Explicit Descent on Elliptic Curves

by Thomas Womack, MMath

Thesis submitted to The University of Nottingham for
the degree of Doctor of Philosophy, July 2003

Contents

1 Introduction **6**

1.1 Minimisation and reduction 8
1.2 Selmer groups, and Ш 9
1.3 n-coverings 10
1.3.1 Adding points with Riemann-Roch 11
1.4 Generalities about intersections of two quadrics 13
1.4.1 Some notation 13
1.4.2 Constructing intersections with given points 15
1.4.3 The invariants 15
1.5 Obtaining S_n: two approaches to descent 16
1.6 Some notation for algebras 18
1.7 Checking local solubility 19

2 Four-descent **21**

2.1 Outline of the argument 21
2.2 The algebraic stage 22
2.2.1 Recap of the two-descent for elliptic curves 22
2.2.2 The action of E on elements of S_2 24
2.2.3 What if we adjoin a root? 26
2.2.4 The Main Theorem for two-coverings 31
2.3 Putting it together – explicit four-descent over $\mathbb{Q}$ 35
2.3.1 From algebraic four-descendents to four-coverings 37
2.3.2 Optimising the leading coefficient 38
2.4 Checking p-adic solvability by local images 40
2.4.1 The value of $[E(\mathbb{Q}_p) : 2E(\mathbb{Q}_p)]$ 41
2.4.2 The size of $\operatorname{im}\mathcal{C}(\mathbb{Q}_p)$ in $L_p^\times/\mathbb{Q}_p^\times(L_p^\times)^2$ 43
2.4.3 Checking at the infinite prime 44
2.4.4 Checking at finite primes 44
2.5 Minimisation of four-coverings 46
2.5.1 Constructive proof of theorem 2.5.1 47

2.5.2 $r_{\text{gen}} = 0$ . . . 49
2.5.3 $r_{\text{gen}} = 1$ . . . 49
2.5.4 $r_{\text{gen}} = 2$ . . . 50
2.5.5 $r_{\text{gen}} = 3$ . . . 52
2.6 Reducing four-coverings . . . 55
2.6.1 Two forms of naïve reduction . . . 55
2.6.2 Stoll's reduction method . . . 56
2.6.3 Canonical forms for four-coverings . . . 57
2.6.4 Practical notes . . . 58
2.7 Finding rational points on a four-covering . . . 58
2.7.1 Parametrizing a $\mathbb{P}^3$ quadric via a rational point . . . 60
2.7.2 Some invariants of the ternary quartic . . . 61
2.7.3 The `LinearMinimise` approach . . . 62
2.8 Sieving for points on homogenous ternary polynomials . . . 63
2.9 p-adic Elkies search on a pair of quaternary quadratics . . . 65
2.10 Working with several descendents . . . 67
2.10.1 How large is $\mathfrak{D}_4$? . . . 68
2.10.2 Some worked examples . . . 69
2.10.3 How long does this all take? . . . 70
2.10.4 Ridiculously large generators . . . 71

3 Survey work on the distribution of ranks and of Tate-Shaferevich groups 73
3.1 Tools and implementations . . . 74
3.1.1 The `ecsieve` algorithm . . . 74
3.1.2 Analytic ranks . . . 75
3.2 Curves with many integral points . . . 76
3.2.1 Experiments performed . . . 76
3.2.2 How are the point counts distributed? . . . 77
3.2.3 What effect does changing I_x have? . . . 77
3.2.4 How are conductors distributed for curves with many integral points? . . . 79
3.2.5 The distribution of conductor with rank . . . 79
3.2.6 Distribution of point count with rank . . . 80
3.2.7 Smallest-observed conductors for given ranks . . . 80
3.2.8 Other work in this field . . . 83
3.2.9 What about regulators? . . . 83
3.3 Curves with non-trivial torsion groups . . . 85
3.4 Mordell curves . . . 87
3.4.1 Finding large-rank examples with `ecsieve` . . . 87

