held in Nancy*, *July* 19–23, 2010, Lecture Notes in Computer Science, no. 6197, Berlin, Springer, 2010. MR 2011g:11002

[24] Everett W. Howe and Kiran S. Kedlaya (eds.), *Algorithmic number theory*: *Proceedings of the* 10*th Biennial International Symposium* (*ANTS-X*) *held in San Diego*, *July* 9–13, 2012, The Open Book Series, no. 1, Berkeley, Mathematical Sciences Publishers, 2013, THIS VOLUME.

[25] Sorina Ionica and Antoine Joux, *Pairing the volcano*, in Hanrot et al. [23], 2010, pp. 201–208. MR 2011m:11127

[26] ———, *Pairing the volcano*, Math. Comp. **82** (2013), no. 281, 581–603. MR 2983037

[27] David Jao, Stephen D. Miller, and Ramarathnam Venkatesan, *Do all elliptic curves of the same order have the same difficulty of discrete log*?, in Roy [38], 2005, pp. 21–40. MR 2007e:94060

[28] David Jao and Vladimir Soukharev, *A subexponential algorithm for evaluating large degree isogenies*, in Hanrot et al. [23], 2010, pp. 219–233. MR 2011h:11144

[29] Lars Knudsen (ed.), *Advances in cryptology—EUROCRYPT* 2002: *Proceedings of the* 21*st International Annual Conference on the Theory and Applications of Cryptographic Techniques held in Amsterdam*, *April* 28–*May* 2, 2002, Lecture Notes in Computer Science, no. 2332, Berlin, Springer, 2002. MR 2003m:94074

[30] David Russell Kohel, *Endomorphism rings of elliptic curves over finite fields*, Ph.D. thesis, University of California, Berkeley, 1996, p. 117. http://search.proquest.com/docview/304241260 MR 2695524

[31] Serge Lang, *Elliptic functions*, 2nd ed., Graduate Texts in Mathematics, no. 112, Springer, New York, 1987. MR 88c:11028

[32] H. W. Lenstra, Jr. and Carl Pomerance, *A rigorous time bound for factoring integers*, J. Amer. Math. Soc. **5** (1992), no. 3, 483–516. MR 92m:11145

[33] J. Miret, D. Sadornil, J. Tena, R. Tomàs, and M. Valls, *Isogeny cordillera algorithm to obtain cryptographically good elliptic curves*, in Brankovic et al. [5], 2007, pp. 127–131.

[34] François Morain, *Edwards curves and CM curves*, 2009. arXiv 0904.2243 [math.NT]

[35] Arnold K. Pizer, *Ramanujan graphs and Hecke operators*, Bull. Amer. Math. Soc. (N.S.) **23** (1990), no. 1, 127–137. MR 90m:11063

[36] Alfred J. van der Poorten and Andreas Stein (eds.), *Algorithmic number theory*: *Proceedings of the* 8*th International Symposium* (*ANTS-VIII*) *held in Banff*, *AB*, *May* 17–22, 2008, Lecture Notes in Computer Science, no. 5011, Berlin, Springer, 2008. MR 2009h:11002

[37] Michael O. Rabin, *Probabilistic algorithms in finite fields*, SIAM J. Comput. **9** (1980), no. 2, 273–280. MR 81g:12002

[38] Bimal Roy (ed.), *Advances in cryptology—ASIACRYPT* 2005: *Proceedings of the* 11*th International Conference on the Theory and Application of Cryptology and Information Security held in Chennai*, *December* 4–8, 2005, Lecture Notes in Computer Science, no. 3788, Berlin, Springer, 2005. MR 2007a:94218

[39] René Schoof, *Elliptic curves over finite fields and the computation of square roots mod* p, Math. Comp. **44** (1985), no. 170, 483–494. MR 86e:11122

[40] ———, *Nonsingular plane cubic curves over finite fields*, J. Combin. Theory Ser. A **46** (1987), no. 2, 183–211. MR 88k:14013

[41] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 106, Springer, New York, 1986. MR 87g:11070

[42] ———, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 151, Springer, New York, 1999. MR 96b:11074

[43] Andrew V. Sutherland, *Computing Hilbert class polynomials with the Chinese remainder theorem*, Math. Comp. **80** (2011), no. 273, 501–538. MR 2011k:11177

[44] ———, *Accelerating the CM method*, LMS J. Comput. Math. **15** (2012), 172–204. MR 2970725

[45] ———, *Identifying supersingular elliptic curves*, LMS J. Comput. Math. **15** (2012), 317–325. MR 2988819

[46] ———, *On the evaluation of modular polynomials*, in Howe and Kedlaya [24], 2013, pp. 531–535.

[47] John Tate, *Endomorphisms of abelian varieties over finite fields*, Invent. Math. **2** (1966), 134–144. MR 34 #5829

[48] Jacques Vélu, *Isogénies entre courbes elliptiques*, C. R. Acad. Sci. Paris Sér. A-B **273** (1971), A238–A241. http://gallica.bnf.fr/ark:/12148/bpt6k56191248/f52.image MR 45 #3414

ANDREW V. SUTHERLAND: drew@math.mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.531

msp

On the evaluation of modular polynomials

Andrew V. Sutherland

We present two algorithms that, given a prime ℓ and an elliptic curve $E/\mathbb{F}_q$, directly compute the polynomial $\Phi_\ell(j(E), Y) \in \mathbb{F}_q[Y]$ whose roots are the j-invariants of the elliptic curves that are ℓ-isogenous to E. We do not assume that the modular polynomial $\Phi_\ell(X, Y)$ is given. The algorithms may be adapted to handle other types of modular polynomials, and we consider applications to point counting and the computation of endomorphism rings. We demonstrate the practical efficiency of the algorithms by setting a new point-counting record, modulo a prime q with more than 5,000 decimal digits, and by evaluating a modular polynomial of level $\ell = 100{,}019$.

1. Introduction

Isogenies play a crucial role in the theory and application of elliptic curves. A standard method for identifying (and computing) isogenies uses the classical modular polynomial $\Phi_\ell \in \mathbb{Z}[X, Y]$, which parametrizes pairs of ℓ-isogenous elliptic curves in terms of their j-invariants. More precisely, over a field $\mathbb{F}$ of characteristic not equal to ℓ, the modular equation $\Phi_\ell(j_1, j_2) = 0$ holds if and only if j_1 and j_2 are the j-invariants of elliptic curves defined over $\mathbb{F}$ that are related by a cyclic isogeny of degree ℓ. In practical applications, $\mathbb{F}$ is typically a finite field $\mathbb{F}_q$, and ℓ is a prime, as we shall assume throughout. For the sake of simplicity we assume that q is prime, but this is not essential.

A typical scenario is the following: We are given an elliptic curve $E/\mathbb{F}_q$ and wish to determine whether E admits an ℓ-isogeny defined over $\mathbb{F}_q$, and if so, to identify one or all of the elliptic curves that are ℓ-isogenous to E. This can be achieved by computing the instantiated modular polynomial

$$\phi_\ell(Y) = \Phi_\ell(j(E), Y) \in \mathbb{F}_q[Y],$$

MSC2010: primary 11Y16; secondary 11G15, 11G20.
Keywords: elliptic curves, isogenies, point counting, SEA algorithm.

and finding its roots in $\mathbb{F}_q$ (if any). Each root is the j-invariant of an elliptic curve that is ℓ-isogenous to E over $\mathbb{F}_q$, and every such j-invariant is a root of $\phi_\ell(Y)$.

For large ℓ the main obstacle to obtaining ϕ_ℓ is the size of Φ_ℓ, which is $O(\ell^3 \log \ell)$ bits; storing Φ_ℓ requires several gigabytes for $\ell \approx 10^3$, and many terabytes for $\ell \approx 10^4$ — see [8, Table 1]. In practice, alternative modular polynomials that are smaller than Φ_ℓ by a large constant factor are often used, but their size grows at the same rate, and this quickly becomes the limiting factor, as noted in [15, §5.2] and elsewhere. The 2009 INRIA Project-Team TANC report states:

> "...computing modular polynomials remains the stumbling block for new point counting records. Clearly, to circumvent the memory problems, one would need an algorithm that directly obtains the polynomial specialized in one variable." [26, p. 9]

Here we present just such an algorithm (two in fact), based on the isogeny volcano approach of [8]. Our basic strategy is to compute the instantiated modular polynomial $\phi(Y) = \Phi_\ell(j(E), Y)$ modulo many "suitable" primes p and apply the explicit Chinese remainder theorem modulo q (see Section 2.4 and Section 2.5 for a discussion of the explicit CRT and suitable primes). However, two key issues arise.

First, if we simply lift the j-invariant $j(E)$ from $\mathbb{F}_q \simeq \mathbb{Z}/q\mathbb{Z}$ to $\mathbb{Z}$ and reduce the result modulo p, when we instantiate $\Phi_\ell(j(E), Y)$ the powers of $j(E)$ we compute may correspond to integers that are much larger than the coefficients of Φ_ℓ, forcing us to use many more CRT primes than we would otherwise need. We address this issue by instead exponentiating in $\mathbb{F}_q$, lifting the powers to $\mathbb{Z}$, and then reducing them modulo p. This yields our first algorithm, which is well-suited to situations where q is much larger than ℓ, say $\log q \approx \ell$, as in point-counting applications.

Second, to achieve the optimal space complexity we must avoid computing $\Phi_\ell \bmod p$. Indeed, if $\log q \approx \log \ell$, then $\Phi_\ell \bmod p$ will not be much smaller than $\Phi_\ell \bmod q$. Our second algorithm uses an online approach to avoid storing all the coefficients of $\Phi_\ell \bmod p$ simultaneously. This algorithm is well-suited to situations where $\log q$ is not dramatically larger than $\log \ell$, say $O(\log \ell)$ or $O(\log^2 \ell)$. This occurs, for example, in algorithms that compute the endomorphism ring of an elliptic curve [3], or algorithms to evaluate isogenies of large degree [27].

Under the generalized Riemann hypothesis (GRH), our first algorithm has an expected running time of $O(\ell^3 \log^3 \ell \operatorname{llog} \ell)$ and uses $O(\ell^2 \log \ell + \ell \log q)$ space, assuming $\log q = O(\ell \log \ell)$.[1] This time complexity is the same as (and in practice is faster than) the time to compute Φ_ℓ, and the space complexity is reduced by up to a factor of ℓ. When $\log q \approx \ell$ the space complexity is nearly optimal: quasilinear

[1] See Theorem 4 for a more precise bound. Throughout, we write $\operatorname{llog} n$ for $\log \log n$ and $\operatorname{lllog} n$ for $\log \log \log n$.

in the size of ϕ_ℓ. The second algorithm uses $O(\ell^3(\log q + \log \ell)\log^{1+o(1)}\ell)$ time and $O(\ell \log q + \ell \log \ell)$ space, under the GRH. Its space complexity is optimal for $q = \Omega(\ell)$, and when $\log q = O(\log^{2-\epsilon}\ell)$ its time complexity is better than the time to compute Φ_ℓ. For $\log q \gg \log^2 \ell$ its running time becomes less attractive and the first algorithm may be preferred; alternatively, see Section 3.4 for a hybrid approach.

In conjunction with the SEA algorithm, the first algorithm allows us to compute the cardinality of an elliptic curve modulo a prime q with a heuristic[2] running time of $O(n^4 \log^3 n \log\log n)$, using $O(n^2 \log n)$ space, where $n = \log q$. To our knowledge, all alternative approaches applicable to prime fields increase at least one of these bounds by a factor of n or more. The running time is competitive with SEA implementations that rely on precomputed modular polynomials (as can be found in Magma [4] and PARI [32]), and can easily handle much larger values of q.

As an important practical optimization, we also evaluate modular polynomials $\phi_\ell^f(Y) = \Phi_\ell^f(f(E), Y)$ defined by modular functions $f(z)$ other than the j-function. This includes the Weber ƒ-function, whose modular polynomials are smaller than the classical modular polynomial by a factor of 1728 and can be computed much more quickly (by roughly the same factor). This speedup also applies when computing ϕ_ℓ^f.

To demonstrate the capability of the new algorithms, we use a modified version of the SEA algorithm to count points on an elliptic curve modulo a prime of more than 5,000 decimal digits, and evaluate a modular polynomial of level $\ell = 100{,}019$ modulo a prime of more than 25,000 decimal digits.

2. Background

This section contains a brief summary of background material that can be found in standard references such as [31; 39; 40], or in the papers [8; 42], both of which exploit isogeny volcanoes using a CRT-based approach, as we do here. For the sake of brevity, we recall only the results we need, and only in the generality necessary.

To simplify the presentation, we assume throughout that $\mathbb{F}_p$ and $\mathbb{F}_q$ denote prime fields with $\ell \neq p, q$, and, where relevant, that q is sufficiently large (typically $q > 2\ell$). But this assumption is not needed for our main result; Algorithms 1 and 2 work correctly for any prime q (even $q = \ell$), and can be extended to handle nonprime q.

2.1. ***Isogenies.*** Let E be an elliptic curve defined over a field $\mathbb{F}$. Recall that an *isogeny* is a nonconstant morphism $\psi : E \to \widetilde{E}$ of elliptic curves that is also a group

[2]The heuristic relates to the distribution of Elkies primes and is a standard assumption made when using the SEA algorithm; without it there is no advantage over Schoof's algorithm.

homomorphism from $E(\overline{\mathbb{F}})$ to $\widetilde{E}(\overline{\mathbb{F}})$. The kernel of an isogeny is a finite subgroup of $E(\overline{\mathbb{F}})$, and when ψ is separable, the size of its kernel is equal to its degree. Conversely, every finite subgroup G of $E(\overline{\mathbb{F}})$ is the kernel of a separable isogeny (defined over the fixed field of the stabilizer of G in $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$). We say that ψ is *cyclic* if its kernel is cyclic, and call ψ an *N-isogeny* when it has degree N. Note that an isogeny of prime degree $\ell \neq \mathrm{char}\,\mathbb{F}$ is necessarily cyclic and separable.

The classical modular polynomial Φ_N is the minimal polynomial of the function $j(Nz)$ over the field $\mathbb{C}(j)$, where $j(z)$ is the modular j-function. As a polynomial in two variables, $\Phi_N \in \mathbb{Z}[X, Y]$ is symmetric in X and Y and has the defining property that the roots of $\Phi_\ell(j(E), Y)$ are precisely the j-invariants of the elliptic curves $\widetilde{E}$ that are related to E by a cyclic N-isogeny. In this paper $N = \ell$ is prime, in which case $\Phi_\ell(X, Y)$ has degree $\ell + 1$ in each variable.

If E is given by a short Weierstrass equation $Y^2 = X^3 + a_4X + a_6$, then ψ can be expressed in the form

$$\psi(x, y) = \left(\psi_1(x),\ cy\frac{d}{dx}\psi_1(x)\right)$$

for some $c \in \overline{\mathbb{F}}^*$. When $c = 1$ we say that ψ and its image are *normalized*. Given a finite subgroup G of $E(\overline{\mathbb{F}})$, a normalized isogeny with G as its kernel can be constructed using Vélu's formulae [45], along with an explicit equation for its image $\widetilde{E}$. Conversely, suppose we are given a root $\tilde{\jmath} = j(\widetilde{E})$ of $\phi_\ell(Y) = \Phi_\ell(j(E), Y)$, and also the values of $\Phi_X(j, \tilde{\jmath})$, $\Phi_Y(j, \tilde{\jmath})$, $\Phi_{XX}(j, \tilde{\jmath})$, $\Phi_{XY}(j, \tilde{\jmath})$, and $\Phi_{YY}(j, \tilde{\jmath})$, where $j = j(E)$ and

$$\Phi_X = \frac{\partial}{\partial X}\Phi_\ell, \qquad \Phi_Y = \frac{\partial}{\partial Y}\Phi_\ell,$$
$$\Phi_{XX} = \frac{\partial^2}{\partial X^2}\Phi_\ell, \quad \Phi_{XY} = \frac{\partial^2}{\partial X\,\partial Y}\Phi_\ell, \quad \Phi_{YY} = \frac{\partial^2}{\partial Y^2}\Phi_\ell.$$

To this data we may apply an algorithm of Elkies [13] that computes an equation for $\widetilde{E}$ that is the image of a normalized ℓ-isogeny $\psi : E \to \widetilde{E}$, along with an explicit description of its kernel: the monic polynomial $h_\ell(X)$ whose roots are the abscissae of the nontrivial points in $\ker\psi$; see [19, Algorithm 27]. The quantities $\Phi_{XX}(j, \tilde{\jmath})$, $\Phi_{XY}(j, \tilde{\jmath})$, and $\Phi_{YY}(j, \tilde{\jmath})$ are not strictly necessary; the equation for $\widetilde{E}$ depends only on j, $\tilde{\jmath}$, $\Phi_X(j, \tilde{\jmath})$ and $\Phi_Y(j, \tilde{\jmath})$, and we may then apply algorithms of Bostan et al. [5] to compute $h_\ell(X)$, and an equation for ψ, directly from E and $\widetilde{E}$.

2.2. ***Explicit CM theory.*** Recall that the endomorphism ring of an ordinary elliptic curve E over a finite field $\mathbb{F}_p$ is isomorphic to an order $\mathcal{O}$ in an imaginary quadratic field K. In this situation E is said to have *complex multiplication* (CM) by $\mathcal{O}$. The elliptic curve $E/\mathbb{F}_p$ is the reduction of an elliptic curve $\widehat{E}/\mathbb{C}$ that also has

CM by $\mathcal{O}$. The j-invariant of $\widehat{E}$ generates the ring class field $K_{\mathcal{O}}$ of $\mathcal{O}$, and its minimal polynomial over K is the *Hilbert class polynomial* $H_{\mathcal{O}} \in \mathbb{Z}[X]$, whose degree is the class number $h(\mathcal{O})$.[3] The prime p splits completely in $K_{\mathcal{O}}$, and $H_{\mathcal{O}}$ splits completely in $\mathbb{F}_p[X]$. For $p > 3$, the prime p splits completely in $K_{\mathcal{O}}$ if and only if it satisfies the norm equation $4p = t^2 - v^2 D$, where $D = \operatorname{disc} \mathcal{O}$, and for $D < -4$ the integers $t = t(p)$ and $v = v(p)$ are uniquely determined up to sign.

We define the set

$$\operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p) = \big\{ j(E) : E/\mathbb{F}_p \text{ with } \operatorname{End}(E) \simeq \mathcal{O} \big\},$$

which consists of the roots of $H_{\mathcal{O}}$ in $\mathbb{F}_p$. Let $\iota : \mathcal{O} \hookrightarrow \operatorname{End}(E)$ denote the normalized embedding (so $\iota(\alpha)^* \omega = \alpha\omega$ for all $\alpha \in \mathcal{O}$ and invariant differentials ω on E; see [40, Proposition II.1.1, p. 97]). The ideals of $\mathcal{O}$ act on $\operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ via isogenies as follows. Let $\mathfrak{a}$ be an $\mathcal{O}$-ideal of norm N, and define $E[\mathfrak{a}] = \bigcap_{\alpha \in \mathfrak{a}} \ker \iota(\alpha)$. There is a separable N-isogeny from E to $\widetilde{E} = E/E[\mathfrak{a}]$, and the action of $\mathfrak{a}$ sends $j(E)$ to $j(\widetilde{E})$. Principal ideals act trivially, and this induces a regular action of the class group $\operatorname{Cl}(\mathcal{O})$ on $\operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p)$. Thus $\operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ is a principal homogeneous space, a *torsor*, for $\operatorname{Cl}(\mathcal{O})$.

Writing the $\operatorname{Cl}(\mathcal{O})$-action on the left, we note that if $\mathfrak{a}$ has prime norm ℓ, then $\Phi_\ell(j, [\mathfrak{a}]j) = 0$ for all $j \in \operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p)$. For ℓ not dividing $v(p)$, the polynomial $\phi_\ell(Y) = \Phi_\ell(j, Y)$ has either one or two roots in $\mathbb{F}_p$, depending on whether ℓ ramifies or splits in K. In the latter case, the two roots $[\mathfrak{a}]j$ and $[\mathfrak{a}^{-1}]j$ can be distinguished using the Elkies kernel polynomial $h_\ell(X)$, as described in [6, §5] and [20, §3].

2.3. ***Polycyclic presentations.*** In order to efficiently realize the action of $\operatorname{Cl}(\mathcal{O})$ on $\operatorname{Ell}_{\mathcal{O}}(\mathbb{F}_p)$, it is essential to represent elements of $\operatorname{Cl}(\mathcal{O})$ in terms of a set of generators with small norm. We will choose $\mathcal{O}$ so that $\operatorname{Cl}(\mathcal{O})$ is generated by ideals of norm bounded by $O(1)$, via [8, Theorem 3.3], but these generators will typically not be independent. Thus, as explained in [42, §5.3], we use polycyclic presentations.

Any sequence of generators $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$ for a finite abelian group G defines a polycyclic series

$$1 = G_0 \lhd G_1 \lhd \cdots \lhd G_{k-1} \lhd G_k = G,$$

with $G_i = \langle \alpha_1, \ldots, \alpha_i \rangle$, in which every quotient $G_i/G_{i-1} \simeq \langle \alpha_i \rangle$ is cyclic. We associate to $\boldsymbol{\alpha}$ the sequence of *relative orders* $r(\boldsymbol{\alpha}) = (r_1, \ldots, r_k)$ defined by $r_i = [G_i : G_{i-1}]$. Every element $\beta \in G$ has a unique $\boldsymbol{\alpha}$*-representation* of the form

$$\beta = \boldsymbol{\alpha}^{\boldsymbol{e}} = \alpha_1^{e_1} \cdots \alpha_k^{e_k} \quad (0 \le e_i < r_i).$$

[3] As in [1], we call $H_{\mathcal{O}}$ a Hilbert class polynomial even when $\mathcal{O}$ is not the maximal order.

We also associate to $\boldsymbol{\alpha}$ the matrix of power relations $s(\boldsymbol{\alpha}) = [s_{ij}]$ defined by

$$\alpha_i^{r_i} = \alpha_1^{s_{i,1}} \alpha_2^{s_{i,2}} \cdots \alpha_{i-1}^{s_{i,i-1}} \quad (0 \le s_{ij} < r_j),$$

with $s_{ij} = 0$ for $i \le j$.

We call $\boldsymbol{\alpha}$, together with $r(\boldsymbol{\alpha})$ and $s(\boldsymbol{\alpha})$, a (*polycyclic*) *presentation* for G, and if all the r_i are greater than 1, we say that the presentation is *minimal*. A generic algorithm to compute a minimal polycyclic presentation is given in [42, Algorithm 2.2]. Having constructed such an $\boldsymbol{\alpha}$, we can efficiently enumerate $G = \mathrm{Cl}(\mathbb{O})$ (or the torsor $\mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_q)$, given a starting point), by enumerating $\boldsymbol{\alpha}$-representations.

2.4. ***Explicit CRT.*** Let $p_1, \ldots, p_n$ be primes with product M, let $M_i = M/p_i$, and let $a_i M_i \equiv 1 \bmod p_i$. If $c \in \mathbb{Z}$ satisfies $c \equiv c_i \bmod p_i$, then $c \equiv \sum_i c_i a_i M_i \bmod M$. If $M > 2|c|$, this congruence uniquely determines c. This is the usual CRT method.

Now suppose $M > 4|c|$ and let q be a prime (or any integer). Then we may apply the *explicit CRT mod q* [2, Theorem 3.1] to compute

$$c \equiv \left(\sum_i c_i a_i M_i - rM \right) \bmod q, \tag{1}$$

where r is the closest integer to $\sum_i c_i a_i / p_i$; when computing r, it suffices to approximate each $c_i a_i / p_i$ to within $1/(4n)$, by [2, Theorem 2.2].