3.4.2 Deeper investigations with two-descents 88
3.4.3 Exploring the inexact realm using four-descents 89
3.5 A Mordell-like family, avoiding the origin 91
3.5.1 Results from four-descent 92
3.6 How useful has four-descent been in practice? 94
3.7 An unexpected lack of independence 96

4 Invariant-theory-based 2-descent over $\mathbb{Z}[i]$ 98
4.1 The model – reduction of positive definite quadratic forms 98
4.2 Hermitian forms: definitions, group actions and contravariant maps 100
4.3 Bounds on reduced Hermitian positive definite quadratic forms . 102
4.3.1 Bounds on A, B, C for $\psi([A, B, C]) \in \mathfrak{F}$ 105
4.3.2 Computing the reduced form of an H.p.d.q.f. 105
4.4 $\mathrm{SL}_2(\mathcal{O}_K)$-reducing higher-degree forms with complex coefficients 106
4.4.1 Bounding coefficients for reduced polynomials 107
4.5 Explicit bounds on the leading coefficient, in the cubic and quartic cases 108
4.5.1 Cubics 108
4.5.2 Quartics 109
4.6 Applications to descent 110

5 Directions for further work 112

A Invariants and covariants of a binary quartic form 114
A.1 The map $\theta_P : \mathcal{C} \to E$ 115

B The invariant map up from a four-covering 116
B.1 Following the maps down 117

C Computer programs used in and developed during this work 119
C.1 Sample data sets 120

Abstract

This thesis presents three pieces of work. The first gives full details of a practical technique for performing second 2-descent on elliptic curves over $\mathbb{Q}$ without 2-torsion, by associating with the two-descendent curve $\mathcal{C} : y^2 = f(x)$ an algebra $L = \mathbb{Q}[x]/(f(x))$ such that the four-coverings arising as two-coverings of $\mathcal{C}$ correspond to elements of $M = L^\times/\mathbb{Q}^\times(L^\times)^2$, finding the coset $\mathfrak{D}_4^{\mathrm{alg}}$ of M in which such four-coverings lie, checking which elements of $\mathfrak{D}_4^{\mathrm{alg}}$ give everywhere-locally-solvable four-coverings, constructing these four-coverings as explicit intersections of quadrics, minimising and reducing the intersections to give equivalent four-coverings of much more convenient form, and finally presenting two ways to find points efficiently on such four-coverings; the algebraic part also works over number fields $K \neq \mathbb{Q}$.

The second investigates the distribution of the ranks and the Tate-Shaferevich groups of elliptic curves from three families (curves with many small integral points, Mordell curves, and curves of the form $y^2 = x^3 + 17x + K$) using the four-descent of the previous part, and a novel algorithm for finding curves with many small integral points. It presents the Mordell curves $y^2 = x^3 \pm K$ with smallest K and rank six, and the curves of smallest conductor currently known for ranks five through nine, several of which are new discoveries; it presents previously-unobtainable experimental results on the distribution of the structure of Ш.

The third section uses some refinements of nineteenth-century results on Hermitian quadratic forms to overcome the one remaining obstacle to an invariant-theory-based 2-descent algorithm over the Gaussian integers $\mathbb{Z}[i]$, and to provide a reduction algorithm which may be useful for improving the algebraic-number-theory-based 2-descent over $\mathbb{Z}[i]$.

Chapter 1

Introduction

There are three largely independent chapters to this thesis, of which the most important is the first, which presents work which has led to a practical implementation of second 2-descent on elliptic curves without 2-torsion. The second introduces a new algorithm for finding curves with many small integral points, and presents the results of much experimental work, using this algorithm, second 2-descent and other approaches, on the distribution of ranks of elliptic curves and of the structure of Tate-Shaferevich groups. The third uses some refinements of nineteenth-century results on Hermitian quadratic forms to overcome the one remaining obstacle to an invariant-theory-based 2-descent algorithm over the Gaussian integers $\mathbb{Z}[i]$, and to provide a reduction algorithm which may be useful for improving the algebraic-number-theory-based 2-descent over $\mathbb{Z}[i]$.