As described in [42, §6], we may use the explicit CRT to simultaneously compute $c \bmod q$ for many integers c (the coefficients of ϕ_ℓ, for example), using an *online algorithm*. We first precompute the a_i and $a_i M_i \bmod q$. Then, for each prime p_i, we determine the values c_i for all the coefficients c (by computing $\phi_\ell \bmod p_i$), update two partial sums for each coefficient, one for $\sum c_i a_i M_i \bmod q$ and one for $\sum c_i a_i / p_i$, and then discard the c_i. When the computations for all the p_i have been completed (these may be performed in parallel), we compute r and apply (1) for each coefficient. The space required by the partial sums is just $O(\log q)$ bits per coefficient. See [42, §6] for further details, including algorithms for each step.

2.5. ***Modular polynomials via isogeny volcanoes.*** For distinct primes ℓ and p, we define the *graph of ℓ-isogenies* $\Gamma_\ell(\mathbb{F}_p)$, with vertex set $\mathbb{F}_p$ and edges (j_1, j_2) present if and only if $\Phi_\ell(j_1, j_2) = 0$. Ignoring the connected components of 0 and 1728, the ordinary components of $\Gamma_\ell(\mathbb{F}_p)$ are *ℓ-volcanoes* [18; 30], a term we take to include cycles as a special case [42]. In this paper we focus on ℓ-volcanoes of a particular form, for which we can compute $\Phi_\ell \bmod p$ very quickly, via [8, Algorithm 2.1].

Let $\mathbb{O}$ be an order in an imaginary quadratic field K with maximal order $\mathbb{O}_K$, let ℓ be an odd prime not dividing $[\mathbb{O}_K : \mathbb{O}]$, let $\mathbb{O}'$ be the order of index l in $\mathbb{O}$, and assume

$D = \operatorname{disc} \mathbb{O} < -4$. Let p be a prime of the form $4p = t^2 - \ell^2 v^2 D$ with $\ell \nmid v$ and $p \equiv 1 \bmod \ell$. Then p splits completely in the ring class fields of $\mathbb{O}$ and $\mathbb{O}'$, but not in the ring class field of the order of index ℓ^2 in $\mathbb{O}$. The requirement $p \equiv 1 \bmod \ell$ ensures that for $j(E) \in \mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)$ we can choose E so that $E[\ell] \subset E(\mathbb{F}_p)$, which is critical to the efficiency of both the algorithm in [8] and our algorithms here.

The components of $\Gamma_\ell(\mathbb{F}_p)$ that intersect $\mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)$ are isomorphic ℓ-volcanoes with two levels: the *surface*, whose vertices lie in $\mathrm{Ell}_{\mathbb{O}}(\mathbb{F}_p)$, and the *floor*, whose vertices lie in $\mathrm{Ell}_{\mathbb{O}'}(\mathbb{F}_p)$. Each vertex on the surface is connected to $1 + \left(\frac{D}{\ell}\right) = 0, 1$ or 2 *siblings* on the surface, and $\ell - \left(\frac{D}{\ell}\right)$ *children* on the floor. An example with $\ell = 7$ and $\left(\frac{D}{\ell}\right) = 1$ is shown below:

Provided that $h(\mathbb{O}) \geq \ell + 2$, this set of ℓ-volcanoes contains enough information to completely determine $\Phi_\ell \bmod p$. This is the basis of the algorithm in [8, Algorithm 2.1], which we adapt here. Selecting a sufficiently large set of such primes p allows one to compute Φ_ℓ over $\mathbb{Z}$ (via the CRT), or modulo an arbitrary prime q (via the explicit CRT). In order to achieve the best complexity bounds, it is important to choose both the order $\mathbb{O}$ and the primes p carefully. We thus introduce the following definitions, in which c_1 and c_2 are fixed constants that do not depend on ℓ or $\mathbb{O}$. (In our implementation we used $c_1 = 1.5$ and $c_2 = 256$.)

Definition 1. Let $\mathbb{O}$ be a quadratic order with discriminant $D = u^2 D_0 < 0$, with D_0 fundamental, and let $c_1, c_2 > 1$ be constants. We say that $\mathbb{O}$ is *suitable* for ℓ if

(1) $\ell + 2 \leq h(\mathbb{O}) \leq c_1 \ell$,

(2) $4 < |D_0| \leq c_2^2$,

(3) $\ell^2 \leq |D| \leq c_2^2 \ell^2$,

(4) $\gcd(u, 2\ell D) = 1$, and

(5) $l_0 < \min(c_2, \ell)$ for all primes $l_0 \mid u$.

This definition combines the criteria in [8, Definition 4.2] and [8, Theorem 5.1]. Provided that c_1 and c_2 are not too small, suitable orders exist for every odd prime ℓ; with $c_1 = 4$ and $c_2 = 16$, for example, we may use orders with $D = -7 \cdot 3^{2n}$ for all $\ell > 3$. Ideally we want c_1 to be as close to 1 as possible, but this makes it harder to find suitable orders. For the asymptotic analysis, any values of c_1 and c_2 will do.

Definition 2. A prime p is *suitable* for ℓ and $\mathbb{O}$ if $p \equiv 1 \bmod \ell$ and p satisfies $4p = t^2 - \ell^2 v^2 D$ for some $t, v \in \mathbb{Z}$ with $\ell \nmid v$ and $\omega(v) \leq 2 \log(\log v + 3)$.

The function $\omega(v)$ counts the distinct prime divisors of v. The bound on $\omega(v)$ ensures that if $\mathbb{O}$ is suitable for ℓ then many small primes split in $\mathbb{O}$ and do not divide u or v. Such primes allow us to more efficiently enumerate $\mathrm{Cl}(\mathbb{O})$ and $\mathrm{Cl}(\mathbb{O}')$.

2.6. ***Selecting primes with the GRH.*** In order to apply the isogeny volcano method to compute $\Phi_\ell \bmod q$ (or $\phi_\ell \bmod q$, as we shall do), we need a sufficiently large set S of suitable primes p. We deem S to be sufficiently large whenever

$$\sum_{p \in S} \log p \geq B + \log 4,$$

where B is an upper bound on the logarithmic height of the integers whose reductions mod q we wish to compute with the explicit CRT. For $\Phi_\ell(X,Y) = \sum_{i,j} a_{ij} X^i Y^j$, we may bound $h(\Phi_\ell) = \log \max_{i,j} |a_{ij}|$ using

$$h(\Phi_\ell) \leq 6\ell \log \ell + 18\ell, \tag{2}$$

$$h(\Phi_\ell) \leq 6\ell \log \ell + 16\ell + 14\sqrt{\ell} \log \ell, \tag{3}$$

as proved in [9]. (We prefer the latter bound when $\ell > 3187$.)

Heuristically (and in practice), it is easy to construct the set S. Given an order $\mathcal{O}$ of discriminant D suitable for ℓ, we fix $v = 2$ if $D \equiv 1 \bmod 8$ and $v = 1$ otherwise, and for increasing $t \equiv 2 \bmod \ell$ of correct parity we test whether $p = (t^2 - v^2 \ell^2 D)/4$ is prime. We add each prime value of p to S, and stop when S is sufficiently large.

Unfortunately, we cannot prove that this method will find *any* primes, even under the GRH. Instead, we use Algorithm 6.2 in [8], which picks an upper bound x and generates random integers t and v in suitable intervals to obtain candidate primes $p = (t^2 - v^2 \ell^2 D)/4 \leq x$ that are then tested for primality. The algorithm periodically increases x, so its expected running time is $O(B^{1+\epsilon})$, even without the GRH. To ensure that the bound on $\omega(v)$ in Definition 2 is satisfied, unsuitable v's are discarded; this occurs with negligible probability.

Under the GRH, there are effective constants $c_3, c_4 > 0$ such that $x \geq c_3 \ell^6 \log^4 \ell$ guarantees at least $c_4 \ell^3 \log^3 \ell$ suitable primes less than x, by [8, Theorem 4.4]. Asymptotically, this is far more than the $O(\ell)$ primes we need to compute $\Phi_\ell \bmod q$. Here we may consider larger values of B, and in general, $x = O(B^2 + \ell^6 \log^4 \ell)$ suffices. We note that S contains $O(B/\log B)$ primes (unconditionally), and under the GRH we have $\log p = O(\log B + \log \ell)$ for all $p \in S$.

3. Algorithms

Let q be a prime and let E be an elliptic curve over $\mathbb{F}_q$. A simple algorithm to compute $\phi_\ell(Y) = \Phi_\ell(j(E), Y) \in \mathbb{F}_q[Y]$ with the explicit CRT works as follows. Let $\hat{\jmath}$ be the integer in $[0, q-1]$ corresponding to $j(E) \in \mathbb{F}_q \simeq \mathbb{Z}/q\mathbb{Z}$. For a sufficiently large set S of suitable primes p, compute $\Phi_\ell(X,Y) \bmod p$ using the isogeny volcano algorithm and evaluate $\Phi_\ell(\hat{\jmath}, Y) \bmod p$ to obtain $\bar{\phi}_\ell \in \mathbb{F}_p[Y]$, and use the explicit CRT mod q to eventually obtain $\phi_\ell \in \mathbb{F}_q[Y]$.

This naïve algorithm suffers from two significant defects. The most serious is that we may now require a much larger set S than is needed to compute $\Phi_\ell \bmod q$.

Compared to the coefficients of Φ_ℓ, which have height $h(\Phi_\ell) = O(\ell \log \ell)$ bounded by inequalities (2) and (3), we now need to use the $O(\ell \log \ell + \ell \log q)$ bound

$$h(\Phi_\ell(\hat{\jmath}, Y)) \leq h(\Phi_\ell) + (\ell + 1) \log q + \log(\ell + 2), \tag{4}$$

since $\Phi_\ell(\hat{\jmath}, Y)$ involves powers of $\hat{\jmath}$ up to $\hat{\jmath}^{\ell+1}$.

If $\log q$ is comparable to $\log \ell$, then the difference between the bounds in inequalities (2) and (3) and the bound in inequality (4) may be negligible. But when $\log q$ is comparable to ℓ, using the bound in inequality (4) increases the running time dramatically. This issue is addressed by Algorithm 1.

The second defect of the naïve algorithm is that although its space complexity may be significantly better than the $O(\ell^2 \log q)$ space required to compute $\Phi_\ell \bmod q$, it is still quasiquadratic in ℓ. But the size of ϕ_ℓ is linear in ℓ, so we might hope to do better, and indeed we can. This is achieved by Algorithm 2.

A hybrid approach that combines aspects of both algorithms is discussed in Section 3.4.

3.1. ***Algorithm 1.*** The increase in the height bound from inequalities (2) and (3) to inequality (4) is caused by the fact that *we are exponentiating in the wrong ring*. Rather than lifting $j(E) \in \mathbb{F}_q$ to the integer $\hat{\jmath}$ and computing powers of its reduction in $\mathbb{F}_p$ (which simulates powering in $\mathbb{Z}$), we should instead compute powers $j(E)$, $j(E)^2, \ldots, j(E)^{\ell+1}$ in $\mathbb{F}_q$, lift these values to integers $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{\ell+1}$, and work with their reductions in $\mathbb{F}_p$, as in [43, §4.4] (a similar strategy is used in [28]). Of course the reductions of $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{\ell+1}$ need not correspond to powers of any particular element in $\mathbb{F}_p$; nevertheless, if we simply replace each occurrence of X^i in the modular polynomial $\Phi_\ell(X, Y) \bmod p$ with $\hat{x}_i \bmod p$, we achieve the same end result using a much smaller height bound.

We now present Algorithm 1 to compute $\phi(Y) = \phi_\ell(Y) = \Phi_\ell(j(E), y)$. If desired, the algorithm can also compute the polynomials

$$\phi_X(Y) = \frac{\partial \Phi_\ell}{\partial X}(j(E), Y) \quad \text{and} \quad \phi_{XX}(Y) = \frac{\partial^2 \Phi_\ell}{\partial X^2}(j(E), Y),$$

which may be used to compute normalized isogenies, as described in Section 3.8.

Algorithm 1.

Input: An odd prime ℓ, a prime q, and $j(E) \in \mathbb{F}_q$.

Output: The polynomial $\phi(Y) = \Phi_\ell(j(E), Y) \in \mathbb{F}_q[Y]$, and, optionally, $\phi_X(Y)$ and $\phi_{XX}(Y)$.

1. Select an order $\mathbb{O}$ suitable for ℓ and a set of suitable primes S (see Section 2.6), using the height bound $B = 6\ell \log \ell + 18\ell + \log q + 3 \log(\ell + 2)$.
2. Compute the Hilbert class polynomial $H_\mathbb{O}(X)$ via [42, Algorithm 2].

3. Perform CRT precomputation mod q using S (see Section 2.4).
4. Compute integers $\widehat{x}_i \in [0, q-1]$ such that $\widehat{x}_i \equiv j(E)^i \bmod q$, for $0 \le i \le \ell+1$.
5. For each prime $p \in S$:
   (a) Compute $\Phi_\ell(X, Y) \bmod p$ using $H_{\mathbb{O}}$, via [8, Algorithm 2.1].
   (b) Compute
   $$\overline{\phi}(Y) = \sum_{i,j} a_{ij} \widehat{x}_i Y^j \bmod p,$$
   where $\Phi_\ell(X, Y) = \sum_{i,j} a_{ij} X^i Y^j$.
   (c) (Optional.) Compute
   $$\overline{\phi}_X(Y) = \sum_{i,j} i a_{ij} \widehat{x}_i Y^j \bmod p$$
   and
   $$\overline{\phi}_{XX}(Y) = \sum_{i,j} i(i-1) a_{ij} \widehat{x}_i Y^j \bmod p.$$
   (d) Update CRT sums for each coefficient c_i of $\overline{\phi}$ (and of $\overline{\phi}_X$ and $\overline{\phi}_{XX}$).
6. Perform CRT postcomputation to obtain ϕ (and ϕ_X and ϕ_{XX}) mod q.
7. Output ϕ and (optionally) ϕ_X and ϕ_{XX}.

Proposition 3. *The output $\phi(Y)$ of Algorithm 1 is equal to $\Phi_\ell(j(E), Y)$ (and $\phi_X(Y) = (\partial\Phi_\ell/\partial X)(j(E), Y)$ and $\phi_{XX}(Y) = (\partial^2\Phi_\ell/\partial X^2)(j(E), Y)$).*

Proof. Let $\varphi = \Phi_\ell(\widehat{\jmath}, Y) \in \mathbb{F}_q[Y]$. Let $\widehat{x}_i \in \mathbb{Z}$ be as in step 4. Write Φ_ℓ as $\sum_{i,j} a_{ij} X^i Y^j$, with $a_{ij} \in \mathbb{Z}$ and let $\widehat{\phi} = \sum_{i,j} a_{ij} \widehat{x}_i Y^j \in \mathbb{Z}[Y]$. Then $\varphi \equiv \widehat{\phi} \bmod q$, and $\overline{\phi} \equiv \widehat{\phi} \bmod p$. To prove $\phi = \varphi$, we only need to show $h(\widehat{\phi}) \le B$. We have

$$\left| \sum_i a_{ij} \widehat{x}_i \right| \le (\ell+2) q \exp h(\Phi_\ell),$$

for $0 \le j \le \ell+1$, hence $h(\widehat{\phi}) \le B$. The proofs for ϕ_X and ϕ_{XX} are analogous. We note that the last term in B can be reduced to $\log(\ell+2)$ if ϕ_X and ϕ_{XX} are not being computed. □

Theorem 4. *Assume the GRH. Then the expected running time of Algorithm 1 is $O(\ell^2 B \log^2 B \operatorname{llog} B)$, where $B = O(\ell \log \ell + \log q)$ is as specified in step 1. The algorithm uses $O(\ell \log q + \ell^2 \log B)$ space.*

Proof. We use $\mathsf{M}(n) = O(n \log n \operatorname{llog} n)$ to denote the cost of multiplication [35]. For step 1, we assume the time spent selecting $\mathbb{O}$ is negligible (as noted in Section 2.5, one may simply choose orders with discriminants of the form $D = -7 \cdot 3^{2n}$), and under the GRH the expected time to construct S is $O(B^{1+\epsilon})$, using $O(B)$ space, as explained in Section 2.6. Step 2 uses $O(\ell^{2+\epsilon})$ expected time and $O(\ell(\log \ell + \log q))$ space, by [42, Theorem 1], since $h(D) = O(\ell)$. An analysis as in [42, §6.3] shows

that the total cost of all CRT operations is $O(\ell \mathsf{M}(B) \log B)$ time and $O(\ell \log q)$ space. Step 4 uses $O(\ell \mathsf{M}(\log q))$ time and $O(\ell \log q)$ space.

The set S contains $O(B/\log B)$ primes p, and under the GRH, $\log p = O(\log B)$; see Section 2.6. Step 5(a) dominates the cost per p, taking $O(\ell^2 \log^3 B \operatorname{llog} B)$ expected time and $O(\ell^2 \log B)$ space, by [8]. This yields an $O(\ell^2 B \log^2 B \operatorname{llog} B)$ bound for step 5, which dominates, and the total space is $O(\ell \log q + \ell^2 \log B)$. □

When $\log q = \Theta(\ell)$, the time bound in Theorem 4 reduces to $O(\ell^3 \log^3 \ell \operatorname{llog} \ell)$, the same as the time to compute $\Phi_\ell \bmod q$, and the space bound is $O(\ell \log \ell \log q)$, which is within an $O(\log \ell)$ factor of the best possible.

3.2. ***Algorithm 2.*** We now present Algorithm 2, which for $q > \ell$ has optimal space complexity $O(\ell \log q)$. When q is reasonably small, say $\log q = o(\log^2 \ell)$, Algorithm 2 is also faster than Algorithm 1, but when $\log q$ is large it may be much slower, since it uses the same height bound — inequality (4) — as the naïve approach (see Section 3.4 for a hybrid approach). The computation of $\overline{\phi} \in \mathbb{F}_p[Y]$ is more intricate, so we present it separately as Algorithm 2.1. Unlike Algorithm 1, it is not so easy to also compute ϕ_X and ϕ_{XX}, but an alternative method to compute normalized isogenies using Algorithm 2 is given in Section 3.8.

Algorithm 2.

Input: An odd prime ℓ, a prime q, and $j(E) \in \mathbb{F}_q$.

Output: The polynomial $\phi(Y) = \Phi_\ell(j(E), Y) \in \mathbb{F}_q[Y]$.

1. Select an order $\mathbb{O}$ suitable for ℓ and a suitable set of primes S (see Section 2.6), using the height bound $B = 6\ell \log \ell + 18\ell + (\ell + 1) \log q + \log(\ell + 2)$.
2. Compute the Hilbert class polynomial $H_\mathbb{O}$ via [42, Algorithm 2].
3. Perform precomputation for the explicit CRT mod q using S.
4. Let $\widehat{\jmath}$ be the integer in $[0, q-1]$ congruent to $j(E) \bmod q$.
5. For each prime $p \in S$:
   (a) Compute $\overline{\phi}(Y) = \Phi_\ell(\widehat{\jmath}, Y) \bmod p$ using $\mathbb{O}$ and $H_\mathbb{O}$ via Algorithm 2.1.
   (b) Update CRT sums for each coefficient c_i of $\overline{\phi}$.
6. Perform postcomputation for the explicit CRT to obtain $\phi \in \mathbb{F}_q[X]$.
7. Output ϕ.

Proposition 5. *The output $\phi(Y)$ of Algorithm 2 is equal to $\Phi_\ell(j(E), Y)$.*

Proof. This follows immediately from Proposition 7 below and the bound

$$h(\Phi_\ell(\widehat{\jmath}, Y)) = \log \max_j \left| \sum_i a_{ij} \widehat{\jmath}^{\,i} \right| \le \log(\ell + 2) + (\ell + 1) \log q + h(\Phi_\ell) \le B$$

on the height of $\Phi_\ell(\widehat{\jmath}, Y) \in \mathbb{Z}[Y]$. □

Theorem 6. *Assume the GRH and that* $\log q = O(\ell^k)$ *for some constant* k. *The expected running time of Algorithm 2 is* $O(\ell^3(\log q + \log \ell)\log \ell \operatorname{llog}^2 \ell \operatorname{lllog}^2 \ell)$ *and it uses* $O(\ell \log q + \ell \log \ell)$ *space.*

Proof. As in the proof of Theorem 4, the expected running time is dominated by the time to compute $\bar{\phi}(Y)$, which by Theorem 8 is $O(\ell^2 \log^2 p \operatorname{llog}^2 p \operatorname{lllog}^2 p)$. There are $O(B/\log B)$ primes $p \in S$, and under the GRH we have $\log p = O(\log B) = O(\log \ell)$. The space complexity is dominated by the $O(B) = O(\ell \log \ell + \ell \log q)$ size of S. □

3.3. ***Algorithm 2.1.*** The algorithm in [8, Algorithm 2.1] computes $\Phi_\ell \bmod p$ by enumerating the sets $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ and $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$, where $\mathcal{O}' = \mathbb{Z} + \ell\mathcal{O}$, the latter of which contains approximately ℓ^2 elements. To achieve a space complexity that is quasilinear in ℓ, we cannot afford to store the entire set $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$. We must compute $\Phi_\ell(\widehat{\jmath}, Y) \bmod p$ using an online algorithm, processing each $j_k \in \mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ as we enumerate it, and then discarding it. Let us consider how this may be done.

Let $y_1, \ldots, y_{h(\mathcal{O})}$ be the elements of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$, as enumerated using a polycyclic presentation $\boldsymbol{\alpha}$ for $\mathrm{Cl}(\mathcal{O})$. Each y_i is ℓ-isogenous to a set S_i of *siblings* in $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$, and to a set C_i of *children* in $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$; see Section 2.5. Thus we have

$$\Phi_\ell(X, y_i) = \Bigg(\prod_{\widetilde{\jmath} \in S_i} (X - \widetilde{\jmath})\Bigg)\Bigg(\prod_{\widetilde{\jmath} \in C_i} (X - \widetilde{\jmath})\Bigg).$$

The siblings can be readily identified in our enumeration of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ using the CM action (see Section 2.2). To identify the children, we need to be able to determine, for any given $j \in \mathcal{O}'$, the set C_i in which it lies. Each C_i is a subset of the torsor $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ corresponding to a coset of the subgroup $C \subset \mathrm{Cl}(\mathcal{O}')$ generated by the ideals of norm ℓ^2; indeed, two children have the same parent if and only if they are related by an isogeny of degree ℓ^2 (the composition of two ℓ-isogenies).

The group $\mathrm{Cl}(\mathcal{O}')$ acts on the cosets of C, and we need to compute this action explicitly in terms of the polycyclic presentation $\boldsymbol{\beta}$ used to enumerate $\mathrm{Cl}(\mathcal{O}')$. This problem is addressed by a generic group algorithm in the next section that computes a polycyclic presentation $\boldsymbol{\gamma}$ for the quotient $\mathrm{Cl}(\mathcal{O}')/C$, along with the $\boldsymbol{\gamma}$-representation of the image of each generator in $\boldsymbol{\beta}$.

As we enumerate the elements j_k of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$, starting from a child j_1 of y_1 obtained via Vélu's algorithm, we keep track of the element $\delta_k \in \mathrm{Cl}(\mathcal{O}')$ whose action sends j_1 to j_k. The image of δ_k in $\mathrm{Cl}(\mathcal{O}')/C$ is the coset of C corresponding to the set C_i containing j_k, and we simply identify C_i with the i-th element of $\mathrm{Cl}(\mathcal{O}')/C$ as enumerated by $\boldsymbol{\gamma}$ (in the lexicographic ordering of $\boldsymbol{\gamma}$-representations).