Let E be an elliptic curve, and let K be some number field. Traditionally, the problem of finding $E(K)$ has been split into two parts: computing the rank of the Mordell-Weil group, and finding an explicit set of points to generate the group. And, traditionally, K has been taken to equal $\mathbb{Q}$; progress on any field of degree > 2 was very limited until the work of Simon [66] in mid-2000, and even with that work there is often great trouble in finding explicit generators.

Methods based on the L-series of the elliptic curve have had some significant experimental success for the first part, at least for small ($< 10^{15}$ on contemporary computers) conductors, and Kolyvagin, Rubin and others [41] have proved that the analytic and actual ranks of the Mordell-Weil group are equal for $r = 0, 1$; for higher ranks, however, the methods rely on still-conjectural parts of the Birch–Swinnerton-Dyer (B–S-D) conjecture.

In the rank one case, an analytic method based on Heegner points [28] may be used to find the generator. For a curve of conductor N and a generator of height h, the computation required is $O(Nh)$; a number of terms proportional

to N must be calculated to $O(\exp(-h))$ precision. For $N < 10^6$ this technique is reasonably practical.

But, if the rank is ≥ 2 or the conductor at all large, none of this analytic approach will help for the second task of actually finding rational points. And in any case it would be nice to have an algorithm which does not require B–S-D, since there seems no hope of an imminent proof of their conjecture.

Searching for solutions of $y^2 = f(x)$ using a large number of quadratic sieves modulo prime powers, and (at least when f is of even degree) p-adic arguments to restrict the prime factors of the denominator, can be made exceptionally fast; on $y^2 = f(x)$ with f cubic, where the points are all of the form $\left(x/z^2, y/z^3\right)$, rational points with naïve height up to 15 (that is, four or five digits in x and z) can be found in 15 seconds or so. But the direct search still takes time exponential in the height, so is impractical for naïve heights much above 20.

Fortunately, for each prime p we have an embedding $E(\mathbb{Q})/pE(\mathbb{Q}) \to S_p(E)$ into the p-Selmer group. $S_p(E)$ is a p-group, consisting of equivalence classes of principal homogenous spaces $\mathcal{C}$ (genus-1 curves birationally isomorphic to E), each equipped with a degree-p rational map $j_{\mathcal{C}} : \mathcal{C} \to E$, and such that each $\mathcal{C}$ possesses points over all local fields $\mathbb{Q}_p$; a process known as p-descent allows us explicitly to find a representative for each of its elements. The elements of $S_p(E)$ all have points over $\mathbb{Q}_\ell$ for all ℓ; the image of $E(\mathbb{Q})$ in $S_p(E)$ consists of those $\mathcal{C}$ which have a point over $\mathbb{Q}$ itself.

Any curve in S_p which has a rational point P corresponds to a coset $j_{\mathcal{C}}(P) + pE(\mathbb{Q})$; the curves on which there are no rational points represent elements of the Tate-Shaferevich group $Ш[p](E)$. The difficulty, of course, is in distinguishing curves without a rational point from curves with a rational point beyond the reach of our searching.

The 2-descent is the only one currently widely used in practice, though work of Stoll, Cremona, Fisher and others [26] is making 3-descent possible. Unfortunately, $Ш[2]$ is often non-empty: of the 2.4 million curves examined in section 3.4.2, more than $330,000$ appear to have $Ш[2] \neq \{1\}$.

This thesis revolves around a second 2-descent, which we call a four-descent. For each $\mathcal{C} \in S_2(E)$ on which we failed to find a rational point, we compute a group $\mathfrak{D}_4$ consisting of curves $\mathcal{H}$ with degree-2 rational maps $\psi_{\mathcal{H}} : \mathcal{H} \to \mathcal{C}$, such that each rational point on $\mathcal{C}$ is the image of a rational point on precisely one of the $\mathcal{H}$. If empty, this indicates that $\mathcal{C}$ represented an element of $Ш[2]$ and so had no rational points; if not, we can search for points on each of its elements. If they do not exist, we have a representative for an element of $Ш[4]$; if we find one, we can lift it to a much larger point $j_{\mathcal{C}}(\psi_{\mathcal{H}}(P))$ on the original elliptic curve.