Thus we can compute the polynomials $\phi_i(X) = \Phi_\ell(X, y_i)$ as we enumerate $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ by accumulating a partial product of linear factors for each ϕ_i. But since

our goal is to evaluate $z_i = \phi_i(\hat{\jmath}) \bmod p$, we simply substitute $x = \hat{\jmath} \bmod p$ into each linear factor, as we compute it, and accumulate the partial product in z_i.

Having computed the values z_i for $1 \le i \le \ell + 2$, we interpolate the unique polynomial $\phi(Y)$ of degree at most $\ell + 1$ for which $\phi(y_i) = z_i$, using Lagrange interpolation. This polynomial must be $\Phi_\ell(\hat{\jmath}, Y)$. We now give the algorithm.

Algorithm 2.1.

Input: An odd prime ℓ, a suitable order $\mathbb{O}$, a suitable prime p, and $x \in \mathbb{F}_p$.

Output: The polynomial $\phi(Y) = \Phi_\ell(x, Y) \in \mathbb{F}_p[Y]$.

1. Compute presentations $\boldsymbol{\alpha}$ of $\mathrm{Cl}(\mathbb{O})$ and $\boldsymbol{\beta}$ of $\mathrm{Cl}(\mathbb{O}')$ suitable for p.
2. Represent generators of the subgroup $C \subset \mathrm{Cl}(\mathbb{O}')$ defined above in terms of $\boldsymbol{\beta}$.
3. Compute the presentation $\boldsymbol{\gamma}$ of $\mathrm{Cl}(\mathbb{O}')/C$ derived from $\boldsymbol{\beta}$, via Algorithm 3.
4. Find a root w_1 of $H_\mathbb{O} \bmod p$ (compute $H_\mathbb{O} \bmod p$ if needed).
5. Enumerate $\mathrm{Ell}_\mathbb{O}(\mathbb{F}_p)$ as $w_1, w_2, \ldots, w_{h(\mathbb{O})}$ using $\boldsymbol{\alpha}$.
6. Obtain $j_1 \in \mathrm{Ell}_{\mathbb{O}'}(\mathbb{F}_p)$ from w_1 using Vélu's algorithm.
7. Set $z_i \leftarrow 1$ and $y_i \leftarrow \texttt{null}$ for $1 \le i \le \ell + 2$.
8. For each $j_k = \delta_k j_1$ in $\mathrm{Ell}_{\mathbb{O}'}(\mathbb{F}_p)$ enumerated using $\boldsymbol{\beta}$:
   (a) Compute the index i of δ_k in the $\boldsymbol{\gamma}$-enumeration of $\mathrm{Cl}(\mathbb{O}')/C$. If $i > \ell + 2$ then proceed to the next j_k, skipping steps (b) and (c) below.
   (b) If $y_i = \texttt{null}$ then set y_i to the ℓ-parent of j_k (via Vélu's algorithm) and for each ℓ-sibling $\tilde{\jmath}$ of y_i in $\mathrm{Ell}_\mathbb{O}(\mathbb{F}_p)$ set $z_i \leftarrow z_i(x - \tilde{\jmath})$.
   (c) Set $z_i \leftarrow z_i(x - j_k)$.
9. Interpolate $\phi \in \mathbb{F}_p[Y]$ such that $\deg \phi \le \ell + 1$ and $\phi(y_i) = z_i$ for $1 \le i \le \ell + 2$.
10. Output ϕ.

The value `null` assigned to y_i in step 7 is used to indicate that the value of y_i is not yet known. Each y_i is eventually set to a distinct $w_j \in \mathrm{Ell}_\mathbb{O}(\mathbb{F}_p)$.

Remark. In practical implementations, Algorithm 2 selects the primes $p \in S$ so that the presentations $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are the same for every p and precomputes them (the only reason they might not be the same is the presence of prime ideals whose norm divides $v = v(p)$, but in practice we fix $v \le 2$, as discussed in Section 2.6).

Proposition 7. *Algorithm 2.1 outputs* $\phi(Y) = \Phi_\ell(x, Y) \bmod p$.

Proof. Let $\varphi(Y) = \Phi_\ell(x, Y)$. It follows from the discussion above that Algorithm 2.1 computes $z_i = \Phi_\ell(x, y_i)$ for $1 \le i \le \ell + 2$. Thus $\phi(y_i) = z_i = \varphi(y_i)$ for $\ell + 2$ values $y_i \in \mathrm{Ell}_\mathbb{O}(\mathbb{F}_p)$, and these values are necessarily distinct. The polynomials ϕ and φ both have degree at most $\ell + 1$, therefore $\phi = \varphi$. □

Theorem 8. *Assume the GRH. Algorithm 2.1 runs in $O(\ell^2 n^2 \log^2 n \operatorname{llog}^2 n)$ expected time using $O(\ell n)$ space, where $n = \log p$.*

Proof. The time complexity is dominated by step 8, which enumerates the $O(\ell^2)$ elements of $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ using $\boldsymbol{\beta}$. By [8, Theorem 5.1] and the suitability of $\mathcal{O}$ and p, we may assume each $\beta_i = [\mathfrak{b}_i]$, where $\mathfrak{b}_i$ has prime norm $b_i = O(\log n \operatorname{llog} n)$. Using Kronecker substitution and probabilistic root-finding [21], the expected time to find the (at most 2) roots of $\Phi_{b_i}(j_k, Y)$ is $O(n\mathsf{M}(n \log n \operatorname{llog} n))$, which dominates the cost for each j_k. Applying $\mathsf{M}(n) = O(n \log n \operatorname{llog} n)$ and multiplying by ℓ^2 yields the desired time bound. Taking into account $h(\mathcal{O}) = O(\ell)$ and $p > \ell$, the computation of $H_{\mathcal{O}}$ mod p uses $O(\ell n)$ space, by [42, Theorem 1], and this bounds the total space. □

3.4. ***A hybrid approach.*** Algorithm 2 achieves an essentially optimal space complexity, but its time complexity is attractive only when $\log q$ is not too large, say $\log q = O(\log^2 \ell)$. Algorithm 1 has an excellent time complexity, but achieves an optimal space complexity only when $\log q$ is very large, say $\log q = \Omega(\ell \log \ell)$. To address the intermediate range, we present a hybrid approach suggested by Daniel Kane that has the same space complexity as Algorithm 2 and a time complexity that is within a polylogarithmic factor of the time complexity of Algorithm 1.

The strategy is to replace the computation of $\overline{\phi}(Y) = \sum_{i,j} a_{ij} \hat{x}_i Y^j \bmod p$ in step 5 of Algorithm 1 with Algorithm 2.2 below. Algorithm 2.2 is similar to Algorithm 2.1, but rather than accumulating $\ell + 2$ values z_i in parallel, we compute them individually by enumerating the each of the sets C_i of children y_i in turn.

Algorithm 2.2.

Input: An odd prime ℓ, suitable order $\mathcal{O}$, suitable prime p, and $x_1, \ldots, x_{\ell+1} \in \mathbb{F}_p$.

Output: $\phi(Y) = \sum_{i,j} a_{ij} x_i Y^j \in \mathbb{F}_p[Y]$, where $\Phi_\ell(X, Y) = \sum_{i,j} a_{ij} X^i Y^j$.

1. Compute presentations $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ as in Algorithm 2.1.
2. Find a root y_1 of $H_{\mathcal{O}}$ mod p (compute $H_{\mathcal{O}}$ mod p if needed).
3. Enumerate $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_q)$ as $y_1, y_2, \ldots, y_{h(\mathcal{O})}$ using $\boldsymbol{\alpha}$.
4. Obtain $j_1 \in \mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_q)$ from y_1 using Vélu's algorithm.
5. For i from 1 to $\ell + 2$ do the following:
   (a) Use $\boldsymbol{\alpha}$ to compute the set S_i of siblings of y_i in $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$.
   (b) Use $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to compute the set C_i of children of y_i in $\mathrm{Ell}_{\mathcal{O}'}(\mathbb{F}_p)$ (see below).
   (c) Compute $\phi_i(X) = \prod_{\tilde{\jmath} \in S_i} (X - \tilde{\jmath}) \prod_{\tilde{\jmath} \in C_i} (X - \tilde{\jmath}) = \sum c_{ik} X^k \bmod p$.
   (d) Compute $z_i = \sum_k c_{ik} x_k$.
6. Interpolate $\phi \in \mathbb{F}_p[Y]$ such that $\deg \phi \le \ell + 1$ and $\phi(y_i) = z_i$ for $1 \le i \le \ell + 2$.
7. Output ϕ.

To compute the set C_i in step 5(b), for each $\widetilde{J} \in C_i$ we determine the $\delta \in \mathrm{Cl}(\mathbb{O}')$ for which $\widetilde{J} = \delta j_1$. Under the GRH, it follows from [11, Theorem 2.1] that we can express δ in the form $\delta = [\mathfrak{p}_1 \cdots \mathfrak{p}_t]$, where the ideals $\mathfrak{p}_i$ have prime norms bounded by $\log^c \ell$, for any $c > 2$, with $t = O(\log \ell)$. Assuming $\log p = O(\log \ell)$, this implies that we can compute each $\widetilde{J}$ in $O(\log^{6+\epsilon} \ell)$ expected time, for any $\epsilon > 0$.

Proposition 9. *Algorithm 2.2 outputs* $\phi(Y) = \sum_{i,j} a_{ij} x_i Y^j$.

Proof. Let $\varphi(y) = \sum_{i,j} a_{ij} x_i Y^j$. The roots of $\phi_i(X)$ are the roots of $\Phi_\ell(X, y_i)$, thus $\sum_k c_{ik} X_k = \sum_{k,j} a_{kj} X^k y_i^j$, and we have $c_{ik} = \sum_j a_{kj} y_i^j$. It follows that $\phi(y_i) = z_i = \sum_k \sum_j a_{kj} x_k y_i^j = \varphi(y_i)$. Since $\phi(Y)$ and $\varphi(Y)$ both have degree at most $\ell + 1$ and agree at $\ell + 2$ distinct values y_i, they must be equal. □

Theorem 10. *Assume the GRH and that* $\log q = O(\ell \log \ell)$. *If Algorithm 1 uses Algorithm 2.2 to compute* $\bar{\phi}(Y)$ *in step* 5, *its expected running time is* $O(\ell^3 \log^{6+\epsilon} \ell)$ *using* $O(\ell \log q + \ell \log \ell)$ *space.*

Proof. It suffices to show that if $\log p = O(\log \ell)$, then Algorithm 2.2 runs in $O(\ell^2 \log^{6+\epsilon} \ell)$ expected time using $O(\ell \log \ell)$ space. The space bound is clear. For the time bound, the cost of step 5(b) is $O(\ell \log^{6+\epsilon} \ell)$ (see above), yielding an $O(\ell^2 \log^{6+\epsilon} \ell)$ bound on the expected time for step 5, which dominates. □

The extra logarithmic factors make the hybrid approach significantly slower than Algorithm 1 in practice, but it does allow us to achieve an essentially optimal space complexity with a quasicubic running time across the entire range of parameters.

3.5. ***Computing a polycyclic presentation for a quotient group.*** We now give a generic algorithm to derive a polycyclic presentation $\boldsymbol{\gamma}$ for a quotient of finite abelian groups G/H. This presentation can be used to efficiently compute in G/H, and to compute the image of elements of G, as required by Algorithm 2.1.

Algorithm 3.

Input: A minimal polycyclic presentation $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ for a finite abelian group G and a subgroup $H = \langle \alpha_1, \ldots, \alpha_t \rangle$, with each α_i specified in terms of $\boldsymbol{\beta}$.

Output: A polycyclic presentation $\boldsymbol{\gamma}$ for G/H, with $\gamma_i = [\beta_i]$ for each $\beta_i \in \boldsymbol{\beta}$.

1. Derive a polycyclic presentation $\boldsymbol{\alpha}$ for H from $\alpha_1, \ldots, \alpha_t$ by using [42, Algorithm 2.2].
2. Enumerate H using $\boldsymbol{\alpha}$ and create a lookup table T_H to test membership in H.
3. Derive a polycyclic presentation $\boldsymbol{\gamma}$ for G/H from $[\beta_1], \ldots, [\beta_k]$ by using [42, Algorithm 2.2], using T_H as described below.
4. Output $\boldsymbol{\gamma}$, with relative orders $r(\boldsymbol{\gamma})$ and relations $s(\boldsymbol{\gamma})$.

The polycyclic presentation $\boldsymbol{\gamma}$ output by Algorithm 3 is not necessarily minimal. It can be converted to a minimal presentation by simply removing those γ_i with $r(\gamma_i) = 1$, but for the purpose of computing the image in G/H of elements of G represented in terms of $\boldsymbol{\beta}$, it is better not to do so.

Algorithm 2.2 of [42] requires a TABLELOOKUP function that searches for a given group element in a table of distinct group elements. In Algorithm 3 above, the elements of G are uniquely represented by their $\boldsymbol{\beta}$-representations, but elements of G/H are represented as equivalence classes $[\delta]$, with $\delta \in G$, which is not a unique representation. To implement the TABLELOOKUP function for G/H, we do the following: Given $[\delta_0] \in G/H$ and a table $T_{G/H}$ of distinct elements $[\delta_i]$ in G/H, we test whether $\delta_0\delta_i^{-1} \in H$, for each $[\delta_i] \in T$. With a suitable implementation of T_H (such as a hash table or balanced tree), membership in H can be tested in $O(\log|G|)$ time, which is dominated by the $O(\log^2|G|)$ time to compute $\delta_0\delta_i^{-1}$.

Once Algorithm 3 completes, the problem of uniquely representing elements of G/H is solved: Every element of G/H has a unique $\boldsymbol{\gamma}$-representation.

Theorem 11. *Algorithm 3 runs in $O(n\log^2 n)$ time using $O((m + n/m)\log n)$ space, where $n = |G|$ and $m = |H|$.*

Proof. The time complexity is dominated by the n/m calls to the TABLELOOKUP function performed by [42, Algorithm 2.2] in step 3, each of which performs m operations in G (using $\boldsymbol{\beta}$-representations) and m lookups in T_H, yielding a total cots of $O(n\log^2 n)$. The space bound is the size of T_H plus the size of $T_{G/H}$. □

3.6. ***Other modular functions.*** For a modular function g of level N and a prime $\ell \nmid N$, the modular polynomial Φ_ℓ^g is the minimal polynomial of the function $g(\ell z)$ over the field $\mathbb{C}(g)$. For suitable functions g, the isogeny volcano algorithm for computing $\Phi_\ell(X, Y)$ can be adapted to compute $\Phi_\ell^g(X, Y)$, as described in [8, §7]. There are some restrictions: Φ_ℓ^g must have degree $\ell + 1$ in both X and Y, and we require some additional constraints on the suitable orders $\mathcal{O}$ that we use. Specifically, we require that there is a generator τ of $\mathcal{O}$ for which $g(\tau)$ lies in the ring class field $K_\mathcal{O}$. In this case we say that $g(\tau)$ is a *class invariant*, and we let $H_\mathcal{O}^g(X)$ denote its minimal polynomial over K; see [7; 14; 16] for algorithms to compute $H_\mathcal{O}^g(X)$. We also require the polynomial $H_\mathcal{O}^g$ to be defined over $\mathbb{Z}$.

With this setup, there is then a one-to-one correspondence between the roots $j(\tau)$ of $H_\mathcal{O}$ and the roots $g(\tau)$ of $H_\mathcal{O}^g$ in which $\Psi^g(g(\tau), j(\tau)) = 0$, where Ψ^g is the minimal polynomial of g over $\mathbb{C}(j)$; note that Ψ^g does not depend on ℓ and is assumed to be given. The class group $\mathrm{Cl}(\mathcal{O}) \simeq \mathrm{Gal}(K_\mathcal{O}/K)$ acts compatibly on both sets of roots, and this allows us to compute Φ_ℓ^g modulo suitable primes p using essentially the same algorithm that is used to compute $\Phi_\ell \bmod p$. In particular, we can enumerate the set $\mathrm{Ell}_\mathcal{O}^g(\mathbb{F}_p) = \{x \in \mathbb{F}_p : H_\mathcal{O}^g(x) = 0\}$ using a polycyclic

presentation $\boldsymbol{\alpha}$ for $\mathrm{Cl}(\mathcal{O})$, provided that we exclude from $\boldsymbol{\alpha}$ generators whose norm divides the level of g, and similarly for $\mathrm{Ell}^g_{\mathcal{O}'}(\mathbb{F}_p)$, where $\mathcal{O}' = \mathbb{Z} + \ell\mathcal{O}$.

Thus Algorithms 1 and 2 can both be adapted to compute instantiated modular polynomials $\phi^g(Y) = \Phi^g_\ell(x, Y) \bmod q$. Some effort may be required to determine the correspondence between $\mathrm{Ell}_{\mathcal{O}}(\mathbb{F}_p)$ and $\mathrm{Ell}^g_{\mathcal{O}}(\mathbb{F}_p)$ in cases where $\Psi^g(X, j(E))$ or $\Psi^g(g(E), Y)$ has multiple roots in $\mathbb{F}_p$; this issue arises when we need to compute a child or parent using Vélu's algorithm. There are several techniques for resolving such ambiguities; see [8, §7.3] and especially [16], which explores this issue in detail.

We emphasize that the point x at which we are evaluating $\Phi^g_\ell(x, Y)$ may be *any* element of $\mathbb{F}_q$; it need not correspond to the "g-invariant" of an elliptic curve.[4] This permits a very useful optimization that speeds up our original version of Algorithm 1 for computing $\phi_\ell(Y) = \phi^j_\ell(Y)$ by a factor of at least 9, as we now explain.

3.7. ***Accelerating the computation of $\boldsymbol{\phi_\ell(Y)}$ using $\boldsymbol{\gamma_2}$.*** Let $\gamma_2(z)$ be the unique cube root of $j(z)$ with integral Fourier expansion, a modular function of level 3 that yields class invariants for $\mathcal{O}$ whenever $3 \nmid \mathrm{disc}\,\mathcal{O}$. As noted in [8, §7.2], for $\ell > 3$ the modular polynomial $\Phi^{\gamma_2}_\ell$ can be written as

$$\Phi^{\gamma_2}_\ell(X, Y) = R(X^3, Y^3)Y^e + S(X^3, Y^3)XY + T(X^3, Y^3)X^2Y^{2-e}, \quad (5)$$

with $e = \ell + 1 \bmod 3$ and $R, S, T \in \mathbb{Z}[X, Y]$. We then have the identity

$$\Phi_\ell = R^3Y^e + (S^3 - 3RST)XY + TX^2Y^{2-e}. \quad (6)$$

When computing $\Phi^{\gamma_2}_\ell \bmod p$ with the isogeny volcano algorithm, one can exploit (5) to speed up the computation by at least a factor of 3. In addition, the integer coefficients of $\Phi^{\gamma_2}_\ell$ are also smaller than those of Φ_ℓ by roughly a factor of 3; we may use the height bound $h(\Phi^{\gamma_2}_\ell) \le 2\ell \log \ell + 8\ell$ from [8, Equation 18].

Let us consider how we may modify Algorithm 1 to exploit (6), thereby accelerating the computation of $\phi_\ell(Y) = \Phi_\ell(x, Y) \bmod q$, where $x = j(E) \in \mathbb{F}_q$. Let $r(Y) = R(x, Y) \bmod q$, and similarly define s and t in terms of S and T. Rather than computing $\Phi_\ell \bmod p$ in step 5(a), we compute $\Phi^{\gamma_2}_\ell \bmod p$ and derive R, S, and T from (5). We then compute polynomials $\bar{r}$, $\bar{s}$, and $\bar{t} \bmod p$ instead of $\bar{\phi}$ in step 5(b). Finally, we recover r, s, and $t \bmod q$ in step 6 via the explicit CRT and output

$$\phi = r^3Y^e + x(s^3 - 3rst)Y + x^2t^3Y^{2-e} \quad (7)$$

[4] Every $x \in \mathbb{F}_q$ is $j(E)$ for some $E/\mathbb{F}_q$, and when E is ordinary, $j(E)$ is the reduction of some $j(\tau) = j(\widehat{E})$ with $\mathbb{Z}[\tau] = \mathcal{O} \simeq \mathrm{End}(E)$. But $g(\tau)$ might not be a class invariant for this $\mathcal{O}$.

in step 7. Adjusting the height bound B appropriately, this yields a speedup of nearly a factor of 9. Note that we are not assuming $x = j(E)$ has a cube root in $\mathbb{F}_q$, or that $\mathrm{End}(E) \simeq \mathbb{O}$ satisfies $3 \nmid \mathrm{disc}\,\mathbb{O}$; the identity (7) holds for all x.

We can similarly compute ϕ_X and ϕ_{XX}. To simplify the formulas, let us define $U = (S^3 - 3RST)$ and $u = U(x, Y) \bmod q$. Define

$$r'(Y) = \frac{\partial R}{\partial X}(x, Y) \quad \text{and} \quad r''(Y) = \frac{\partial^2 R}{\partial X^2}(x, Y),$$

and similarly for s, t, and u. Note that u, u', and u'' can all be easily expressed in terms of $r, r', r'', s, s', s'', t, t'$, and t''. We replace the computation of $\bar{\phi}_X$ and $\bar{\phi}_{XX}$ in step 5(c) with analogous computations of $\bar{r}', \bar{r}'', \bar{s}', \bar{s}'', \bar{t}'$, and $\bar{t}''$ mod p. We then obtain $r, r', r'', s, s', s'', t, t'$, and t'' via the explicit CRT mod q and apply

$$\begin{aligned}\phi_X &= 3r^2r'Y^e + (u + xu')Y + (2xt^3 + 3x^2t^2t')Y^{2-e},\\ \phi_{XX} &= (6rr'r' + 3r^2r'')Y^e\\ &\qquad + (2u' + u'')Y + (2t^3 + 12xt^2t' + 6x^2tt't' + 3x^2t^2t'')Y^{2-e}.\end{aligned}$$

3.8. ***Normalized isogenies.*** We now explain how Algorithms 1 and 2 may be used to compute normalized isogenies ψ, first using j-invariants, and then using g-invariants. Throughout this section $j = j(E) \in \mathbb{F}_q$ denotes the j-invariant of a given elliptic curve $E/\mathbb{F}_q$, defined by $y^2 = x^3 + Ax + B$, and $\phi(Y) = \Phi_\ell(j, Y)$. We use $\widetilde{j} = j(\widetilde{E})$ to denote a root of $\phi(Y)$ in $\mathbb{F}_q$. Our goal is to compute an equation for the image of $\psi : E \to \widetilde{E}$, and the kernel polynomial $h_\ell(X)$ for ψ.

3.8.1. *Algorithm 1.* When computing ϕ, we also compute the optional outputs ϕ_X and ϕ_{XX}, and then

$$\phi_Y(Y) = \frac{d}{dY}\phi(Y), \quad \phi_{YY}(Y) = \frac{d}{dY}\phi_Y(Y), \quad \text{and} \quad \phi_{XY} = \frac{d}{dY}\phi_X(Y).$$

We then compute the quantities $\Phi_*(j, \widetilde{j}) = \phi_*(\widetilde{j})$, for $* = X, Y, XX, XY, YY$, as defined in Section 2.1, and apply Elkies's algorithm [19, Algorithm 27] to compute $\widetilde{E}$ and $h_\ell(X)$.

3.8.2. *Algorithm 2.* Having computed ϕ and obtained $\widetilde{j}$, we run Algorithm 2 *again*, this time with the input $\widetilde{j}$, obtaining $\widetilde{\phi}(Y) = \Phi_\ell(\widetilde{j}, Y)$, which we now regard as $\widetilde{\phi}(X) = \Phi_\ell(X, \widetilde{j})$, by the symmetry of Φ_ℓ. We then compute

$$\Phi_X(j, \widetilde{j}) = \left(\frac{d}{dX}\widetilde{\phi}\right)(j) \quad \text{and} \quad \Phi_Y(j, \widetilde{j}) = \left(\frac{d}{dY}\phi\right)(\widetilde{j}),$$

as well as the quantities

$$j' = \frac{18B}{A}j, \quad \widetilde{j}' = \frac{-\Phi_X(j, \widetilde{j})}{\ell\Phi_Y(j, \widetilde{j})}j', \quad \widetilde{m} = \frac{\widetilde{j}'}{\widetilde{j}}, \quad \text{and} \quad \widetilde{k} = \frac{\widetilde{j}'}{1728 - \widetilde{j}},$$

as in [19, Algorithm 27]. The normalized equation for $\widetilde{E}$ is then

$$y^2 = x^3 + \frac{\ell^4 \widetilde{m}\widetilde{k}}{48}x + \frac{\ell^6 \widetilde{m}^2 \widetilde{k}}{864},$$

and the fastElkies′ algorithm in [5] may be used to compute $h_\ell(X)$.

3.8.3. *Handling g-invariants.* We assume that $g(E)$ is known to be a class invariant (see Section 3.9 below). Let $g = g(E)$, $\phi^g(Y) = \Phi_\ell^g(g, Y)$, and let $\widetilde{g} = g(\widetilde{E})$ denote a root of $\phi^g(Y)$ in $\mathbb{F}_q$. In the case of Algorithm 1 we compute

$$\Phi_X^g(g, \widetilde{g}) = \phi_X^g(\widetilde{g}) \quad \text{and} \quad \Phi_Y^g(g, \widetilde{g}) = \left(\frac{d}{dY}\phi^g\right)(\widetilde{g}),$$

and in the case of Algorithm 2 we make a second call with input $\widetilde{g}$ to obtain $\widetilde{\phi}^g(X) = \Phi_\ell^g(X, \widetilde{g})$ as above. We then compute

$$\Phi_X^g(g, \widetilde{g}) = \left(\frac{d}{dX}\widetilde{\phi}^g\right)(g) \quad \text{and} \quad \Phi_Y^g(g, \widetilde{g}) = \left(\frac{d}{dY}\phi^g\right)(\widetilde{g}).$$

We assume the modular equation $\Psi_\ell^g(G, J) = 0$ relating $g(z)$ to $j(z)$ can be solved for $j(z)$ (for the $g(z)$ considered in [8], we have $\deg_J \Psi^g(G, J) \le 2$), and let $F(G)$ satisfy $\Psi_\ell^g(F(J), J) = 0$ and $F' = dF/dG$.

To compute the normalized equation for $\widetilde{E}$, we proceed as in Section 3.8.2, except now

$$\widetilde{j}' = \frac{-\Phi_X^g(g, \widetilde{g})F'(\widetilde{g})}{\ell\Phi_Y^g(g, \widetilde{g})/F'(g)}j'.$$

The fastElkies′ algorithm in [5] may then be used to compute h_ℓ, or, in the case of Algorithm 1, one may derive the trace of h_ℓ using $\Phi_{XX}^g(g, \widetilde{g})$, $\Phi_{XY}^g(g, \widetilde{g})$, and $\Phi_{YY}^g(g, \widetilde{g})$ as in Section 3.8.1, and compute h_ℓ as usual. We omit the details.

3.9. ***Verifying that g(E) is a class invariant.*** Let $E/\mathbb{F}_q$ be an elliptic curve that is not supersingular (see [44] for fast tests), with $\mathrm{End}(E) \simeq \mathcal{O}$. As in Section 3.6, we call an element $g(E)$ of $\mathbb{F}_q$ a *class invariant* if

(1) $H_{\mathcal{O}}^g(X)$ splits into linear factors in the ring class field of O, and

(2) $g(E)$ is a common root of $H_{\mathcal{O}}^g(X)$ and $\Psi^g(X, j(E))$.

For practical applications, we would like to determine whether $g(E)$ is a class invariant without computing $\mathcal{O}$ (indeed, the application may be to compute $\mathcal{O}$). This is often easy to do, at least as far as condition (1) is concerned. As noted in Section 3.6, condition (1) can typically be guaranteed by constraints involving $D = \mathrm{disc}\,\mathcal{O}$ and the level N of g. Verifying condition (2) is more difficult, in general, but it can be easily addressed in particular cases if we know that $\Psi^g(X, j(E))$

either has a unique root in $\mathbb{F}_q$ (which then must also be a root of $H^g(\mathbb{O})$ once condition (1) is satisfied), or that all its roots in $\mathbb{F}_q$ are roots of $H^g(\mathbb{O})$, or of $H^{\bar{g}}(\mathbb{O})$ for some $\bar{g}$ with $\Phi_\ell^{\bar{g}} = \Phi_\ell^g$. In the latter case we may not determine $g(E)$ uniquely, but for the purposes of computing a normalized ℓ-isogeny this does not matter, any choice will work.

Taking $\gamma_2 = \sqrt[3]{j}$ as an example, condition (1) holds when $\left(\frac{D}{3}\right) \neq 0$, which means $j(E)$ is on the surface of its 3-volcano and has either 0 or 2 siblings. This can be easily determined using [18] or [42, §4.1]. If we have $q \equiv 2 \bmod 3$, the polynomial $\Psi^g(X, j(E)) = X^3 - j(E)$ has a unique root $g(E)$ in $\mathbb{F}_q$ and condition (2) also holds. (There are techniques to handle $q \equiv 1 \bmod 3$ — see [7], for example — but they assume that $\mathbb{O}$ is known.)

As a second example, consider the Weber ƒ-function, which is related to the j-function by $\Psi^{\mathfrak{f}}(X, J) = (X^{24} - 16)^3 - X^{24}J$. Now we require $\left(\frac{D}{3}\right) \neq 0$ and $\left(\frac{D}{2}\right) = 1$. The latter is equivalent to $j(E)$ being on the surface of its 2-volcano with 2 siblings. If we also have $q \equiv 11 \bmod 12$, then $\Psi^{\mathfrak{f}}(X, j(E))$ has exactly two roots $\mathfrak{f}(E)$ and $-\mathfrak{f}(E)$, by [8, Lemma 7.3], and either may be used since $\Phi_\ell^{\mathfrak{f}} = \Phi_\ell^{-\mathfrak{f}}$.

For a more general solution, having verified condition (1), we may simply compute instantiated polynomials $\phi(Y) = \Phi_\ell(x, Y)$ for *every* root x of $\Psi^g(X, j(E))$ in $\mathbb{F}_q$. This can be done at essentially no additional cost, and we may then attempt to compute a normalized isogeny corresponding to each root x, which we validate by computing the dual isogeny (using the normalization factor $c = \ell$ rather than 1) and checking whether the composition corresponds to scalar multiplication by ℓ using randomly generated points in $E(\mathbb{F}_q)$. The cost of these validations is negligible compared to the cost of computing $\phi(Y)$ for even one x.

As a final remark, we note that in applications such as point counting where one is only concerned with the isogeny class of E, in cases where condition (1) is not satisfied, one may be able to obtain an isogenous $\widetilde{E}$ for which condition (1) holds by simply climbing to the surface of the relevant ℓ_0-volcanoes for the primes $\ell_0 \mid N$ (we regard N as fixed so ℓ_0 is small; $\ell_0 = 2, 3$ in the examples above).

4. Applications

In this section we analyze the use of Algorithms 1 and 2 in two particular applications: counting points and computing endomorphism rings.

Recall that for an elliptic curve $E/\mathbb{F}_q$, an odd prime ℓ is called an *Elkies prime* whenever $\phi(Y) = \Phi_\ell(j(E), Y)$ has a root in $\mathbb{F}_q$. This holds if and only if $t^2 - 4q$ is a square mod ℓ, where $t = q + 1 - \#E(\mathbb{F}_q)$. It follows from the Chebotarev density theorem that the set of Elkies primes for E has density $1/2$. The complexity of the Schoof-Elkies-Atkin algorithm [36] for computing $\#E(\mathbb{F}_q)$ depends critically

on the number of *small* Elkies primes, specifically, the least $L = L(E)$ for which

$$\sum_{\text{Elkies primes } \ell \le L(E)} \log \ell > \log(4\sqrt{q}). \tag{8}$$

On average, one expects $L \approx \log q$, but even under the GRH the best proven bound is $L = O(\log^{2+\epsilon} q)$; see Appendix A of [34] by Satoh and Galbraith. This yields a complexity bound that is actually slightly *worse* than Schoof's original algorithm.

For practical purposes, the heuristic assumption $L(E) = O(\log q)$ is often used when analyzing the complexity of the SEA algorithm. This assumption holds for almost all elliptic curves [38], but it is known to fail in infinitely many cases [37]. We instead adopt the following weaker heuristic.

Heuristic 12. There exists a constant c such that for all sufficiently large q we have $L(E) \le c \log q \operatorname{llog} q$ for every elliptic curve $E/\mathbb{F}_q$.

Theorem 13. *Assume the GRH and Heuristic 12. Let $E/\mathbb{F}_q$ be an elliptic curve over a prime field $\mathbb{F}_q$ and let $n = \log q$. There is a Las Vegas algorithm to compute $\#E(\mathbb{F}_q)$ that runs in $O(n^4 \log^3 n \operatorname{llog} n)$ expected time using $O(n^2 \log n)$ space.*

Proof. Apply the SEA algorithm, using Algorithm 1 to compute $\phi(Y) = \Phi_\ell(j(E), Y)$ (and also ϕ_X and ϕ_{XX}), and ignore the Atkin primes, as in [38, Algorithm 1], for example. There are $O(n/\log n)$ primes in the sum (8), and under Heuristic 12, they are bounded by $L = O(n \log n)$. It follows from [38, Table 1] that the expected time to process each Elkies prime given ϕ is $O(n^3 \log^3 n \operatorname{llog}^2 n)$, which is dominated by the time to compute ϕ, as is the space. The theorem then follows from Theorem 4. □

A common application of the SEA algorithm is to search for random curves of prime (or near prime) order, for use in cryptographic applications. As shown in [38], we no longer need Heuristic 12 to do this; we can assume $L(E) = O(\log q)$ for a randomly chosen elliptic curve. Additionally, since we expect to count points on many curves ($\approx \log q$), we can take advantage of *batching*, whereby we extend Algorithm 1 to take multiple inputs $j(E_1) \in \mathbb{F}_{q_1}, \ldots, j(E_k) \in \mathbb{F}_{q_k}$ and produce corresponding outputs for each (the $\mathbb{F}_{q_i}$ may coincide, but they need not). Provided $k = O(\log \ell)$, this does not change the time complexity (relative to the largest $\mathbb{F}_{q_i}$), since the most time-consuming steps depend only on ℓ, not $j(E)$, and the space complexity is increased by at most a factor of k.[5]

Let $E_{a,b}$ denote the elliptic curve defined by $y^2 = x^3 + ax + b$, and for any real number $x > 3$, let $T(x)$ denote the set of all triples (q, a, b) with $q \in [x, 2x]$ prime, $a, b \in \mathbb{F}_q$, and $\#E_{a,b}$ prime. The following result strengthens [38, Theorem 3].

[5] These remarks also apply to Algorithm 2.

Theorem 14. *There is a Las Vegas algorithm that, given x, outputs a random triple $(q,a,b) \in T(x)$ and the prime $\#E_{a,b}(\mathbb{F}_q)$, with q uniformly distributed over the primes in $[x,2x]$ and (a,b) uniformly distributed over the pairs $(c,d) \in \mathbb{F}_q^2$ for which $\#E_{c,d}(\mathbb{F}_q)$ is prime. Under the GRH, its expected running time is $O(n^5 \log^2 n \operatorname{llog} n)$ using $O(n^2 \log^2 n)$ space, where $n = \log x$.*

Proof. We modify the algorithm in [38] to use Algorithm 1, operating on batches of $O(\log n)$ inputs at a time. One then obtains an $O(n^4 \log n \operatorname{llog} n)$ bound on the average time to compute $\#E_{a,b}(\mathbb{F}_q)$ for primes $q \in [x,2x]$, and a space complexity of $O(n^2 \log^2 n)$. The theorem then follows from the proof of [38, Theorem 3]. □

A second application of Algorithms 1 and 2 is in the computation of the endomorphism ring of an ordinary elliptic curve. The algorithm in [3] achieves a heuristically subexponential running time of $L[1/2, \sqrt{3}/2]$ using $L[1/2, 1/\sqrt{3}]$ space. Algorithms 1 and 2 both improve the space complexity bound to $L[1/2, 1/\sqrt{12}]$, which is significant, since space is the limiting factor in these computations. Algorithm 2 also provides a slight improvement to the time complexity that is not visible in the $L[\alpha, c]$ notation but may be practically useful. These remarks also apply to the algorithm in [27] for evaluating isogenies of large degree.

5. Computations

Using a modified version of the SEA algorithm incorporating Algorithm 1, we determined the number of points on the elliptic curve

$$y^2 = x^3 + 2718281828x + 3141592653,$$

modulo the 5011-digit prime $q = 16219299585 \cdot 2^{16612} - 1$. The algorithm ignored the Atkin primes and computed the trace of Frobenius t modulo 700 Elkies primes, the largest of which was $\ell = 11681$; see [41] for details, including the exact value of t, which is too large to print here. The computation was distributed over 32 cores (3.0 GHz AMD Phenom II), and took about 6 weeks. Table 1 gives the time taken for various parts of the computation.

For $\ell = 11681$, the size of $\phi_\ell^{\mathfrak{f}}(Y) = \Phi_\ell^{\mathfrak{f}}(\mathfrak{f}(E), Y)$ was under 20MB and took about two hours to compute on a single core. As can be seen in Table 1, the computation of $\phi_\ell^{\mathfrak{f}}$ accounted for less than 3% of the total running time, despite being the asymptotically dominant step. This is primarily due to the use of the Weber $\mathfrak{f}$-invariant; with a less advantageous invariant (in the worst case, the j-invariant with the optimization of Section 3.7), the time spent computing ϕ_ℓ would have been comparable to or greater than the time spent on the remaining steps. But in any case the computation would still have been quite feasible.

To demonstrate the scalability of the algorithm, we computed $\phi_\ell^{\mathfrak{f}}(Y)$ for an elliptic curve $E/\mathbb{F}_q$, with $\ell = 100019$ and $q = 2^{86243} - 1$. Running on 32 cores

Task	CPU days
Computing $\phi_\ell^{\mathfrak{f}}(Y)$ with Algorithm 1	32
Computing $X^q \bmod \phi_\ell$ (using [24])	995
Computing h_ℓ using [19, Algorithm 27]	3
Computing Y^q and $X^q \bmod h_\ell, E$ using [22]	326
Computing the eigenvalue λ_ℓ using BSGS	22
Total	1378

Table 1. Breakdown of time spent computing $\#E(\mathbb{F}_q)$ for a 5011-bit prime q. The computation was performed on 32 cores of a 3.0 GHz AMD Phenom II.

(Algorithms 1 and 2 are both easily parallelized), this computation took less than a week. We note that the size of the instantiated modular polynomial $\phi_\ell^{\mathfrak{f}}$ (and ϕ_ℓ) is almost exactly one gigabyte, whereas the size of $\Phi_\ell^{\mathfrak{f}}$ is many terabytes, and we estimate that the size of Φ_ℓ is around 20 or 30 petabytes.

Acknowledgments

I am grateful to David Harvey for his assistance with the algorithms for fast polynomial arithmetic used in the computations described in Section 5, and to Daniel Kane for suggesting the hybrid approach outlined in Section 3.4. I also thank the referees for their comments and helpful suggestions.

This work was partially supported by NSF grant DMS-1115455.

References

[1] Juliana Belding, Reinier Bröker, Andreas Enge, and Kristin Lauter, *Computing Hilbert class polynomials*, in van der Poorten and Stein [33], 2008, pp. 282–295. MR 2009j:11200

[2] Daniel J. Bernstein and Jonathan P. Sorenson, *Modular exponentiation via the explicit Chinese remainder theorem*, Math. Comp. **76** (2007), no. 257, 443–454. MR 2007f:11142

[3] Gaetan Bisson and Andrew V. Sutherland, *Computing the endomorphism ring of an ordinary elliptic curve over a finite field*, J. Number Theory **131** (2011), no. 5, 815–831. MR 2012a:11080

[4] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system*, *I*: *The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[5] A. Bostan, F. Morain, B. Salvy, and É. Schost, *Fast algorithms for computing isogenies between elliptic curves*, Math. Comp. **77** (2008), no. 263, 1755–1778. MR 2009k:11207

[6] Reinier Bröker, *A p-adic algorithm to compute the Hilbert class polynomial*, Math. Comp. **77** (2008), no. 264, 2417–2435. MR 2009j:11093

[7] ———, *p-adic class invariants*, LMS J. Comput. Math. **14** (2011), 108–126. MR 2801172

[8] Reinier Bröker, Kristin Lauter, and Andrew V. Sutherland, *Modular polynomials via isogeny volcanoes*, Math. Comp. **81** (2012), no. 278, 1201–1231. MR 2012m:11180

[9] Reinier Bröker and Andrew V. Sutherland, *An explicit height bound for the classical modular polynomial*, Ramanujan J. **22** (2010), no. 3, 293–313. MR 2011g:11123

[10] D. A. Buell and J. T. Teitelbaum (eds.), *Computational perspectives on number theory*: *Proceedings of the conference in honor of A. O. L. Atkin held at the University of Illinois*, *Chicago*, *IL*, *September* 1995, AMS/IP Studies in Advanced Mathematics, no. 7, Providence, RI, American Mathematical Society, 1998. MR 98g:11001

[11] Andrew M. Childs, David Jao, and Vladimir Soukharev, *Constructing elliptic curve isogenies in quantum subexponential time*, 2012. arXiv 1012.4019v2 [quant-ph]

[12] Jean-Guillaume Dumas (ed.), *ISSAC* 2006: *Proceedings of the* 2006 *International Symbolic and Algebraic Computation held in Genova*, *July* 9–12, 2006, New York, ACM Press, 2006. MR 2289094

[13] Noam D. Elkies, *Elliptic and modular curves over finite fields and related computational issues*, in Buell and Teitelbaum [10], 1998, pp. 21–76. MR 99a:11078

[14] Andreas Enge, *The complexity of class polynomial computation via floating point approximations*, Math. Comp. **78** (2009), no. 266, 1089–1107. MR 2010h:11097

[15] ———, *Computing modular polynomials in quasi-linear time*, Math. Comp. **78** (2009), no. 267, 1809–1824. MR 2010b:11171

[16] Andreas Enge and Andrew V. Sutherland, *Class invariants by the CRT method*, in Hanrot et al. [23], 2010, pp. 142–156. MR 2012d:11246

[17] Claus Fieker and David R. Kohel (eds.), *Algorithmic number theory*: *Proceedings of the* 5*th International Symposium* (*ANTS-V*) *held at the University of Sydney*, *July* 7–12, 2002, Lecture Notes in Computer Science, no. 2369, Berlin, Springer, 2002. MR 2004j:11002

[18] Mireille Fouquet and François Morain, *Isogeny volcanoes and the SEA algorithm*, in Fieker and Kohel [17], 2002, pp. 276–291. MR 2005c:11077

[19] Steven D. Galbraith, *Mathematics of public key cryptography*, Cambridge University Press, 2012. MR 2931758

[20] Steven D. Galbraith, Florian Hess, and Nigel P. Smart, *Extending the GHS Weil descent attack*, in Knudsen [29], 2002, pp. 29–44. MR 2004f:94060

[21] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, 2nd ed., Cambridge University Press, 2003. MR 2004g:68202

[22] P. Gaudry and F. Morain, *Fast algorithm for computing the eigenvalue in the Schoof-Elkies-Atkin algorithm*, in Dumas [12], 2006, pp. 109–115. MR 2289108

[23] Guillaume Hanrot, François Morain, and Emmanuel Thomé (eds.), *Algorithmic number theory*: *Proceedings of the* 9*th Biennial International Symposium* (*ANTS-IX*) *held in Nancy*, *July* 19–23, 2010, Lecture Notes in Computer Science, no. 6197, Berlin, Springer, 2010. MR 2011g:11002

[24] David Harvey, *A cache-friendly truncated FFT*, Theoret. Comput. Sci. **410** (2009), no. 27–29, 2649–2658. MR 2010g:68327

[25] IEEE (ed.), *Proceedings of the* 49*th IEEE Symposium on Foundations of Computer Science held in Philadelphia*, *October* 25–28, 2008, Los Alamitos, CA, Institute of Electrical and Electronics Engineers, IEEE Computer Society, 2008.

[26] INRIA Project-Team TANC, 2009 *activity report*, 2009. http://raweb.inria.fr/rapportsactivite/RA2009/tanc/tanc.pdf

[27] David Jao and Vladimir Soukharev, *A subexponential algorithm for evaluating large degree isogenies*, in Hanrot et al. [23], 2010, pp. 219–233. MR 2011h:11144

[28] Kiran S. Kedlaya and Christopher Umans, *Fast Modular Composition in any Characteristic*, in IEEE [25], 2008, pp. 146–155.

[29] Lars Knudsen (ed.), *Advances in cryptology—EUROCRYPT* 2002: *Proceedings of the* 21*st International Annual Conference on the Theory and Applications of Cryptographic Techniques held in Amsterdam*, *April* 28–*May* 2, 2002, Lecture Notes in Computer Science, no. 2332, Berlin, Springer, 2002. MR 2003m:94074

[30] David Russell Kohel, *Endomorphism rings of elliptic curves over finite fields*, Ph.D. thesis, University of California, Berkeley, 1996, p. 117. http://search.proquest.com/docview/304241260 MR 2695524

[31] Serge Lang, *Elliptic functions*, 2nd ed., Graduate Texts in Mathematics, no. 112, Springer, New York, 1987. MR 88c:11028

[32] The PARI Group, *PARI/GP* (*version* 2.4.3), 2011. http://pari.math.u-bordeaux.fr/

[33] Alfred J. van der Poorten and Andreas Stein (eds.), *Algorithmic number theory*: *Proceedings of the* 8*th International Symposium* (*ANTS-VIII*) *held in Banff*, *AB*, *May* 17–22, 2008, Lecture Notes in Computer Science, no. 5011, Berlin, Springer, 2008. MR 2009h:11002

[34] Takakazu Satoh, *On p-adic point counting algorithms for elliptic curves over finite fields*, in Fieker and Kohel [17], 2002, pp. 43–66. MR 2004k:11098

[35] A. Schönhage and V. Strassen, *Schnelle Multiplikation großer Zahlen*, Computing (Arch. Elektron. Rechnen) **7** (1971), 281–292. MR 45 #1431

[36] René Schoof, *Counting points on elliptic curves over finite fields*, J. Théor. Nombres Bordeaux **7** (1995), no. 1, 219–254. MR 97i:11070

[37] Igor Shparlinski, *On the product of small Elkies primes*, 2013. arXiv 1301.0035 [math.NT]

[38] Igor E. Shparlinski and Andrew V. Sutherland, *On the distribution of Atkin and Elkies primes*, 2011. arXiv 1112.3390 [math.NT]

[39] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 106, Springer, New York, 1986. MR 87g:11070

[40] ———, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 151, Springer, New York, 1999. MR 96b:11074

[41] Andrew V. Sutherland, *Genus* 1 *point counting records over prime fields*, 2010. http://math.mit.edu/∼drew/SEArecords.html

[42] ———, *Computing Hilbert class polynomials with the Chinese remainder theorem*, Math. Comp. **80** (2011), no. 273, 501–538. MR 2011k:11177

[43] ———, *Accelerating the CM method*, LMS J. Comput. Math. **15** (2012), 172–204. MR 2970725

[44] ———, *Identifying supersingular elliptic curves*, LMS J. Comput. Math. **15** (2012), 317–325. MR 2988819

[45] Jacques Vélu, *Isogénies entre courbes elliptiques*, C. R. Acad. Sci. Paris Sér. A-B **273** (1971), A238–A241. http://gallica.bnf.fr/ark:/12148/bpt6k56191248/f52.image MR 45 #3414

ANDREW V. SUTHERLAND: drew@math.mit.edu
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

THE OPEN BOOK SERIES **1** (2013)
Tenth Algorithmic Number Theory Symposium
dx.doi.org/10.2140/obs.2013.1.557

msp

Constructing and tabulating dihedral function fields

Colin Weir, Renate Scheidler, and Everett W. Howe

We present algorithms for constructing and tabulating degree-ℓ dihedral extensions of $\mathbb{F}_q(x)$, where $q \equiv 1 \bmod 2\ell$. We begin with a Kummer-theoretic algorithm for constructing these function fields with prescribed ramification and fixed quadratic resolvent field. This algorithm is based on the proof of our main theorem, which gives an exact count for such fields. We then use this construction method in a tabulation algorithm to construct all degree-ℓ dihedral extensions of $\mathbb{F}_q(x)$ up to a given discriminant bound, and we present tabulation data. We also give a formula for the number of degree-ℓ dihedral extensions of $\mathbb{F}_q(x)$ with discriminant divisor of degree $2(\ell-1)$, the minimum possible.