Experiment suggests that $Ш[4] = Ш[2]$ for about 85% of curves with $Ш[2] \neq$

{1}, and that, on curves with rational points, four-descent will find points of canonical height up to about 140.

1.1 Minimisation and reduction

The terms "minimisation" and "reduction" describe two allied tasks, both of which are extremely important for making descent calculations practical. Let $F^{(1)}$ and $F^{(2)}$ be spaces equipped with a left G-action. A *covariant function* $f : F^{(1)} \to F^{(2)}$ is a function with, for all $g \in G$, $g \cdot f(A) = f(g \cdot A)$; a *contravariant function* has $g \cdot f(A) = f(g^{-1} \cdot A)\ \forall g \in G$.

Consider a space F of forms in n variables, on which there is an action of some $\mathrm{GL}_n(\mathbb{Q})$ – for example, bivariate homogeneous polynomials, where $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ acts by $M \cdot f(x, y) = f(ax+by, cx+dy)$. We often find functions on the forms which, under that action, are scaled by some power of the determinant of M – in the example case, among them is the discriminant Δ, where

$$\Delta\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot f(x, y)\right) = (\det M)^{n(n-1)}\Delta(f).$$

On a pencil of forms, with generators say $\langle g_1, g_2, \ldots g_r \rangle$, there is an additional action of $\mathrm{GL}_r(\mathbb{Z})$ which replaces the generators with linear combinations of them.

Suppose we wish to consider only integral forms (that is, forms with coefficients taken from the ring of integers of some number field K); naturally we can scale any form with K-rational coefficients until its coefficients become integral, but this will increase the invariants. *Minimisation* is a $\mathfrak{p}$-adic process in which, by transformations in $\mathrm{GL}_n(K)$, we try to make the $\mathfrak{p}$-valuations of certain covariant functions as small as possible whilst keeping the form $\mathfrak{p}$-integral. The covariants for an integral form will be integers, so the minimum valuation is well-defined at any prime $\mathfrak{p}$. Over fields with class number 1, it is possible to minimise at $\mathfrak{p}$ using only transformations with determinant a power of π (where $\pi\mathcal{O}_K = \mathfrak{p}$), meaning that there exists a *globally minimal* version of a form, whose invariant has minimal valuation at **all** primes.

Similarly, for any n-ary form with integer coefficients, we can pick any positive-definite function of the coefficients, and then consider how this function's value behaves as you move from a form f around the set of forms equivalent to it under the action of $\mathrm{SL}_n(\mathbb{Z})$: again, this set will have a minimum, and *reduction* consists of picking a function such that finding this minimum is practical, and then finding the minimum. The idea here is to make the coefficients

of the form smaller: the hope is that this also makes the coordinates of points lying on the hypersurface cut out by the form smaller, and so the points easier to find.

For at least three kinds of objects, a good reduction theory exists already: lattices (by the LLL algorithm, though there is a problem with uniqueness in dimensions higher than four), points in the upper half-plane $\mathfrak{H}^2$ under Möbius transforms by $\mathrm{SL}_2(\mathbb{Z})$, and points in upper half-space $\mathfrak{H}^3$ under Möbius transforms by $\mathrm{SL}_2(\mathcal{O}_K)$ for $K = \mathbb{Z}[i]$ or $K = \mathbb{Z}[\sqrt{-2}]$. So a common approach to reduction is to find a covariant or contravariant function ψ from the set of objects to be reduced to one of these sets of objects with a known reduction theory, and define X as reduced iff $\psi(X)$ is. Given the covariance, if we can find M such that $M \cdot \psi(X)$ is the reduced form of $\psi(X)$, then $\psi(M \cdot X) = M \cdot \psi(X)$ and we can use $M \cdot X$ as the reduced form of X; if ψ were contravariant, we would use $M^{-1} \cdot X$.