1. Introduction

Two important problems in algebraic and algorithmic number theory are the construction of global fields of a fixed discriminant or prescribed ramification — with its curve analogue of constructing Galois covers of fixed genus — and the tabulation of global fields with a certain Galois group up to some discriminant or genus bound. The latter problem goes hand in hand with asymptotic estimates for the number of such fields; for example, estimates for cubic number fields were first given in [11] and for quartics in [2]. There is a sizable body of literature on construction, tabulation, and asymptotic counts of number fields; a comprehensive survey of known results can be found in [6], and extensive tables of data are available at [19].

Far less is known in the function field setting; only the asymptotic counts for cubic [10] and abelian [39] extensions have been proved. However, there is a general program described by Ellenberg and Venkatesh [37] for formulating these asymptotic estimates for both number fields and function fields. In particular, they point out the "alarming gap between theory and experiment" in asymptotic predictions for number fields. In the case of cubic number fields, this inconsistency led

MSC2010: primary 11R58; secondary 11Y40.
Keywords: function field, Galois group, dihedral extension, construction, tabulation.

Roberts [23] to conjecture the secondary term in the theorem of Davenport and Heilbronn in [11]. His conjecture was later proved independently by Bhargava, Shankar, and Tsimerman [3] and by Taniguchi and Thorne [35]. In the function field setting, however, there is practically no experimental data to potentially identify a similar such gap. The only known algorithm for constructing all cubic function fields with a given squarefree discriminant is that of [18], although recently Pohst [22] showed how to construct all non-Galois cubic extensions of $\mathbb{F}_q(x)$ with a given discriminant, which also leads to such an algorithm. Tabulation methods for certain classes of cubic function fields can be found in [26] and [25].

This paper represents a next step toward function field tabulation. We present a method for constructing all degree-ℓ extensions of $\mathbb{F}_q(x)$ with prescribed ramification and with Galois group isomorphic to the dihedral group of order 2ℓ, in the case where $q \equiv 1 \bmod 2\ell$. We use a Kummer-theoretic approach inspired by the methods of Cohen [7; 8] for number fields. This construction method can be converted into a tabulation algorithm in the usual manner via iteration. However, we are able to use the automorphism group $\mathrm{PGL}(2,q)$ of $\mathbb{F}_q(x)$ to effect significant improvements. Note that this technique is unique to the function field setting, as there are no nontrivial automorphisms of the rational numbers. Exploiting $\mathbb{F}_q(x)$-automorphisms reduces the number of constructions by a factor of order q^3 compared to the naïve approach. We present our improved tabulation procedure along with numerical data obtained from an implementation in Magma [5]. It is important to note that in the special case $\ell = 3$, our algorithm generates complete tables of non-Galois cubic function fields over $\mathbb{F}_q(x)$ up to a given discriminant bound.

2. Preliminaries

Let ℓ be an odd prime and let $\mathbb{F}_q$ be a finite field of characteristic coprime to 2ℓ. We denote by K the rational function field over $\mathbb{F}_q$ and by K^{sep} a separable closure of K. In this paper, a *function field* will always mean a subfield L of K^{sep} that contains K as a subfield of finite index, and by the *Galois group* of L we mean the Galois group of its Galois closure over K.

Suppose F/E is a finite extension of functions fields. Let $\mathrm{Places}(F)$ denote the set of places of F, and let $e(P'|P)$ and $f(P'|P)$ denote the ramification index and relative degree of a place $P' \in \mathrm{Places}(F)$ lying over $P \in \mathrm{Places}(E)$, respectively. The *norm* of a place $P' \in \mathrm{Places}(F)$ is the divisor

$$\mathrm{N}_{F/E}(P') := f(P' \mid P)\, P,$$

and the *conorm* of $P \in \mathrm{Places}(E)$ is

$$\mathrm{Con}_{F/E}(P) := \sum_{P' \mid P} e(P' \mid P)\, P'.$$

Then $\mathrm{N}_{F/E}(\mathrm{Con}_{F/E}(P)) = [F : E]\,P$. These definitions extend additively to divisors. We will also use $\mathrm{N}_{F/E}$ to denote the norm map on elements of F. Proposition 7.8 in [24] shows that this is reasonable: The norm of a principal divisor (α) of F is the principal divisor $(\mathrm{N}_{F/E}(\alpha))$ of E. Restricting to the cases where the characteristic is different from 2 and ℓ guarantees that for the field extensions we will consider, there are no wildly ramified places. Thus, for the extensions F/E we will work with, the *different* is given by

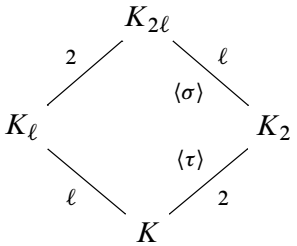
$$\mathrm{Diff}_{F/E} := \sum_{P \in \mathrm{Places}(E)} \sum_{P' \mid P} (e(P' \mid P) - 1)\, P'.$$

The *discriminant divisor* of F/E is defined as

$$\Delta_{F/E} := \mathrm{N}_{F/E}(\mathrm{Diff}_{F/E}) = \sum_{P \in \mathrm{Places}(E)} \sum_{P' \mid P} (e(P' \mid P) - 1) f(P' \mid P)\, P.$$

When $E = K$, we drop E from the notation and simply write Δ_F. Note that $\deg \Delta_{F/E} = \deg \mathrm{Diff}_{F/E}$, so one can replace $\mathrm{Diff}_{F/E}$ by $\Delta_{F/E}$ in the Hurwitz genus formula ([32, Theorem 3.4.13]). For these reasons, we will henceforth describe the ramification of a function field in terms of its discriminant divisor.

Let K_ℓ be a degree-ℓ function field with Galois group $\mathcal{D}_\ell$, the dihedral group with 2ℓ elements, and construct the dihedral extension $K_{2\ell}$ as the Galois closure of K_ℓ over K:



(1)

Here K_2 is the fixed field of the unique index-2 subgroup $\mathcal{C}_\ell$ of $\mathcal{D}_\ell$ and K_ℓ is the fixed field of an element of order 2 in $\mathcal{D}_\ell$. We note that there are ℓ such elements in $\mathcal{D}_\ell$, which give ℓ subfields of $K_{2\ell}$ conjugate to K_ℓ. The field K_2 is called the *quadratic resolvent field* of K_ℓ; we write $K_2 = \mathrm{QuadRes}\, K_\ell$. We let τ denote a generator of $\mathrm{Gal}(K_2/K)$ and σ a generator of $\mathrm{Gal}(K_{2\ell}/K_2)$.

3. Description of all degree-ℓ dihedral function fields

Our first goal is to count the number of ℓ-tuples of conjugate dihedral degree-ℓ function fields with a given discriminant divisor and quadratic resolvent field. There is a one-to-one correspondence between nonconjugate dihedral degree-ℓ function

fields K_ℓ and their Galois closures $K_{2\ell}$. Consequently, instead of counting degree-ℓ dihedral extensions, we count the number of dihedral Galois fields $K_{2\ell}$. We do so via construction: Given a quadratic function field K_2 and discriminant divisor Δ, we construct all degree-ℓ cyclic extensions $K_{2\ell}$ of K_2 such that $\mathrm{Gal}(K_{2\ell}/K) = \mathcal{D}_\ell$ and all conjugate index-2 subfields K_ℓ of $K_{2\ell}$ have discriminant divisor $\Delta_{K_\ell} = \Delta$.

Since $q \equiv 1 \bmod \ell$, all cyclic ℓ-extensions of K_2 are Kummer extensions — that is, extensions of the form $K_2(\sqrt[\ell]{\alpha})$ for some $\alpha \in K_2^\times \setminus (K_2^\times)^\ell$. In Section 3A we give necessary and sufficient conditions on α for $K_2(\sqrt[\ell]{\alpha})$ to be Galois over K with group $\mathcal{D}_\ell$. In Section 3B, we use virtual units to decompose $K_2^\times/(K_2^\times)^\ell$ in a way that allows us to determine the elements α that correspond to nonisomorphic dihedral function fields. With this information, in Section 3C we compute the discriminant divisor of $K_\ell \subset K_2(\sqrt[\ell]{\alpha})$ in terms of (α) and Δ_{K_2}. Next, in Section 3D we give a constructive proof of the main theorem: an exact count of the number of nonconjugate dihedral degree-ℓ extensions of K with a given quadratic resolvent field K_2 and discriminant divisor. We close in Section 3E by showing how to give explicit equations for the function fields we construct.

3A. ***Kummer theory.*** Let ℓ be a prime and let F be a field that contains the ℓ-th roots of unity. A *degree-ℓ Kummer extension* of F is an extension of the form $F(\theta)$, where θ^ℓ is an element of $F \setminus F^\ell$.

Theorem 3.1 (See [38, Theorem 5.8.5, Proposition 5.8.7, and Theorem 5.8.12]). *Let ℓ be a prime and let F be a field that contains the ℓ-th roots of unity.*

(1) *Let $F' = F(\theta)$ be a Kummer extension of F, with $\theta^\ell = \alpha \in F \setminus F^\ell$. Then the minimal polynomial of θ is $T^\ell - \alpha$, and F' is a degree-ℓ Galois extension of F.*

(2) *Every degree-ℓ Galois extension F' of F is a Kummer extension.*

(3) *Let $F' = F(\sqrt[\ell]{\alpha})$ and $F'' = F(\sqrt[\ell]{\beta})$ be two Kummer extensions of F. Then $F' \cong F''$ if and only if $\alpha = \beta^j \gamma^\ell$ for some $\gamma \in F^\times$ and some $j \in \mathbb{Z}$ with $1 \le j \le \ell - 1$.*

(4) *Suppose F is a function field. Let $F' = F(\sqrt[\ell]{\alpha})$ be a Kummer extension, let P be a place of F, and let P' be a place of F' lying over P. Then*

$$e(P' \mid P) = \frac{\ell}{\gcd(\ell, v_P(\alpha))},$$

where v_P is the additive valuation associated to P.

Note in particular that statement (3) gives a bijection between the Kummer extensions of F and the nontrivial cyclic subgroups of $F^\times/(F^\times)^\ell$.

Now suppose we are given an odd prime ℓ and a prime power $q \equiv 1 \bmod 2\ell$, and let K be the rational function field over $\mathbb{F}_q$. We construct dihedral degree-ℓ

function fields over K with a given quadratic resolvent field K_2 by starting with the field K_2 and constructing, via Kummer's theorem, cyclic degree-ℓ extensions of K_2 that are Galois over K with Galois group $\mathcal{D}_\ell$. Our next proposition allows us to recognize when we have such an extension. Before stating the proposition, we note that the norm map from K_2 to K induces a norm map $K_2^\times/(K_2^\times)^\ell \to K^\times/(K^\times)^\ell$, and that the inclusion $K^\times \subset K_2^\times$ induces a conorm map $K^\times/(K^\times)^\ell \to K_2^\times/(K_2^\times)^\ell$.

Proposition 3.2. *Let K_2/K be a quadratic function field and let $K_2(\theta)$ be a Kummer extension of K_2, where $\theta^\ell = \alpha \in K_2^\times \setminus (K_2^\times)^\ell$. Let C be the cyclic subgroup of $K_2^\times/(K_2^\times)^\ell$ generated by the class of α. If C is contained in the image of the conorm map, then $K_2(\theta)$ is a cyclic Galois extension of K; if C is contained in the kernel of the norm map, then $K_2(\theta)$ is a Galois extension of K with group $\mathcal{D}_\ell$; and otherwise, $K_2(\theta)$ is not a Galois extension of K.*

Proof. Since K_2 is Galois over K, the group $\mathrm{Gal}(K^{\mathrm{sep}}/K)$ acts on $K_2^\times/(K_2^\times)^\ell$, and this action reflects the action of $\mathrm{Gal}(K^{\mathrm{sep}}/K)$ on the set of Kummer extensions of K_2 in K^{sep}. Thus, the field $L = K_2(\theta)$ is Galois over K if and only if $\omega(C) = C$ for all $\omega \in \mathrm{Gal}(K^{\mathrm{sep}}/K)$, and this will be the case if and only if $\tau(C) = C$ for the nontrivial automorphism τ of K_2 over K.

Suppose $\tau(C) = C$, so that L/K is Galois. Since τ^2 is the identity on C, we have $\tau(\alpha) = \alpha^i \gamma^\ell$ for some $\gamma \in K_2$ and $i = \pm 1$. Let ω be an element of order 2 in $\mathrm{Gal}(L/K)$, so that ω is a lift of τ. If $i = 1$ then we have $(\omega(\theta)/\theta)^\ell = \gamma^\ell$, so $\omega(\theta) = \theta\gamma\zeta$ for some ℓ-th root of unity $\zeta \in K$; replacing γ with $\gamma\zeta$, we may assume that $\zeta = 1$ and $\omega(\theta) = \theta\gamma$. Then

$$\theta = \omega^2(\theta) = \omega(\theta) \cdot \omega(\gamma) = \theta\gamma \cdot \omega(\gamma)$$

so $1 = N_{K_2/K}(\gamma)$. By Hilbert 90, we have $\gamma = \varepsilon/\tau(\varepsilon)$ for some $\varepsilon \in K_2$. Since $\tau(\alpha) = \alpha\gamma^\ell$, we find that $\alpha\varepsilon^\ell$ is fixed by τ, so the image of α in $K_2^\times/(K_2^\times)^\ell$ lies in the image of the conorm. On the other hand, if $i = -1$ then $\gamma^\ell = N_{K_2/K}(\alpha) \in K$. Since $\gamma \in K_2$ and K_2 is a quadratic extension of K, we must have $\gamma \in K$. Thus the image of α in $K_2^\times/(K_2^\times)^\ell$ lies in the kernel of the norm. We see that if C is neither in the image of the conorm nor in the kernel of the norm, then $K_2(\theta)$ is not Galois over K; this is the final statement of the proposition.

If C is in the image of the conorm, then $\alpha = \beta\gamma^\ell$ for some $\beta \in K$ and $\gamma \in K_2$. Then $K_2(\theta)$ is the composition of the quadratic extension K_2/K with the Kummer extension $K(\sqrt[\ell]{\beta})/K$, so $K_2(\theta)$ is Galois over K with cyclic Galois group.

Finally, suppose C is killed by the norm map, so that $N_{K_2/K}(\alpha) = \gamma^\ell$ for some $\gamma \in K$. Then $\tau(\alpha) = \gamma^\ell/\alpha$, so $\tau(C) = C$, and L is Galois over K. If we again let ω be an element of order 2 in $\mathrm{Gal}(L/K)$, then $\omega(\theta) = \gamma\zeta/\theta$ for some ℓ-th root of unity $\zeta \in K$. If we let σ be a generator of $\mathrm{Gal}(L/K_2)$, we find that $\omega\sigma\omega = \sigma^{-1}$, so $\mathrm{Gal}(L/K) \cong \mathcal{D}_\ell$. □

Elements of K_2 whose norms are ℓ-th powers in K have divisors of a specific type, described below.

Proposition 3.3. *Let $\alpha \in K_2^\times$. If $\mathrm{N}_{K_2/K}(\alpha) = \gamma^\ell$ for some $\gamma \in K^\times$, then the principal divisor of α takes the form*

$$(\alpha) = \ell E' + \sum_{i=1}^{(\ell-1)/2} i(D_i' - D_{-i}'),$$

where E' is a divisor of K_2, the D_i' are squarefree effective divisors of K_2 with pairwise disjoint support, and where $\tau(D_i') = D_{-i}'$ for all i. Consequently, every place of K lying under a place in the support of some D_i' splits in K_2.

Proof. Let P' be a place in the support of the principal divisor (α), and set $n_P = v_P((\alpha))$. Then by the division algorithm we can uniquely write $n_P = q\ell + r$ for some $q, r \in \mathbb{Z}$ with $|r| \le (\ell-1)/2$. Repeating this for all places in the support of (α), we see that the divisor of α can be written uniquely as

$$(\alpha) = \ell E' + \sum_{i=1}^{(\ell-1)/2} i(D_i' - D_{-i}'),$$

where the D_i' are squarefree effective divisors with pairwise disjoint support. Applying the norm map $\mathrm{N}_{K_2/K}$ to (α), we obtain

$$\begin{aligned}(\mathrm{N}_{K_2/K}(\alpha)) &= (\alpha) + (\tau(\alpha)) \\ &= \ell(E' + \tau(E')) + \sum_{i=1}^{(\ell-1)/2} i(D_i' - D_{-i}' + \tau(D_i') - \tau(D_{-i}')).\end{aligned}$$

As $\mathrm{N}_{K_2/K}(\alpha) = \gamma^\ell$, we see that

$$i(D_i' - D_{-i}' + \tau(D_i') - \tau(D_{-i}')) = 0 \quad \text{for } 1 \le i \le (\ell-1)/2.$$

This shows that $D_i' = 0$ if and only if $D_{-i}' = 0$. If $D_i' \ne 0$, then D_i' and D_{-i}' are effective and have disjoint support, forcing $D_i' = \tau(D_{-i}')$. □

3B. ***Virtual unit decomposition.*** Theorem 3.1 states that elements of $K_2^\times$ that generate the same subgroup of $K_2^\times/(K_2^\times)^\ell$ produce the same Kummer extension. We wish to construct distinct dihedral function fields by constructing distinct Kummer extensions of K_2. To that end, we decompose the group $K_2^\times/(K_2^\times)^\ell$ using a function field definition of virtual units, as inspired by H. Cohen's work on number fields [7]. In particular, we show how to construct a basis for the kernel of the norm map $K_2^\times/(K_2^\times)^\ell \to K^\times/(K^\times)^\ell$.

We define the (ℓ-)*virtual units* of K_2 to be the elements of the set

$$V_\ell = \{\alpha \in K_2^\times : (\alpha) \in \ell \operatorname{Div}^0 K_2\}.$$

The map from V_ℓ to $\operatorname{Div}^0 K_2$ that sends α to $(\alpha)/\ell$ induces a map from V_ℓ to $(\operatorname{Pic}^0 K_2)[\ell]$, the ℓ-torsion of the degree-0 divisor class group of K_2; this leads to the exact sequence

$$1 \longrightarrow \mathbb{F}_q^\times/(\mathbb{F}_q^\times)^\ell \longrightarrow V_\ell/(K_2^\times)^\ell \longrightarrow (\operatorname{Pic}^0 K_2)[\ell] \longrightarrow 0.$$

We also have an exact sequence

$$1 \longrightarrow V_\ell/(K_2^\times)^\ell \longrightarrow K_2^\times/(K_2^\times)^\ell \longrightarrow K_2^\times/V_\ell \longrightarrow 1. \tag{2}$$

To better understand the final term of this sequence, we set

$$I_\ell = \operatorname{Prin} K_2/(\operatorname{Prin} K_2 \cap \ell \operatorname{Div}^0 K_2)$$

and define a map $\varphi\colon K_2^\times \to I_\ell$ by $\varphi(\alpha) = (\alpha) + \operatorname{Prin} K_2 \cap \ell \operatorname{Div}^0 K_2$. Then φ is surjective and $\ker\varphi = V_\ell$, so $K_2^\times/V_\ell \cong I_\ell$. All told, we obtain this diagram of exact sequences, which represents a virtual unit decomposition:

$$\begin{array}{ccccccccc}
 & & 1 & & 1 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
1 & \longrightarrow & \mathbb{F}_q^\times/(\mathbb{F}_q^\times)^\ell & \longrightarrow & \mathbb{F}_q^\times/(\mathbb{F}_q^\times)^\ell & \longrightarrow & 1 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
1 & \longrightarrow & V_\ell/(K_2^\times)^\ell & \longrightarrow & K_2^\times/(K_2^\times)^\ell & \longrightarrow & K_2^\times/V_\ell & \longrightarrow & 1 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & (\operatorname{Pic}^0 K_2)[\ell] & \longrightarrow & \operatorname{Prin} K_2/\ell \operatorname{Prin} K_2 & \longrightarrow & I_\ell & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array} \tag{3}$$

The middle vertical sequence here shows that the divisor map from $K_2^\times/(K_2^\times)^\ell$ to $\operatorname{Prin} K_2/\ell \operatorname{Prin} K_2$ has kernel $\mathbb{F}_q^\times/(\mathbb{F}_q^\times)^\ell$. However, by Proposition 3.2, Kummer extensions of K_2 that are Galois over K with group $\mathcal{D}_\ell$ correspond to nontrivial cyclic subgroups of the kernel of the norm map from $K_2^\times/(K_2^\times)^\ell$ to $K^\times/(K^\times)^\ell$. We now describe how the divisor map behaves on this kernel.

Let H be the group

$$H = \left\{\alpha \in K_2^\times : \mathrm{N}_{K_2/K}(\alpha) \in (K^\times)^\ell\right\}, \tag{4}$$

so that $H/(K_2^\times)^\ell$ is the kernel of the norm map from $K_2^\times/(K_2^\times)^\ell$ to $K^\times/(K^\times)^\ell$.

Proposition 3.4. *The map*

$$H/(K_2^\times)^\ell \longrightarrow \operatorname{Prin} K_2/\ell \operatorname{Prin} K_2$$

(*induced from the divisor map*) *is injective, and its image is the group*

$$J_\ell = \{(\beta) + \ell \operatorname{Prin} K_2 \in \operatorname{Prin} K_2/\ell \operatorname{Prin} K_2 : \mathrm{N}_{K_2/K}((\beta)) \in \ell \operatorname{Prin} K\}.$$

Proof. Let (H) be the group of divisors of elements in H. First we claim that the sequence

$$1 \longrightarrow (\mathbb{F}_q^\times)^\ell \longrightarrow H \longrightarrow (H) \longrightarrow 0$$

is exact. To see this, note that the map sending an element of H to its divisor is clearly surjective. The kernel of this map is the set $H \cap \mathbb{F}_q^\times$. Let $k \in \mathbb{F}_q^\times$ and suppose $\mathrm{N}_{K_2/K}(k) \in (K^\times)^\ell$. Then $\mathrm{N}_{K_2/K}(k) = k\tau(k) = k^2 \in (K^\times)^\ell$. As squaring is an isomorphism of $\mathbb{F}_q^\times/(\mathbb{F}_q^\times)^\ell$, we have $k \in (\mathbb{F}_q^\times)^\ell$.