1.2 Selmer groups, and Ш

Let $G_K = \mathrm{Gal}(K/\mathbb{Q})$; noting that $H^0(G_K, E) = E(K)$, we form a long exact sequence of Galois cohomology from the trivial diagram

$$0 \longrightarrow E[n] \longrightarrow E \xrightarrow{\times n} E \longrightarrow 0$$

to obtain the Kummer sequence

$$0 \longrightarrow E(K)/nE(K) \longrightarrow H^1(G_K, E[n]) \longrightarrow H^1(G_K, E)[n] \longrightarrow 0$$

The diagram can be specialised at the field K and at all its completions $K_\mathfrak{p}$, giving a diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E(K)/nE(K) & \longrightarrow & H^1(G_K, E[n]) & \longrightarrow & H^1(G_K, E)[n] & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \prod_\mathfrak{p} E(K_\mathfrak{p})/nE(K_\mathfrak{p}) & \longrightarrow & \prod_\mathfrak{p} H^1(G_{K_\mathfrak{p}}, E[n](K_\mathfrak{p})) & \longrightarrow & \prod_\mathfrak{p} H^1(G_{K_\mathfrak{p}}, E(K_\mathfrak{p}))[n] & \longrightarrow & 0
\end{array}$$

The Selmer group $S_n(E, K)$ is then the set of elements of $H^1(G_K, E[n])$ whose image in $\prod_\mathfrak{p} H^1(G_{K_\mathfrak{p}}, E[n](K_\mathfrak{p}))$ arises from an element of $\prod_\mathfrak{p} E(K_\mathfrak{p})/nE(K_\mathfrak{p})$; that is, the set of everywhere locally soluble elements. Going round the diagram,

we have $E(K)/nE(K) \leq S_n(E,K)$; define

$$Ш(K)[n] = S_n(E,K)/(E(K)/nE(K)).$$

Kloosterman [44] has recently proved that, for K sufficiently large (of degree $O(p^4)$), there exist curves defined over K such that $Ш[p]$ is arbitrarily large; for $p = 2, 3, 5$ it is known ([6], [34]) that this holds for $K = \mathbb{Q}$.

The Birch–Swinnerton-Dyer conjecture over $\mathbb{Q}$ asserts that

$$\|Ш\| = \frac{L^{(r)}(E,1)\,\|E_{\mathrm{tors}}\|^2}{\Omega R_E \prod c_p}$$

where $r = \operatorname{rank} E$, E_{tors} is the subgroup of E consisting of elements of finite order, Ω the real period, c_p the local Tamagawa numbers, and R_E the determinant of the height-pairing matrix for the generators of E. The quantity on the right-hand side of this equation can be calculated explicitly, and without inordinate difficulty provided that E has rank zero – for rank one we have to find a generator, and for rank two and above we need to ensure that we find a set of generators which give $E(\mathbb{Q})$ rather than some finite-index subgroup (Prickett [52] is working on this issue) – and will be called $Ш_{\mathrm{anal}}$; see section 3.1.2 for some practical details on the computation.

The following groups will often be mentioned in what is to follow:

Definition 1.2.1. $Ш[2^\infty]$ is the set of elements of Ш of order a power of two. $Ш[2^n]$ is the set of elements of $Ш[2^\infty]$ of order dividing 2^n (so $Ш[2] \subset Ш[4] \subset Ш[8] \ldots$), and ${}_{2^n}Ш$ is the set of elements of $Ш[2^\infty]$ of exact order 2^n.

1.3 n-coverings

The elements of $S_n(E)$ are called *n-coverings of E*, or *elliptic normal n-ics*; this last notation is particularly used in the $n = 5$ case intensely studied by Fisher [36], and often indicates that the objects are considered geometrically rather than arithmetically. I normally call them coverings, but use "descendents" when I wish to emphasise that they are being constructed by a descent process.