It follows from the exact sequence above that the divisor map

$$H/(K_2^\times)^\ell \longrightarrow \operatorname{Prin} K_2/\ell \operatorname{Prin} K_2$$

is injective. Its image is certainly contained in J_ℓ. To complete the proof, we must show that every element of J_ℓ lies in the image of $H/(K_2^\times)^\ell$.

Let $(\beta) + \ell \operatorname{Prin} K_2$ be an element of J_ℓ, where $\beta \in K_2^\times$ satisfies $\mathrm{N}_{K_2/K}((\beta)) \in \ell \operatorname{Prin} K$, say $\mathrm{N}_{K_2/K}((\beta)) = \ell(\gamma)$ for some $\gamma \in K^\times$. Then $\mathrm{N}_{K_2/K}(\beta) = c\gamma^\ell$ for some $c \in \mathbb{F}_q^\times$. If we let $d = c^{(\ell-1)/2}$, then $\mathrm{N}_{K_2/K}(d\beta) = (c\gamma)^\ell$, so $d\beta$ is an element of H whose image in $\operatorname{Prin} K_2/\ell \operatorname{Prin} K_2$ is $(\beta) + \ell \operatorname{Prin} K_2$. □

Proposition 3.5. *The image of* $(\operatorname{Pic}^0 K_2)[\ell]$ *in* $\operatorname{Prin} K_2/\ell \operatorname{Prin} K_2$ *is contained in* J_ℓ.

Proof. Suppose $D' \in \operatorname{Div}^0 K_2$ represents an element of $(\operatorname{Pic}^0 K_2)[\ell]$, so that $\ell D'$ is a principal divisor, say equal to (α) for some $\alpha \in K_2^\times$. Then the divisor of $\mathrm{N}_{K_2/K}(\alpha)$ is also an ℓ-multiple of a principal divisor. □

Let U_ℓ be the image of $H/(K_2^\times)^\ell$ in I_ℓ, so that

$$U_\ell = \{(\alpha) + \operatorname{Prin} K_2 \cap \ell \operatorname{Div}^0 K_2 : \alpha \in H\}.$$

Corollary 3.6. *The bottom row of Diagram* (3) *gives rise to an exact sequence*

$$0 \longrightarrow (\operatorname{Pic}^0 K_2)[\ell] \longrightarrow H/(K_2^\times)^\ell \longrightarrow U_\ell \longrightarrow 0,$$

which splits (*noncanonically*).

Proof. The sequence is obtained from combining the exact sequence

$$0 \longrightarrow (\mathrm{Pic}^0\, K_2)[\ell] \longrightarrow J_\ell \longrightarrow U_\ell \longrightarrow 0$$

of subgroups of the bottom row of Diagram (3) with the isomorphism $H/(K_2^\times)^\ell \cong J_\ell$. The sequence splits because all of the groups are ℓ-torsion. □

This corollary, together with Proposition 3.2, gives us the following theorem:

Theorem 3.7. *There is a one-to-one correspondence between Kummer extensions $K_{2\ell}/K_2$ such that $K_{2\ell}$ is Galois over K with group $\mathcal{D}_\ell$ and the set of nontrivial cyclic subgroups of* $(\mathrm{Pic}^0\, K_2)[\ell] \times U_\ell$.

3C. ***The discriminant divisors of $\mathcal{D}_\ell$ extensions.*** Now that we have established the correspondence of Theorem 3.7 for $\mathcal{D}_\ell$ Kummer extensions $K_{2\ell} = K_2(\sqrt[\ell]{\alpha})$ of K_2, it remains to compute the discriminant divisor of $K_\ell \subset K_2(\sqrt[\ell]{\alpha})$. In particular, we compute the discriminant divisor Δ_{K_ℓ} of K_ℓ in terms of (α) and Δ_{K_2}. We begin by describing the discriminant divisor $\Delta_{K_{2\ell}/K_2}$. Our description is simplified by the introduction of the following terminology.

Suppose α is an element of the group H defined by (4). Let $D'_1, \ldots, D'_{(\ell-1)/2}$ be the divisors arising from the representation of (α) as described in Proposition 3.3. We define the *ramification divisor* of α to be the divisor

$$D'_1 + \tau(D'_1) + \cdots + D'_{(\ell-1)/2} + \tau(D'_{(\ell-1)/2})$$

of K_2, and the *reduced ramification divisor* of α to be the divisor

$$\mathrm{N}_{K_2/K}(D'_1 + \cdots + D'_{(\ell-1)/2})$$

of K. Note that the ramification divisor is the conorm of the reduced ramification divisor.

Lemma 3.8. *Let K_2 be a quadratic function field over K. Suppose that $K_{2\ell} = K_2(\sqrt[\ell]{\alpha})$ is a Kummer extension of K_2 such that $K_{2\ell}/K$ is Galois with Galois group $\mathcal{D}_\ell$. Then*

$$\Delta_{K_{2\ell}/K_2} = (\ell - 1)D',$$

where D' is the ramification divisor of α.

Proof. By Theorem 3.1, for all places P' in the support $\mathrm{Supp}\, D'$ of the divisor D', there is a unique place P'' of $K_{2\ell}$ lying over P' such that $e(P'' \mid P') = \ell$. Furthermore, all other places of K_2 are unramified in $K_{2\ell}$. □

We now compute the degree of the discriminant divisor Δ_{K_ℓ}, which will in turn allow us to compute Δ_{K_ℓ} itself. To that end, we examine the characters of $\mathcal{D}_\ell$. For subgroups G of $\mathcal{D}_\ell$, let $\Psi(G)$ denote the induced character of $\mathcal{D}_\ell$ obtained from the trivial character of G (see [27, Chapter 3]). The fields K, K_2, K_ℓ and

$K_{2\ell}$ of Diagram (1) are the fixed fields of the four subgroups $\mathcal{D}_\ell$, $\mathcal{C}_\ell$, $\mathcal{C}_2$, and 1, respectively. The induced characters of these groups are linearly dependent and satisfy the relation

$$\Psi(1) + 2\Psi(\mathcal{D}_\ell) = 2\Psi(\mathcal{C}_2) + \Psi(\mathcal{C}_\ell).$$

Since the Artin L function of an induced character $\Psi(G)$ is the ζ function of the fixed field of G (see [16, Chapter 8]), we obtain

$$\zeta_{K_{2\ell}}(s)\zeta_K^2(s) = \zeta_{K_\ell}^2(s)\zeta_{K_2}(s).$$

From the functional equation of the ζ function, we have

$$\deg \Delta_{K_{2\ell}} + 2\deg \Delta_K = 2\deg \Delta_{K_\ell} + \deg \Delta_{K_2},$$

and since $\Delta_K = 0$ we find

$$\deg \Delta_{K_{2\ell}} = 2\deg \Delta_{K_\ell} + \deg \Delta_{K_2}. \tag{5}$$

By [32, Corollary 3.4.12(a)] we have $\mathrm{Diff}_{K_{2\ell}} = \mathrm{Con}_{K_{2\ell}/K_2}(\mathrm{Diff}_{K_2}) + \mathrm{Diff}_{K_{2\ell}/K_2}$. Applying norms yields

$$\Delta_{K_{2\ell}} = [K_{2\ell} : K_2]\Delta_{K_2} + \mathrm{N}_{K_2/K}(\Delta_{K_{2\ell}/K_2}).$$

By Lemma 3.8, we obtain

$$\Delta_{K_{2\ell}/K_2} = (\ell - 1)D',$$

where D' is the ramification divisor of any α that defines $K_{2\ell}$ as a Kummer extension of K_2. Let M be the reduced ramification divisor of α. Then

$$\mathrm{N}_{K_2/K}(\Delta_{K_{2\ell}/K_2}) = 2(\ell - 1)M,$$

and (5) can be rewritten as

$$\ell \deg \Delta_{K_2} + 2(\ell - 1)\deg M = 2\deg \Delta_{K_\ell} + \deg \Delta_{K_2}.$$

Thus,

$$\deg \Delta_{K_\ell} = \frac{\ell - 1}{2}\deg \Delta_{K_2} + (\ell - 1)\deg M.$$

Using this information we can now compute the discriminant divisor of K_ℓ.

Theorem 3.9. *With notation as above, we have* $\Delta_{K_\ell} = \frac{\ell-1}{2}\Delta_{K_2} + (\ell - 1)M$.

Proof. Let $E = \frac{\ell-1}{2}\Delta_{K_2} + (\ell - 1)M$. First note that the only places of K ramified in K_ℓ are those lying over places in the support of M and Δ_{K_2} as $K_{2\ell}/K_2/K$ is only ramified at these places. Moreover, for all places $P \in \mathrm{Supp}\, M$ and all $P'' \in \mathrm{Places}(K_{2\ell})$ lying over P, we have $e(P'' \mid P) = \ell$. Similarly, for all places $P \in \mathrm{Supp}\, \Delta_{K_2}$ and all $P'' \in \mathrm{Places}(K_{2\ell})$ lying over P, we have $e(P'' \mid P) = 2$.

As $[K_{2\ell} : K_\ell] = 2 \nmid \ell$, all places $P' \in \text{Places}(K_\ell)$ lying over M must have $e(P' \mid P) = \ell$. Also, for all $P' \in \text{Places}(K_\ell)$ lying over Δ_{K_2}, $e(P' \mid P) \leq 2$. Applying the identity

$$\sum_{P' \mid P} e(P' \mid P) f(P' \mid P) = \ell$$

to any place $P \in \text{Supp}\, \Delta_{K_2}$ allows at most $(\ell - 1)/2$ places $P' \mid P$ to be ramified. Thus, Δ_{K_ℓ} divides E. Since both divisors have the same degree, they must be equal. □

We note that the above proof in fact gives the complete decomposition of the ramified places of K_ℓ / K.

3D. ***The number of $\mathcal{D}_\ell$ function fields.*** We now prove the main result, Theorem 3.10, which provides the number of nonconjugate degree-ℓ dihedral extensions K_ℓ of K with fixed discriminant divisor $\Delta_{K_\ell} = \Delta$ and quadratic resolvent field K_2. We use the correspondence of Theorem 3.7 and the discriminant divisor result of Theorem 3.9. First, we introduce some more notation.

Let $M \in \text{Div}(K)$ be a squarefree effective divisor. Set $N = \#\,\text{Supp}\, M$, and suppose that every place $P_i \in \text{Supp}\, M$, $1 \leq i \leq N$, splits in K_2 as $P_i = P_i' + \tau(P_i')$ with $P_i' \neq \tau(P_i')$. We then define a set $\mathcal{Q}_\ell(M)$ of formal sums by

$$\mathcal{Q}_\ell(M) := \left\{ \sum_{i=1}^{N} n_i (P_i' - \tau(P_i')) : n_i \in (\mathbb{Z}/\ell\mathbb{Z})^\times \right\}.$$

We can view $\mathcal{Q}_\ell(M)$ as a subset of the group

$$\overline{\mathcal{Q}}_\ell(M) = \sum_{i=1}^{N} (\mathbb{Z}/\ell\mathbb{Z})(P_i' - \tau(P_i'));$$

note that the natural map $\text{Div}^0\, K_2 \to \text{Pic}^0\, K_2$ reduces to a homomorphism

$$\rho\colon \overline{\mathcal{Q}}_\ell(M) \longrightarrow \text{Pic}^0\, K_2 / \ell\, \text{Pic}^0\, K_2.$$

We set

$$T_\ell(M) := \{E' \in \mathcal{Q}_\ell(M) : \rho(E') = 0\}. \tag{6}$$

Theorem 3.10. *Let K_2 be a quadratic function field over $K = \mathbb{F}_q(x)$ with discriminant divisor Δ_{K_2}, with $q \equiv 1 \bmod 2\ell$. Let r denote the ℓ-rank of* $\text{Pic}^0\, K_2$, *and let M be a divisor of K that is either zero or a sum of distinct places of K supported away from Δ_{K_2}. Let $\Delta = \frac{\ell-1}{2}\Delta_{K_2} + (\ell - 1)M$.*

(1) *If $M = 0$, then the number of nonconjugate dihedral degree-ℓ function fields K_ℓ / K with discriminant divisor $\Delta_{K_\ell} = \Delta$ and quadratic resolvent field K_2 is $(\ell^r - 1)/(\ell - 1)$.*

(2) *If $M \neq 0$ and some $P \in \operatorname{Supp} M$ is inert in K_2/K, then there are no dihedral degree-ℓ function fields K_ℓ/K with discriminant divisor $\Delta_{K_\ell} = \Delta$ and quadratic resolvent field K_2.*

(3) *Suppose $M \neq 0$ and that all $P_i \in \operatorname{Supp} M$ split in K_2 as $P_i = P_i' + \tau(P_i')$ with $P_i' \neq \tau(P_i')$. Then the number of nonconjugate dihedral degree-ℓ function fields with discriminant divisor $\Delta_{K_\ell} = \Delta$ and quadratic resolvent field K_2 is $\#T_\ell(M)\ell^r/(\ell-1)$, where $T_\ell(M)$ is defined by* (6).

Proof. Let $U_{\ell,M}$ denote the subset of U_ℓ consisting of those classes

$$(\alpha) + \operatorname{Prin} K_2 \cap \ell \operatorname{Div}^0 K_2$$

such that the reduced ramification divisor of α is equal to M. Note that $U_{\ell,M}$ is closed under multiplication by nonzero elements of $\mathbb{Z}/\ell\mathbb{Z}$.

Using the correspondence of Theorem 3.7, the conjugacy classes of dihedral degree-ℓ function fields with discriminant divisor $\Delta_{K_\ell} = \Delta$ and quadratic resolvent field K_2 are in one-to-one correspondence with the number of nontrivial cyclic subgroups of $(\operatorname{Pic}^0 K_2)[\ell] \times U_\ell$ that can be generated by elements (A, B) with $B \in U_{\ell,M}$.

If $M = 0$, then $U_{\ell,M}$ consists of the identity, so $B = 0$ and A can be any nonzero class in $(\operatorname{Pic}^0 K_2)[\ell]$. There are $\ell^r - 1$ such pairs, and they generate $(\ell^r-1)/(\ell-1)$ different cyclic subgroups.

If $M \neq 0$, then $\#U_{\ell,M} = \#T_\ell(M)$. This is because an element α of H gives rise to an element of $U_{\ell,M}$ if and only if its divisor is of the form E' (up to multiples of ℓ) for some E' in $T_\ell(M)$. Thus, there are $\#T_\ell(M)\ell^r$ pairs (A, B) in $(\operatorname{Pic}^0 K_2)[\ell] \times U_\ell$ with $B \in U_{\ell,M}$, and there are $\#T_\ell(M)\ell^r/(\ell-1)$ cyclic subgroups generated by such pairs. □

3E. ***Defining equations.*** In this section, we will write down explicit defining equations for $\mathcal{D}_\ell$ extensions of K constructed as above.

Definition 3.11. Given an integer $n > 0$ and an element γ of K, let $C_{n,\gamma}$ be the polynomial

$$C_{n,\gamma}(X) = \sum_{r=0}^{\lfloor n/2 \rfloor} (-\gamma)^r \frac{n}{n-r} \binom{n-r}{r} X^{n-2r}$$

in $K[X]$. (Note that the coefficient $\frac{n}{n-r}\binom{n-r}{r}$ is in fact an integer, so the definition makes sense in positive characteristic; see [30, Sequence A082985].)

The polynomials $C_{n,\gamma}$ are scaled versions of the Chebyshev polynomials of the first kind, and it follows that if u and v are elements of a field extension L of K that satisfy $uv = \gamma$, then

$$C_{n,\gamma}(u + v) = u^n + v^n.$$

Proposition 3.12. *Let ℓ be an odd prime, let $q \equiv 1 \bmod 2\ell$ be a prime power, and let K_2 be a quadratic extension of $K = \mathbb{F}_q(x)$. Let α be an element of $K_2 \setminus K_2^\ell$ such that $\mathrm{N}_{K_2/K}(\alpha) = \gamma^\ell$ for some $\gamma \in K$, and let $K_{2\ell}$ be the Kummer extension $K_2(\sqrt[\ell]{\alpha})$, so that $K_{2\ell}/K$ is Galois with group $\mathcal{D}_\ell$. Then the roots in $K_{2\ell}$ of the polynomial*

$$C_{\ell,\gamma}(X) - \mathrm{Tr}_{K_2/K}(\alpha)$$

are generators for the index-2 subfields of $K_{2\ell}/K$.

Proof. Let θ be a root of $z^\ell - \alpha$, let σ be a generator of $\mathrm{Gal}(K_{2\ell}/K_2)$, and let τ be an element of $\mathrm{Gal}(K_{2\ell}/K)$ that restricts to the nontrivial element of $\mathrm{Gal}(K_2/K)$. Then $\tau(\theta)$ and γ/θ are both roots of $z^\ell - \tau(\alpha)$, so $\tau'(\theta) = \gamma/\theta$ for some $\tau' = \sigma^i\tau$. Thus, $\theta + \gamma/\theta$ lies in the fixed field of τ' (but does not lie in K, for otherwise θ would lie in a quadratic extension of K).

It follows that

$$C_{\ell,\gamma}(\theta + \gamma/\theta) = \theta^\ell + (\gamma/\theta)^\ell = \alpha + \tau(\alpha) = \mathrm{Tr}_{K_2/K}(\alpha),$$

so one of the roots of $C_{\ell,\gamma}(X) - \mathrm{Tr}_{K_2/K}(\alpha)$ generates an index-2 subfield of $K_{2\ell}/K$. Since all of these index-2 subfields are conjugate to one another, the other roots of the polynomial generate the other fields. □

4. Algorithms and data

4A. ***Construction algorithm.*** The correspondence of Theorem 3.7 can be made explicit, and the proof of Theorem 3.10 is constructive; this leads naturally to Algorithm 4.1 below. This algorithm takes as input a quadratic function field K_2 and an effective squarefree divisor M of K, and outputs all nonconjugate degree-ℓ dihedral function fields with discriminant divisor $\frac{\ell-1}{2}\Delta_{K_2} + (\ell-1)M$ and quadratic resolvent field K_2. Note that K_2 may be the unique degree-2 constant field extension of K, in which case $\Delta_{K_2} = 0$.

Algorithm 4.1 (Constructing all $\mathcal{D}_\ell$ function fields with a given quadratic resolvent and given ramification divisor).

Input: A quadratic extension K_2 of K, an odd prime ℓ, and a squarefree effective divisor M of K with support disjoint from that of Δ_{K_2}.

Output: A set L of defining equations for all the dihedral extensions K_ℓ of K with $\Delta_{K_\ell} = \frac{\ell-1}{2}\Delta_{K_2} + (\ell-1)M$ and with QuadRes $K_\ell = K_2$.

1. Compute fundamental information:
   (a) Compute a basis $\{[B_1], \ldots, [B_r]\}$ of $(\mathrm{Pic}^0 K_2)[\ell]$ and an element ζ of $\mathbb{F}_q^\times \setminus \mathbb{F}_q^{\times\ell}$.
   (b) Set $N \leftarrow \varnothing$; eventually, N will contain the pairs of places of K_2 lying over the support of M.

(c) For $P \in \operatorname{Supp} M$:
   i. Ensure $P = P_0' + P_1'$ in $\operatorname{Div} K_2$; upon failure, return the empty set.
   ii. $N \leftarrow N \cup \{(P_0', P_1')\}$.

(d) Use N to compute the set $\mathcal{Q}_\ell(M)$.

2. Compute functions in H representing elements of $\mathcal{Q}_\ell(M)$ that map into U_ℓ:
   (a) Set $T \leftarrow \varnothing$; eventually, T will contain lifts to H of all elements of $\mathcal{Q}_\ell(M)$ (up to the action of $(\mathbb{Z}/\ell\mathbb{Z})^\times$) that can be lifted to H.
   (b) For $E' \in \mathcal{Q}_\ell(M)$ up to the action of $(\mathbb{Z}/\ell\mathbb{Z})^\times$ such that $\rho(E') = 0$:
      i. Find $\beta \in K_2^\times$ such that $(\beta) \equiv E' \bmod \ell$.
      ii. Repeat $\beta \leftarrow \zeta\beta$ until $\mathrm{N}_{K_2/K}(\beta) \in (K^\times)^\ell$.
      iii. $T \leftarrow T \cup \{\beta\}$.

3. Compute virtual units in H:
   (a) Set $V \leftarrow \varnothing$; eventually V will contain elements of $H \cap V_\ell$ whose images in $V_\ell/(K_2^\times)^\ell$ form a basis for that group.
   (b) For $[B_i]$ in the basis of $(\operatorname{Pic}^0 K_2)[\ell]$ computed in step 1(a):
      i. Find $\eta_i \in K_2$ such that $(\eta_i) = \ell B_i$.
      ii. Repeat $\eta_i \leftarrow \zeta\eta_i$ until $\mathrm{N}_{K_2/K}(\eta_i) \in (K^\times)^\ell$.
      iii. $V \leftarrow V \cup \{\eta_i\}$.

4. Create defining equations:
   (a) Set $L \leftarrow \varnothing$.
   (b) If $M = 0$ then for all nonzero $(z_i) \in (\mathbb{Z}/\ell\mathbb{Z})^r$ up to the action of $(\mathbb{Z}/\ell\mathbb{Z})^\times$:
      i. Compute $\alpha := \prod_{i=1}^r \eta_i^{z_i}$ and $\gamma \in K$ with $\gamma^\ell = \mathrm{N}_{K_2/K}(\alpha)$.
      ii. Let $C(X) \leftarrow C_{\ell,\gamma}(X) - \mathrm{Tr}_{K_2/K}(\alpha)$, as in Proposition 3.12.
      iii. $L \leftarrow L \cup \{C(X)\}$.
   (c) If $M \neq 0$ then for all $\beta \in T$ and for all $(z_i) \in (\mathbb{Z}/\ell\mathbb{Z})^{\#V}$:
      i. Compute $\alpha := \beta \prod_{\gamma_i \in V} \eta_i^{z_i}$ and $\gamma \in K$ with $\gamma^\ell = \mathrm{N}_{K_2/K}(\alpha)$.
      ii. Let $C(X) \leftarrow C_{\ell,\gamma}(X) - \mathrm{Tr}_{K_2/K}(\alpha)$, as in Proposition 3.12.
      iii. $L \leftarrow L \cup \{C(X)\}$.
   (d) Return L.

Algorithm 4.1 is precisely the construction in the proof of Theorem 3.10, and thus computes all elements α such that $K_2(\sqrt[\ell]{\alpha})$ is a Galois dihedral function field. Notice that the repeat loops in steps 2(b)(ii) and 3(b)(ii) will halt, as by Proposition 3.4, there is a unique $\beta \in K_2^\times$ with $(\beta) = B' - \tau(B') - \ell E'$ and $\beta \in H$; similarly for η.

Remarks 4.2. There are several ways to perform Algorithm 4.1 more efficiently.

(1) The generators $[B_1], \ldots, [B_r]$ of $(\mathrm{Pic}^0 K_2)[\ell]$ in step 1(a) can be computed from a set of generators $[A_1], \ldots, [A_h]$ of $\mathrm{Pic}^0 K_2$ chosen so that the order m_i of $[A_i]$ is equal to the i-th invariant factor of the group $\mathrm{Pic}^0 K_2$. Using the $[A_i]$, it is also easy to check whether an element E' of $\mathcal{Q}_\ell(M)$ is in the kernel of the map ρ, and, if so, to obtain an element $\beta \in K_2^\times$ such that $(\beta) \equiv E' \bmod \ell$, as is required in step 2(b)(i). We do this as follows: Suppose D' is a lift of E' to the degree-0 divisor group of K_2. Write $[D'] = d_1[A_1] + \cdots + d_h[A_h]$. Then E' is in the kernel of ρ if and only if ℓ divides d_i whenever m_i is divisible by ℓ. If this is the case, set $e_i = d_i/\ell$ when $\ell \mid m_i$ and $e_i \equiv d_i \ell^{-1} \bmod m_i$ when $\ell \nmid m_i$. Then $D' - \ell(e_1 A_1 + \cdots e_h A_h)$ is principal, and we can compute $\beta \in K_2^\times$ with this divisor; this is the desired β.

(2) When K_2 has positive genus, it is the function field of an elliptic or hyperelliptic curve $y^2 = f(x)$. One could potentially take advantage of faster arithmetic available for the Jacobians of hyperelliptic curves, instead of the slower generic arithmetic in $\mathrm{Pic}^0 K_2$.

Algorithm 4.3 takes as input a pair of effective squarefree divisors D and M of K with disjoint support and uses Algorithm 4.1 to generate all nonconjugate degree-ℓ dihedral function fields K_ℓ with discriminant divisor $\frac{\ell-1}{2} D + (\ell - 1)M$. It takes advantage of the following observation: In order for any degree-ℓ dihedral function fields K_ℓ to exist, D must be the discriminant divisor of a quadratic function field — that is, effective, squarefree, and of even degree. Moreover, all the places in the support of M must be split over the quadratic resolvent field of K_ℓ, which has discriminant divisor D. If $D = 0$, then this field is the unique quadratic constant field extension of K. If D is nonzero, then there are exactly two quadratic function function fields K_2 and K_2' of discriminant divisor D; they are in fact twists of one another. Any place $P \notin \mathrm{Supp}\, D$ splits in K_2 if and only if it is inert in K_2', and vice versa. Thus, if M is nonzero, only one of K_2 and K_2' needs to be considered in the construction of K_ℓ.

Algorithm 4.3 (Constructing all $\mathcal{D}_\ell$ function fields from divisors).

Input: An odd prime ℓ and squarefree effective divisors D and M of K with disjoint support.

Output: A set L of defining equations for all the degree-ℓ dihedral extensions K_ℓ of K with $\Delta_{K_2} = D$ and $\Delta_{K_\ell} = \frac{\ell-1}{2} D + (\ell - 1)M$.

1. If $\deg D$ is even, construct a quadratic field K_2 with discriminant divisor D; otherwise, return "D IS NOT A QUADRATIC DISCRIMINANT DIVISOR".
2. If $D = 0$, get L from Algorithm 4.1 with input K_2, ℓ, M, return L, and stop.
3. Construct the quadratic twist K_2' of K_2.

4. If $M = 0$ then:
   (a) Get L_1 from Algorithm 4.1 with input K_2, ℓ, M.
   (b) Get L_2 from Algorithm 4.1 with input K_2', ℓ, M.
   (c) Return $L_1 \cup L_2$, and stop.
5. Pick $P \in \operatorname{Supp} M$.
6. If $P = P_0' + P_1'$ in $\operatorname{Div} K_2$ then set $K_2'' \leftarrow K_2$; otherwise, set $K_2'' \leftarrow K_2'$.
7. Get L from Algorithm 4.1 with input K_2'', ℓ, M, return L, and stop.

All finite places P of K correspond to irreducible polynomials $f_P(x) \in \mathbb{F}_q[x]$. Therefore, in step 1 we can easily construct $K_2 = K(y)$ as follows: If $D = 0$, then y is simply the square root of a nonsquare in $\mathbb{F}_q$. If $D \neq 0$, then K_2 is the function field of the curve

$$y^2 = \prod_{\substack{P \in \operatorname{Supp} D \\ P \text{ finite}}} f_P(x).$$

4B. ***Tabulation algorithm.*** Algorithm 4.1 constructs all degree-ℓ dihedral function fields with a given discriminant divisor and quadratic resolvent field; by iterating this algorithm, we obtain a procedure for tabulating all degree-ℓ dihedral function fields whose discriminant divisor has degree at most some fixed input bound $B \geq 0$. However, in this context, we can use the automorphism group of K to significantly reduce the number of quadratic function fields that need to be considered.

Recall that $\operatorname{Aut} K = \operatorname{Aut} \mathbb{F}_q(x)$ is isomorphic to $\mathrm{PGL}(2, q)$, the group of fractional linear transformations of x over $\mathbb{F}_q$. The group $\operatorname{Aut} K$ also acts on the set of extension fields of K, and for every $\phi \in \operatorname{Aut} K$ we have $\phi(\Delta_{K_i}) = \Delta_{\phi(K_i)}$. Therefore, instead of applying Algorithm 4.1 to all suitable K_2 and M, we only need to consider a representative from each orbit of $\operatorname{Aut} K$ acting on the set of suitable quadratic function fields K_2. Moreover, for each such field K_2 we need only consider representatives of the action of the stabilizer $\operatorname{Stab} K_2 \subseteq \mathrm{PGL}(2, q)$ on the set of suitable M.

These ideas are captured below in three algorithms. We start with Algorithm 4.4, which, given an integer B, finds orbit representatives for the set of quadratic function fields whose discriminant divisors are of degree at most $2B/(\ell - 1)$.

Recall that every quadratic function field K_2 can be expressed as $K(y)$, where y^2 is equal to either a nonsquare in $\mathbb{F}_q$ or a squarefree polynomial $f(x) \in \mathbb{F}_q[x]$ of degree $2g + 1$ or $2g + 2$, where g is the genus of K_2. In the former case, K_2 is fixed under $\mathrm{PGL}(2, q)$. In the latter case, the action of $\phi \in \mathrm{PGL}(2, q)$ on K_2 does not necessarily preserve the degree of $f(x)$, but $\phi(K_2)$ has the same genus as K_2; in fact, the discriminant divisors of K_2 and $\phi(K_2)$ have the same degree, namely $2g + 2$.

In the following algorithm, we will let $P(q, \ell, B, h)$ denote the set of nonconstant squarefree polynomials $f \in \mathbb{F}_q[x]$ whose degrees satisfy

$$\lceil \deg(f)/2 \rceil \leq \lfloor 2B/(\ell - 1) \rfloor$$

and whose leading coefficient is either 1 or a fixed nonsquare $h \in \mathbb{F}_q$.

Algorithm 4.4 (Constructing a list of $\mathrm{PGL}(2, q)$-orbit representatives for quadratic function fields of bounded discriminant).

Input: A nonnegative integer B, an odd prime ℓ, and a prime power $q \equiv 1 \bmod 2\ell$.

Output: A set R'_B of pairs (f, S) such that each f is a squarefree element of $\mathbb{F}_q[x]$ such that $K_2 := K[y]/(y^2 - f)$ has discriminant divisor of degree at most $2B/(\ell - 1)$, each S is the $\mathrm{PGL}(2, q)$-stabilizer of the class of f in $K^\times/(K^\times)^2$, and such that every quadratic extension K_2 of K with $\deg \Delta_{K_2} \leq 2B/(\ell - 1)$ has exactly one $\mathrm{PGL}(2, q)$-orbit representative in the collection of fields defined by the f.

1. Compute a primitive element h of $\mathbb{F}_q$.
2. Initialize $R'_B \leftarrow \{(h, \mathrm{PGL}(2, q))\}$.
3. Set $L(f) \leftarrow 0$ for all $f \in P(q, \ell, B, h)$.
4. For all $f \in P(q, \ell, B, h)$:
   (a) If $L(f) = 0$ then
      i. $S \leftarrow \varnothing$.
      ii. For all $\phi = \frac{ax+b}{cx+d} \in \mathrm{PGL}(2, q)$:
         - $f_1(x) \leftarrow (cx + d)^{2\lceil (\deg f)/2 \rceil} f(\phi(x))$.
         - If the leading coefficient m of f_1 is a square, replace f_1 with f_1/m; otherwise, replace f_1 with hf_1/m.
         - $L(f_1) \leftarrow 1$.
         - If $f_1 = f$, then $S \leftarrow S \cup \{\phi\}$.
      iii. $R'_B \leftarrow R'_B \cup \{(f, S)\}$.
5. Return R'_B.

Next we have Algorithm 4.5, which constructs minimal polynomials for all dihedral function fields with discriminant divisors $\frac{\ell-1}{2}\Delta_{K_2} + (\ell - 1)M$ for representatives K_2 and M obtained from Algorithm 4.4.

Algorithm 4.5 (Tabulating $\mathrm{PGL}(2, q)$-orbit representatives of dihedral function fields with bounded discriminant).

Input: A nonnegative integer B, an odd prime ℓ, a prime power $q \equiv 1 \bmod 2\ell$, and the set R'_B computed by Algorithm 4.4 on input B, ℓ, q.

Output: A set R_B of triples (L_2, Δ, S') such that each Δ is an effective divisor of K of degree at most B, the group S' is the $\mathrm{PGL}(2,q)$-stabilizer of Δ, the set L_2 consists of equations defining $\mathcal{D}_\ell$ extensions of K with discriminant divisor Δ, and such that every $\mathcal{D}_\ell$ extension of K with discriminant divisor of degree at most B has a unique $\mathrm{PGL}(2,q)$-orbit representative in the collection of fields defined by the elements of the L_2.

1. Initialize $R_B \leftarrow \varnothing$.
2. For $(f, S) \in R'_B$:
   (a) Construct $K_2 = K(x)[y]/(y^2 - f)$ and compute Δ_{K_2}.
   (b) Compute $B' = \lfloor B/(\ell-1) - (\deg \Delta_{K_2})/2 \rfloor$.
   (c) Initialize $\mathcal{M} \leftarrow \varnothing$; eventually, $\mathcal{M}$ will contain all effective squarefree divisors of K with support disjoint from Δ_{K_2} and degree at most B'.
   (d) Compute lists
   $$L_j = \{P \in \mathrm{Places}(K) \setminus \mathrm{Supp}\, \Delta_{K_2} : \deg P = j\}$$
   for $1 \le j \le B'$.
   (e) For i from 0 to B' and for every partition $\mathbf{n} = [n_1, .., n_r]$ of i:
      i. Generate the set $W_{\mathbf{n}} = \left\{\sum_{k=1}^{r} P_k : P_k \in L_{n_k}\right\}$.
      ii. $\mathcal{M} \leftarrow \mathcal{M} \cup W_{\mathbf{n}}$.
   (f) Compute the set $\mathcal{M}_S$ of all pairs (M, S') where each $M \in \mathcal{M}$ is a unique orbit representative of S acting on $\mathcal{M}$ and S' is the stabilizer of M in S.
   (g) For $(M, S') \in \mathcal{M}_S$:
      i. Get L_2 from Algorithm 4.1 on input (K_2, ℓ, M).
      ii. Compute $\Delta = \frac{\ell-1}{2}\Delta_{K_2} + (\ell-1)M$.
      iii. $R_B \leftarrow R_B \cup \{(L_2, \Delta, S')\}$.
3. Return R_B.

Finally, Algorithm 4.6 reapplies $\mathrm{Aut}\, K$ to each of the constructed minimal polynomials to obtain the full list of degree-ℓ dihedral function fields whose discriminant divisor has degree bounded by B.

Algorithm 4.6 (Tabulating the full list of dihedral function fields with bounded discriminant).

Input: A nonnegative integer B, an odd prime ℓ, a prime power $q \equiv 1 \bmod 2\ell$, and the set R_B computed by Algorithm 4.5 on input B, ℓ, q.

Output: A set L_B of defining equations for all the dihedral extensions K_ℓ of K with $\deg \Delta_{K_\ell} \le B$.

1. Initialize $L_B \leftarrow \varnothing$.

2. For $(L, \Delta, S') \in R_B$:
   (a) For all distinct representatives ϕ of cosets in $\mathrm{PGL}(2,q)/S'$ and for all $C(X) \in L$, set $L_B \leftarrow L_B \cup \{(\phi(C(X)), \phi(\Delta))\}$.
3. Return L_B.

4C. ***Numerical results.*** We implemented our algorithms in Magma [5]. In Table 1, we provide data for all odd primes ℓ, prime powers $q \equiv 1 \bmod 2\ell$, and multiples $B > \ell-1$ of $\ell-1$ such that $q^{2B/(\ell-1)+1} < 2^{29}$. The column headed $K_2/\sim$ gives the number of quadratic function fields generated by Algorithm 4.4. The number of function fields constructed by Algorithm 4.5 is given in the column headed $K_\ell/\sim$, and the total number of nonconjugate dihedral degree-ℓ function fields whose discriminant divisor has degree at most B is listed in the column headed K_ℓ. The running times of Algorithms 4.4, 4.5, and 4.6 are listed in the next three columns. For each ℓ, q and B, we also computed the value $R = (q^3-q)T_5/(T_4+T_5+T_6)$, where T_i denotes the running time of Algorithm 4.i for $i = 4, 5, 6$. The quantity R estimates the improvement factor obtained by our tabulation method relative to simply iterating Algorithm 4.1 over all possible quadratic function fields without using the $\mathrm{PGL}(2,q)$ action.

Notice that the improvement factor R is highly varied. For fixed ℓ and B, R tends to decrease as q increases although the improvement still remains significant. Why this decrease occurs is unclear; it may be due to the fact that R is not a sufficiently refined estimate for the actual running time improvement. Overall, the running time of Algorithm 4.1 is dominated by the construction of the set $\mathcal{Q}_\ell(M)$ and obtaining functions for the principal divisors in steps 2(b)(i) and 3(b)(i). Data suggests that as B grows, finding the generators of these principal divisors will tend to dominate the running time. Using Jacobian arithmetic as opposed to divisor arithmetic as suggested in part (2) of Remarks 4.2 improved the performance of our tabulation only very marginally, even for larger parameters.

The entries of columns 4 and 5 of Table 1 differ by a factor that is very close to $\ell-1$; in other words, for the data we collected, it looks like the number of quadratic extensions of K with discriminant degree at most $2B/(\ell-1)$ is about $\ell-1$ times as large as the number of $\mathcal{D}_\ell$ extensions of K with discriminant degree at most B. When $B = 2(\ell-1)$ this is explained by the results of the following section, but we do not know whether it is true in general.

5. A formula for the case $B = 2(\ell-1)$

In this section we give an explicit formula for the number of $\mathcal{D}_\ell$ extensions whose discriminant divisor has degree $2(\ell-1)$.

First we note that there are no $\mathcal{D}_\ell$ extensions with discriminant of degree smaller than $2(\ell-1)$. To see this, suppose K_ℓ is a $\mathcal{D}_\ell$ extension of K with Galois closure

						Running time (seconds)			
ℓ	q	B	$K_2/\sim$	$K_\ell/\sim$	K_ℓ	Alg. 4.4	Alg. 4.5	Alg. 4.6	R
3	7	4	33	17	2,373	0.9	1.1	0.8	132.0
		6	782	472	117,285	25.8	35.2	47.1	109.4
		8	35,010	18,149	5,763,093	1,321.5	2,416.9	2,505.1	130.1
	13	4	61	33	28,470	13.7	3.2	9.5	264.7
		6	4,650	2,564	4,824,534	1,379.5	286.1	1,870.2	176.7
	19	4	81	41	130,131	82.8	7.6	44.5	385.4
	25	4	109	57	390,300	726.8	17.6	149.1	307.3
	31	4	129	65	923,025	821.0	31.7	357.2	779.7
	37	4	157	81	1,873,458	1,983.1	56.5	731.7	1,031.9
	43	4	177	89	3,417,855	4,040.5	100.2	1,341.9	1,452.3
	49	4	205	105	5,763,576	20,544.4	189.6	2,376.5	964.8
5	11	8	45	9	6,655	6.1	1.4	2.7	181.2
		12	2,858	949	1,058,695	461.5	102.9	463.5	132.1
	31	8	109	33	446,865	821.0	29.2	191.0	834.6
	41	8	169	45	1,378,420	3,178.2	80.5	602.0	1,436.2
7	29	12	121	19	219,646	546.8	22.6	94.8	828.9
	43	12	177	29	1,086,911	4,000.5	95.2	567.8	1,622.2
11	23	20	93	8	48,829	192.7	10.1	23.8	541.3
13	53	24	217	21	1,340,794	10,935.6	235.8	742.8	2,945.5
23	47	44	189	11	519,961	5,951.6	184.2	364.9	2,940.5

Table 1. Function field counts for all ℓ and $q \equiv 1 \bmod 2\ell$ with $q^{\frac{2B}{\ell-1}+1} < 2^{29}$, for $B \geq 2(\ell-1)$. For each ℓ, q, and B given in the first three columns, we list in column 4 the number of $\mathrm{PGL}(2,q)$-equivalence classes of quadratic extension of $K = \mathbb{F}_q(x)$ whose discriminants have degree at most $2B/(\ell-1)$. In column 5, we list the number of $\mathrm{PGL}(2,q)$-equivalence classes of $\mathcal{D}_\ell$ extensions of K whose discriminants have degree at most B, and in column 6 we list the total number of such extensions. In the next three columns we give the running times of the algorithms that computed these quantities, and in the final column we give an estimate of the improvement in running time obtained by using the $\mathrm{PGL}(2,q)$ action in our computations. (Computations were carried out on one core of a 2GHz Intel Xeon X7550, with 64GB of available RAM.)

$K_{2\ell}$ and quadratic resolvent K_2. Theorem 3.9 gives $\Delta_{K_\ell} = \frac{\ell-1}{2}\Delta_{K_2} + (\ell-1)M$, where M is as in Section 3C. Quadratic extensions have discriminants of even degree, so $\deg \Delta_{K_\ell}$ is divisible by $\ell-1$. If $\deg \Delta_{K_\ell}$ were zero, K_ℓ/K would be a constant field extension, and would not have Galois group $\mathcal{D}_\ell$. If $\deg \Delta_{K_\ell}$ were $\ell-1$, then either K_2 would have genus 0 and $\deg M = 0$, or K_2/K would be a constant field extension and $\deg M = 1$. In the first case, $K_{2\ell}/K_2$ would be unramified and hence a constant field extension, so K_ℓ/K would also be a constant

field extension, a contradiction. In the second case, M would be a single place of degree 1; since every place in M must split in K_2, and since the places of K that split in a quadratic constant field extension are the places of even degree, we again have a contradiction. On the other hand, there do exist $\mathcal{D}_\ell$ extensions with discriminant divisor of degree $2(\ell-1)$, as the following theorem shows.

Theorem 5.1. *Let ℓ be an odd prime and let q be a prime power with $q \equiv 1 \bmod 2\ell$. For every nonnegative even integer d, let N_d be the number of $\mathcal{D}_\ell$ extensions of K whose discriminant divisors have degree $2(\ell-1)$ and whose quadratic resolvents have discriminant divisors of degree d. Let X be the modular curve $X_1(\ell)$. Then*

$$\frac{N_d}{q^3-q} = \begin{cases} \dfrac{1}{2q+2} & \text{if } d=0, \\ 1 & \text{if } d=2, \\ -2+\dfrac{2\#X(\mathbb{F}_q)}{\ell-1} & \text{if } d=4, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 5.2. For $\ell = 3$, 5, and 7, the modular curve $X_1(\ell)$ has genus 0, so for these values of ℓ the formula for N_4 simplifies to

$$\frac{N_4}{q^3-q} = \frac{2(q-\ell+2)}{\ell-1}.$$

Equations for $X_1(\ell)$ for larger values of ℓ are known. For example, Sutherland [33] gives equations for all $\ell \le 47$; as of this writing, Sutherland's online tables [34] extend the results of [33] up to $\ell = 181$.

Proof of Theorem 5.1. Theorem 3.9 shows that if K_ℓ is a $\mathcal{D}_\ell$ extension of K with quadratic resolvent K_2, and if $\deg \Delta_{K_\ell} = 2(\ell-1)$, then $\deg \Delta_{K_2}$ is 0, 2, or 4.

Let us count the number of $\mathcal{D}_\ell$ extensions K_ℓ such that $\deg \Delta_{K_2} = 0$; that is, such that K_2 is the unique quadratic extension of K obtained by extending the constant field from $\mathbb{F}_q$ to $\mathbb{F}_{q^2}$. In this case, we must have $\deg M = 2$. We know that every place in M splits in K_2, and since the places of K that split in K_2 are precisely the places of even degree, M must consist of a single degree-2 place P.

If $\alpha \in K_2$ gives rise to a $\mathcal{D}_\ell$ extension of K, its divisor is of the shape given in Proposition 3.3, where exactly one of the D_i' with $i > 0$ is nonzero (and consists of a place of K_2 lying over P). Replacing α by a power if necessary, we may assume that D_1' and D_{-1}' are the only nonzero D_i', and we can choose which of the two places above P appears in D_1' and which in D_{-1}'. Since K_2 has genus 0, we can modify α by an ℓ-th power so that the divisor E' from the proposition is 0. If we let x be a generator of K, so that $K_2 \cong \mathbb{F}_{q^2}(x)$, then $\alpha = b(x-c)/(x-c^q)$ for

some $b \in \mathbb{F}_{q^2}$ and $c \in \mathbb{F}_{q^2} \setminus \mathbb{F}_q$, and we see that $N_{K_2/K}(\alpha) = b^{q+1}$. Since this norm is supposed to be an ℓ-th power, b itself must be an ℓ-th power, so we may replace α by α/b. We find that for every degree-2 place P of K, we obtain exactly one $\mathcal{D}_\ell$ extension of K_2, so $N_0 = (q^2 - q)/2$. This leads to the formula for N_0 in the statement of the theorem.

Now let us count the number of $\mathcal{D}_\ell$ extensions K_ℓ such that $\deg \Delta_{K_2} = 2$; that is, such that K_2 is a genus-0 extension K_2 with constant field $\mathbb{F}_q$. Such extensions are obtained by adjoining to K a square root of a polynomial f that is either linear or quadratic with nonzero discriminant; the polynomial is determined by the extension, if we require that its leading coefficient be either 1 or a fixed nonsquare element of $\mathbb{F}_q$. These extensions are of two different types: The ramification points of the cover can either be rational over $\mathbb{F}_q$, or not. There are $q^2 + q$ extensions of the first type, and $q^2 - q$ of the second.

Since $\deg \Delta_{K_2} = 2$, we must have $\deg M = 1$, so M consists of a degree-1 place of K that splits in K_2. The number of such places is equal to half of the number of degree-1 places of K_2 that are not ramified in K_2/K; this is equal to $(q-1)/2$ for extensions with rational ramification, and $(q+1)/2$ for extensions without rational ramification.

As in the case where K_2 was a constant field extension, the Kummer extension $K_{2\ell}/K_2$ is completely determined by the divisor M. Thus, the number of K_ℓ whose quadratic resolvents are genus-0 extensions of K with rational ramification is equal to

$$(q^2 + q) \cdot \frac{q-1}{2} = \frac{q^3 - q}{2},$$

while the number whose quadratic resolvents are genus-0 extensions of K without rational ramification is equal to

$$(q^2 - q) \cdot \frac{q+1}{2} = \frac{q^3 - q}{2}.$$

We thus see that $N_2 = q^3 - q$.

Finally, we count the number of $\mathcal{D}_\ell$ extensions K_ℓ such that $\Delta_{K_2} = 4$; that is, such that K_2 is a genus-1 extension of K. In this case, the degree of M is 0, so that $K_{2\ell}$ is an unramified degree-ℓ Galois extension of K_2.

Let E be an elliptic curve over $\mathbb{F}_q$ and let K_2 be its function field. Let $\operatorname{Aut} E$ (respectively, $\operatorname{Aut}' E$) denote the automorphism group of E in the category of elliptic curves (respectively, in the category of curves). Then

$$\operatorname{Aut}' E \cong E(\mathbb{F}_q) \rtimes \operatorname{Aut} E,$$

where the subgroup $E(\mathbb{F}_q)$ acts on E by translation [29, Proposition X.5.1].

Up to twists, the unramified degree-ℓ Galois extensions of K_2 (with constant field $\mathbb{F}_q$) are in bijection with the index-ℓ subgroups of $E(\mathbb{F}_q)$ (see [28, §VI.6]); by duality, the number of such families of twists is equal to the number of order-ℓ subgroups of $E(\mathbb{F}_q)$, which is equal to

$$\frac{\#E[\ell](\mathbb{F}_q)-1}{\ell-1}.$$

Exactly one twist $z^\ell = f$ in each family has the property that $N_{K_2/K}(f) \in (K^\times)^\ell$. Thus,

$$N_4 = \sum_{E/\mathbb{F}_q} \frac{\#E[\ell](\mathbb{F}_q)-1}{\ell-1} \cdot \#\{\text{degree-2 maps } E \to \mathbb{P}^1 \text{ up to isomorphism}\}; \quad (7)$$

here we say that two degree-2 maps $\pi_1, \pi_2 \colon E \to \mathbb{P}^1$ are isomorphic if there is an $\alpha \in \mathrm{Aut}' E$ such that $\pi_2 = \pi_1 \alpha$.

Given an $E/\mathbb{F}_q$, we will count the number of isomorphism classes of degree-2 maps $E \to \mathbb{P}^1$ in two steps. First, we count the number of $(\mathrm{Aut}' E)$-orbits of index-2 genus-0 subfields of the function field K_2 of E. Then, for each orbit, we fix an orbit representative L and we count the number of isomorphism classes of degree-2 maps $E \to \mathbb{P}^1$ that send the function field K of $\mathbb{P}^1$ to L.

Every index-2 genus-0 subfield of K_2 is the fixed field of an involution in $\mathrm{Aut}' E$ that induces -1 on the Jacobian of E. The involutions that induce -1 on the Jacobian are the maps i_Q, for $Q \in E(\mathbb{F}_q)$, defined by $i_Q(P) = Q - P$. The fixed fields of two such involutions i_{Q_1} and i_{Q_2} lie in the same $(\mathrm{Aut}' E)$-orbit if and only if i_{Q_1} and i_{Q_2} are conjugate in $\mathrm{Aut}' E$; this translates into the condition that $Q_2 - \alpha(Q_1) \in 2E(\mathbb{F}_q)$ for some $\alpha \in \mathrm{Aut}\, E$. Thus, the $(\mathrm{Aut}' E)$-orbits of index-2 genus-0 subfields L are in bijection with the orbits of $E(\mathbb{F}_q)/2E(\mathbb{F}_q)$ under the action of $\mathrm{Aut}\, E$.

Let L be an index-2 genus-0 subfield of K_2, corresponding to an involution i_Q. Let $\mathcal{S}_L$ denote the set of isomorphism classes of degree-2 maps $E \to \mathbb{P}^1$ that send the function field K of $\mathbb{P}^1$ to the subfield L of K_2, and let π be one such map. The group $\mathrm{PGL}(2,q)$ acts transitively on $\mathcal{S}_L$, so to compute $\#\mathcal{S}_L$ it suffices to compute the stabilizer of π. Tracing through the definitions, we see that $\phi \in \mathrm{PGL}(2,q)$ stabilizes π if and only if there is an automorphism α of E (as a curve) such that $\phi\pi = \pi\alpha$. Furthermore, every automorphism α of E whose induced automorphism of K_2 sends L to itself gives rise to a ϕ that stabilizes the isomorphism class of π; also, two such automorphisms $\alpha_1 \neq \alpha_2$ will give rise to distinct ϕ, unless $\alpha_1^{-1}\alpha_2 = i_Q$. We find that we have

$$\begin{aligned}\#\{\phi \in \mathrm{PGL}(2,q) : \phi \text{ stabilizes } \pi\} &= (1/2)\#\{\alpha \in \mathrm{Aut}' E : \alpha \text{ stabilizes } L\} \\ &= (1/2)\#\{\alpha \in \mathrm{Aut}' E : \alpha \text{ commutes with } i_Q\}.\end{aligned}$$

We check that an element $(P, a) \in E(\mathbb{F}_q) \rtimes \operatorname{Aut} E \cong \operatorname{Aut}' E$ commutes with i_Q if and only if $2P = Q - a(Q)$. This shows that for every element of $\operatorname{Aut} E$ that fixes the image of Q in $E(\mathbb{F}_q)/2E(\mathbb{F}_q)$, there are $\#E(\mathbb{F}_q)[2]$ choices for P that give an element of $\operatorname{Aut} E'$ that commutes with i_Q. In other words, if we let O be the $(\operatorname{Aut} E)$-orbit of Q in $E(\mathbb{F}_q)/2E(\mathbb{F}_q)$, then

$$\#\{\alpha \in \operatorname{Aut}' E : \alpha \text{ commutes with } i_Q\} = \#E(\mathbb{F}_q)[2] \frac{\#\operatorname{Aut} E}{\#O}.$$

Putting this all together, we obtain

$$\begin{aligned}\frac{\#\mathcal{S}_L}{\#\operatorname{PGL}(2,q)} &= \frac{1}{\#\{\phi \in \operatorname{PGL}(2,q) : \phi \text{ stabilizes } \pi\}} \\ &= \frac{2}{\#\{\alpha \in \operatorname{Aut}' E : \alpha \text{ commutes with } i_Q\}} \\ &= \frac{2}{\#\operatorname{Aut} E} \frac{\#O}{\#E(\mathbb{F}_q)[2]}.\end{aligned}$$

The total number of degree-2 maps $E \to \mathbb{P}^1$ (up to isomorphism) is equal to the sum $\sum_L \mathcal{S}_L$, where L ranges over a set of representatives for the $(\operatorname{Aut} E')$-orbits of index-2 genus-0 subfields of K_2. Summing over these L is the same as summing over the $(\operatorname{Aut} E)$-orbits O of $E(\mathbb{F}_q)/2E(\mathbb{F}_q)$. Thus,

$$\begin{aligned}\frac{\#\{\text{degree-2 maps } E \to \mathbb{P}^1\}/\cong}{\#\operatorname{PGL}(2,q)} &= \frac{2}{\#\operatorname{Aut} E} \frac{1}{\#E(\mathbb{F}_q)[2]} \sum_{\text{orbits } O} \#O \\ &= \frac{2}{\#\operatorname{Aut} E} \frac{1}{\#E(\mathbb{F}_q)[2]} \#(E(\mathbb{F}_q)/2E(\mathbb{F}_q)) \\ &= \frac{2}{\#\operatorname{Aut} E}.\end{aligned}$$

Combining this with (7) gives

$$\begin{aligned}\frac{N_4}{\#\operatorname{PGL}(2,q)} &= \sum_{E/\mathbb{F}_q} \frac{\#E[\ell](\mathbb{F}_q) - 1}{\ell - 1} \frac{2}{\#\operatorname{Aut} E} \\ &= \frac{2}{\ell - 1} \sum_{E/\mathbb{F}_q} \sum_{P \in E[\ell](\mathbb{F}_q) \setminus \{O\}} \frac{1}{\#\operatorname{Aut} E} \\ &= \frac{2}{\ell - 1} \sum_{(E,P)/\cong} \frac{1}{\#\operatorname{Aut}(E,P)}. \qquad (8)\end{aligned}$$

Let us explain the notation in the final line. The sum is over isomorphism classes of pairs (E, P), where E is an elliptic curve over $\mathbb{F}_q$ and P is a nonzero ℓ-torsion point in $E(\mathbb{F}_q)$; two such pairs (E_1, P_1) and (E_2, P_2) are isomorphic to one another

when there is an isomorphism $E_1 \to E_2$ that takes P_1 to P_2. The automorphism group of a pair (E, P) consists of the automorphisms of E (as an elliptic curve) that fix P.

From [17, Proposition 3.3 on p. 240 and Proposition 2.3 on p. 233], we find that

$$\sum_{(E,P)/\cong} \frac{1}{\#\operatorname{Aut}(E, P)} = \#X(\mathbb{F}_q) - c,$$

where X is the modular curve $X_1(\ell)$ and c is the number of $\mathbb{F}_q$-rational cusps on X. Since $\mathbb{F}_q$ contains the ℓ-th roots of unity, all of the $\ell - 1$ geometric cusps of X are defined over $\mathbb{F}_q$ [31, Theorem 1.3.1, p. 12], so we have $c = \ell - 1$. Combining this with (8) gives the formula for N_4 stated in the theorem. □

6. Conclusions and future work

It is interesting that the number of degree-ℓ dihedral function fields with a given quadratic resolvent K_2 and discriminant divisor $\Delta = \frac{\ell-1}{2}\Delta_{K_2} + (\ell - 1)M$ behaves quite differently depending on whether or not M is trivial. We see from Theorem 3.10 that when $M = 0$, the number of such fields with a given resolvent field K_2 depends exclusively on the ℓ-rank r of $\operatorname{Pic}^0 K_2$. The probability that the divisor class group of K_2 has a certain ℓ-Sylow subgroup is the focus of various heuristics of Cohen-Lenstra type. These are discussed further in [1], [14], [15], and [21], and directly relate to the number of $\mathcal{D}_\ell$ function fields with $M = 0$.

When $M \neq 0$, the number of degree-ℓ dihedral function fields with given quadratic resolvent field K_2 depends additionally on the cardinality of the set $T_\ell(M)$ defined in Section 3D. The natural map $\operatorname{Div}^0 K_2 \to \operatorname{Pic}^0 K_2/\ell \operatorname{Pic}^0 K_2$ is surjective, and when $\#\operatorname{Supp} M$ is greater than r it is reasonable to expect that the map ρ from Section 3D is also surjective, so that a random element of $\overline{\mathcal{Q}}_\ell(M)$ will lie in the kernel of ρ with probability

$$\frac{1}{\#(\operatorname{Pic}^0 K_2/\ell \operatorname{Pic}^0 K_2)} = \frac{1}{\ell^r}.$$

Now, an element of $\mathcal{Q}_\ell(M)$ lies in $T_\ell(M)$ if and only if it is in the kernel of ρ, so we expect $T_\ell(M)$ to contain about $\#\mathcal{Q}_\ell(M)/\ell^r = (\ell-1)^{\#\operatorname{Supp} M}/\ell^r$ elements. From Theorem 3.10, the number of nonconjugate degree-ℓ dihedral function fields with quadratic resolvent K_2 and with discriminant divisor $\Delta = \frac{\ell-1}{2}D + (\ell-1)M$ is $\#T_\ell(M)\ell^r/(\ell-1)$, which we expect to be approximately $(\ell-1)^{\#\operatorname{Supp} M - 1}$. Note that this is independent of r. When $\#\operatorname{Supp} M$ is sufficiently large, our data seems to support this heuristic.

In the case when $\ell = 3$, our algorithm tabulates all non-Galois cubic function fields up to a given degree bound on the discriminant divisor. Galois cubics are

		Number of cubic extensions				
q	B	Non-Galois	Galois	Total	$q^{B-2}(q^2+q+1)$	Ratio
7	4	2,373	85	2,458	2,793	1.136
	6	117,285	1,093	118,378	136,857	1.156
	8	5,763,093	4,117	5,767,210	6,705,993	1.163
13	4	28,470	274	28,744	30,927	1.076
	6	4,824,534	6,826	4,831,360	5,226,663	1.082
19	4	130,131	571	130,702	137,541	1.052
25	4	390,300	976	391,276	406,875	1.040
31	4	923,025	1,489	924,514	954,273	1.032
37	4	1,873,458	2,110	1,875,568	1,926,183	1.027
43	4	3,417,855	2,839	3,420,694	3,500,157	1.023
49	4	5,763,576	3,676	5,767,252	5,884,851	1.020

Table 2. Cubic function field counts compared to asymptotics, for $q \equiv 1 \bmod 3$ and $B \geq 4$ with $q^{B+1} < 2^{29}$. For the q and B given in the first two columns, we list the number of cubic extensions of $\mathbb{F}_q(x)$ with discriminant divisor of degree at most B, subdivided into the counts of non-Galois and Galois extensions. The sixth column gives an estimate for the total number derived from the asymptotic formula (9), and the seventh column gives the ratio between the estimate and the actual number from column 5.

easy to count, so we can find the total number of cubic extensions of K whose discriminant divisors have degree at most some fixed bound. On the other hand, using a result of Datskovsky and Wright [10, Theorem I.1] we can compute an asymptotic formula for the number of cubic extensions:

$$\lim_{\substack{B\to\infty \\ B \text{ even}}} q^{-B} \sum_{\substack{K_3/K \\ \deg \Delta_{K_3} \leq B}} 1 = \frac{q^3}{(q^2-1)(q-1)\zeta_K(3)} = \frac{q^2+q+1}{q^2}. \tag{9}$$

(Note that the term $2\log q$ in [10, Theorem I.1] should be simply $\log q$.) In Table 2 we compare this asymptotic expression to actual computations. For each q and B listed in the first two columns, the entry in column 5 gives the total number of cubic extensions of $\mathbb{F}_q(x)$ with discriminant divisor of degree at most B, broken down into the number of non-Galois extensions (column 3) and Galois extensions (column 4). Column 6 gives the estimate from (9), and column 7 gives the ratio of the estimate to the actual values.

Note that for $B = 4$ we have explicit formulas for the number of cubic extensions:

By Theorem 5.1, the number of non-Galois extensions is

$$(q^3 - q)\left(\frac{1}{2q+2} + 1 + (q-1)\right) = q^4 - \frac{q^2+q}{2},$$

and it is not hard to show that the number of Galois extensions is $(3q^2+3q+2)/2$, so the total number of cubic extension is q^4+q^2+q+1. It follows that for $B = 4$ the ratio in column 7 is equal to

$$1 + \frac{q^3 - q - 1}{q^4 + q^2 + q + 1}.$$

As in the number field setting, the leading term of the asymptotic expression overestimates the actual number of cubic function fields, which leads us to believe that the secondary term has a negative coefficient. An explicit computation of this secondary term is currently underway by Yongqiang Zhao (private communication, 2012).

One obstacle to generating larger amounts of data is the memory intensive nature of Algorithm 4.4 as written. One could obtain most of the results by instead looking for orbit representatives of $\mathrm{PGL}(2,q)$ acting on elliptic and hyperelliptic curves of genus g by iterating over these curves and computing their invariants. One would then only need to store a representative for each set of invariants. This would largely remove the storage requirements of the algorithm; however, it would also be a slower process as additional time must be spent computing these invariants.

For primes $\ell > 3$, no asymptotic estimates on counts of degree-ℓ function fields are known; it may be possible to obtain such estimates by generalizing the work of [9] or adapting the program of [37] to the case $q \equiv 1 \bmod \ell$ by using results in [13], [15], and [21]. It would be very interesting to see if the "gaps" for the number field setting referred to in Section 1 occur here as well. This is research in progress by the first two authors and several others.

We close by noting that our work is readily extendable to the problem of finding $\mathcal{D}_\ell$ extensions of function fields K other than $\mathbb{F}_q(x)$. This should be reasonably straightforward if one restricts to cases where $(\mathrm{Pic}^0 K)[\ell]$ is trivial. Work is also in progress to extend our algorithms to the cases when $q \not\equiv 1 \bmod \ell$. As in [8], one can construct cyclic function fields by adjoining the ℓ-th roots of unity to K, applying Kummer theory to the extension field, and finally taking a fixed field by the Frobenius automorphism of $\mathbb{F}_{q^{\ell-1}}/\mathbb{F}_q$. We expect that one can combine this technique with the work above to construct $\mathcal{D}_\ell$ function fields with $q \not\equiv 1 \bmod \ell$.

Acknowledgments

The first author is supported by NSERC and AITF of Canada. The second author is supported in part by NSERC of Canada.

References

[1] Jeffrey D. Achter, *The distribution of class groups of function fields*, J. Pure Appl. Algebra **204** (2006), no. 2, 316–333. MR 2006h:11132

[2] Manjul Bhargava, *The density of discriminants of quartic rings and fields*, Ann. of Math. (2) **162** (2005), no. 2, 1031–1063. MR 2006m:11163

[3] Manjul Bhargava, Arul Shankar, and Jacob Tsimerman, *On the Davenport-Heilbronn theorems and second order terms*, Invent. Math. **193** (2013), no. 2, 439–499. MR 3090184

[4] Rajendra Bhatia, Arup Pal, G. Rangarajan, V. Srinivas, and M. Vanninathan (eds.), *Proceedings of the International Congress of Mathematicians* (*Hyderabad*, 2010), vol. 2, New Delhi, Hindustan Book Agency, 2010. MR 2012d:00009

[5] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system*, *I*: *The user language*, J. Symbolic Comput. **24** (1997), no. 3-4, 235–265. MR 1484478

[6] H. Cohen, *Constructing and counting number fields*, in Li [20], 2002, pp. 129–138. http://www.mathunion.org/ICM/ICM2002.2/Main/icm2002.2.0129.0138.ocr.pdf MR 2003m:11186

[7] Henri Cohen, *Advanced topics in computational number theory*, Graduate Texts in Mathematics, no. 193, Springer, New York, 2000. MR 2000k:11144

[8] Henri Cohen, Francisco Diaz y Diaz, and Michel Olivier, *On the density of discriminants of cyclic extensions of prime degree*, J. Reine Angew. Math. **550** (2002), 169–209. MR 2004a:11115

[9] Henri Cohen and Anna Morra, *Counting cubic extensions with given quadratic resolvent*, J. Algebra **325** (2011), 461–478. MR 2012b:11168

[10] Boris Datskovsky and David J. Wright, *Density of discriminants of cubic extensions*, J. Reine Angew. Math. **386** (1988), 116–138. MR 90b:11112

[11] H. Davenport and H. Heilbronn, *On the density of discriminants of cubic fields*, *II*, Proc. Roy. Soc. London Ser. A **322** (1971), no. 1551, 405–420. MR 58 #10816

[12] Jean-Marie De Koninck and Claude Levesque (eds.), *Théorie des nombres*: *Proceedings of the International Conference held at the Université Laval*, *Quebec*, *July* 5–18, 1987, Berlin, de Gruyter, 1989. MR 90f:11002

[13] Jordan S. Ellenberg, Akshay Venkatesh, and Craig Westerland, *Homological stability for Hurwitz spaces and the Cohen-Lenstra conjecture over function fields*, 2009. arXiv 0912.0325v2 [math.NT]

[14] Eduardo Friedman and Lawrence C. Washington, *On the distribution of divisor class groups of curves over a finite field*, in De Koninck and Levesque [12], 1989, pp. 227–239. MR 91e:11138

[15] Derek Garton, *Random matrices and Cohen-Lenstra statistics for global fields with roots of unity*, Ph.D. thesis, University of Wisconsin — Madison, 2012. http://digital.library.wisc.edu/1711.dl/VB2JDB7JUO4RZ8N

[16] David Goss, *Basic structures of function field arithmetic*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 35, Springer, Berlin, 1996. MR 97i:11062

[17] Everett W. Howe, *On the group orders of elliptic curves over finite fields*, Compositio Math. **85** (1993), no. 2, 229–247. MR 94a:11089

[18] M. J. Jacobson, Jr., Y. Lee, R. Scheidler, and H. C. Williams, *Construction of all cubic function fields of a given square-free discriminant*, preprint, 2012.

[19] J. Jones, *Number fields* (searchable database), 2012. http://hobbes.la.asu.edu/NFDB

[20] Tatsien Li (ed.), *Proceedings of the International Congress of Mathematicians* (*Beijing*, 2002), vol. 2, Beijing, Higher Education Press, 2002. MR 2003i:00010a

[21] Gunter Malle, *On the distribution of class groups of number fields*, Experiment. Math. **19** (2010), no. 4, 465–474. MR 2011m:11224

[22] Michael E. Pohst, *On computing non-Galois cubic global function fields of prescribed discriminant in characteristic* > 3, Publ. Math. Debrecen **79** (2011), no. 3-4, 611–621. MR 2907993

[23] David P. Roberts, *Density of cubic field discriminants*, Math. Comp. **70** (2001), no. 236, 1699–1705. MR 2002e:11142

[24] Michael Rosen, *Number theory in function fields*, Graduate Texts in Mathematics, no. 210, Springer, New York, 2002. MR 2003d:11171

[25] Pieter Rozenhart, Michael Jacobson, and Renate Scheidler, *Tabulation of cubic function fields via polynomial binary cubic forms*, Math. Comp. **81** (2012), no. 280, 2335–2359. MR 2945159

[26] Pieter Rozenhart and Renate Scheidler, *Tabulation of cubic function fields with imaginary and unusual Hessian*, in van der Poorten and Stein [36], 2008, pp. 357–370. MR 2009m:11213

[27] Jean-Pierre Serre, *Linear representations of finite groups*, Graduate Texts in Mathematics, no. 42, Springer, New York, 1977. MR 56 #8675

[28] ———, *Algebraic groups and class fields*, Graduate Texts in Mathematics, no. 117, Springer, New York, 1988. MR 88i:14041

[29] Joseph H. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, no. 106, Springer, New York, 1986. MR 87g:11070

[30] Neil J. A. Sloane, *The on-line encyclopedia of integer sequences*, 2012. http://oeis.org

[31] Glenn Stevens, *Arithmetic on modular curves*, Progress in Mathematics, no. 20, Birkhäuser, Boston, 1982. MR 87b:11050

[32] Henning Stichtenoth, *Algebraic function fields and codes*, Universitext, Springer, Berlin, 1993. MR 94k:14016

[33] Andrew V. Sutherland, *Constructing elliptic curves over finite fields with prescribed torsion*, Math. Comp. **81** (2012), no. 278, 1131–1147. MR 2012m:11079

[34] ———, *Defining equations for* $X_1(N)$, 2012. http://math.mit.edu/∼drew/X1_altcurves.html

[35] Takashi Taniguchi and Frank Thorne, *Secondary terms in counting functions for cubic fields*, 2011. arXiv 1102.2914v1 [math.NT]

[36] Alfred J. van der Poorten and Andreas Stein (eds.), *Algorithmic number theory*: *Proceedings of the* 8*th International Symposium* (*ANTS-VIII*) *held in Banff*, *AB*, *May* 17–22, 2008, Lecture Notes in Computer Science, no. 5011, Berlin, Springer, 2008. MR 2009h:11002

[37] Akshay Venkatesh and Jordan S. Ellenberg, *Statistics of number fields and function fields*, in Bhatia et al. [4], 2010, pp. 383–402. http://www.mathunion.org/ICM/ICM2010.2/Main/icm2010.2.0383.0402.pdf MR 2012h:11160

[38] Gabriel Daniel Villa Salvador, *Topics in the theory of algebraic function fields*, Mathematics: Theory & Applications, Birkhäuser, Boston, 2006. MR 2007i:11002

[39] David J. Wright, *Distribution of discriminants of abelian extensions*, Proc. London Math. Soc. (3) **58** (1989), no. 1, 17–50. MR 90b:11115

COLIN WEIR: colin_weir@sfu.ca
Department of Mathematics, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

RENATE SCHEIDLER: rscheidl@ucalgary.ca
Department of Mathematics and Statistics, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4, Canada

EVERETT W. HOWE: however@alumni.caltech.edu
Center for Communications Research, 4320 Westerra Court, San Diego, CA 92121-1969, United States

THE OPEN BOOK SERIES 1

Tenth Algorithmic Number Theory Symposium

The Algorithmic Number Theory Symposium (ANTS), held biennially since 1994, is the premier international forum for research in computational number theory. ANTS is devoted to algorithmic aspects of number theory, including elementary, algebraic, and analytic number theory, the geometry of numbers, arithmetic algebraic geometry, the theory of finite fields, and cryptography.

This volume is the proceedings of the tenth ANTS meeting, held July 9–13, 2012, at the University of California, San Diego. It includes revised and edited versions of the 25 refereed papers presented at the conference, together with extended abstracts of two of the five invited talks.

TABLE OF CONTENTS