The shape of the n-coverings is given by the Riemann-Roch theorem; 2-coverings are of the form $y^2 = g(x,z)$ for g a integral binary quartic form, and 3-coverings $f(x,y,z) = 0$ for f a ternary cubic form. For $n \geq 4$, an n-covering is given by a set of $\frac{1}{2}n(n-3)$ quadrics in $\mathbb{P}^{n-1}$; and, conveniently, for $n = 4$ any pair of quadrics satisfying a simple non-singularity condition corresponds to a 4-covering of some elliptic curve. Syzygies between the invariants of n-coverings give us the maps $j : \mathcal{C} \to E$, so given an n-covering we can in principle compute the syzygy and obtain the map; we call j the syzygy map on $\mathcal{C}$.

For $n \geq 5$, a random collection of $\frac{1}{2}n(n-3)$ quadrics in $\mathbb{P}^{n-1}$ need not define an n-covering of any elliptic curve: at $n = 5$, we at least know the subspace which corresponds to elliptic curves, thanks to Buchbaumm, Eisenbud and Fisher [10], [36], [37], who showed that the quadrics for a 5-covering arise as the five 4×4 Pfaffians of a 5×5 matrix of linear forms in $\mathbb{P}^4$, and conversely that the Pfaffians of any such matrix indeed form a 5-covering for some elliptic curve.

The paper [55] collects conveniently the Jacobian maps (which indicate which elliptic curve $E_{\mathcal{C}}$ a n-covering $\mathcal{C}$ corresponds to) and the syzygy maps $j_{\mathcal{C}} : \mathcal{C} \to E_{\mathcal{C}}$ for $n = 2, 3, 4$. For $n = 5$ the situation is less good, since the Jacobian map is too complicated to write down as an explicit polynomial (though $E_{\mathcal{C}}$ can be computed quickly for any given $\mathcal{C}$), and the $j_{\mathcal{C}}$ are as yet unknown: all that is known for this case is in [36].

The method of fermionic Fock spaces used by Anderson in [1] extends in principle to arbitrary values of n. However, it involves symbolic differentiation of some rational functions of large degree, over fields with a large number of indeterminants – it starts in $k = \mathbb{Q}(a_0, a_1, \ldots a_m)$ where the a_i are the variables in the parametrisation of an n-covering, works in the function field K of the n-covering, constructs a k-linear derivation on K, defines a function $\wp$ as a sum of ratios of Jacobian-like determinants of matrices of repeated derivatives, and then derives the g_2 and g_3 of the underlying elliptic curve by solving

$$\left(\frac{\wp'}{4}\right)^2 - 4\wp^3 = -g_2\wp - g_3,$$

also obtaining a map j_D from the n-covering to the curve by solving $\wp = x \circ j_D$.

The computation for this approach is a matter of extremely complicated Gröbner-basis calculations, and in [1] the author's computer was not able to contain the general case for $n \geq 3$.

1.3.1 Adding points with Riemann-Roch

The Riemann-Roch theorem allows us to compute explicitly on any genus-one curve V, at least once we have a base point O giving V the structure of an elliptic curve.

Recall that a divisor on the curve – an element $D \in \operatorname{Div} V$ – is a finite sum $\sum \lambda_i P_i$ of points with multiplicities $\lambda_i \in \mathbb{Z}$; a divisor is *positive* if all the $\lambda_i > 0$, and the *degree* of the divisor is $\sum \lambda_i$. Recall also that you can construct the divisor $\operatorname{Div} f$ of a function f on V as the sum of its zeroes (with multiplicity equal to their multiplicity as zeroes) and its poles (with multiplicity equal to the negation of the degree of the pole), necessarily obtaining a divisor of degree

zero; the divisor of a function is called *principal*. The *Jacobian* of V is defined as $\operatorname{Jac} V = \operatorname{Div}^0 V / \operatorname{Princ} V$, the degree-zero divisors quotiented by the principal divisors.

A divisor is *K-rational* if it is invariant under the action of $\operatorname{Gal}(\overline{K}/K)$; we can therefore talk about the K-rational points on the Jacobian of a curve, by replacing "divisors" by "K-rational divisors" in the definition of Jacobian. These K-rational points on the Jacobian form an Abelian group, analogous to the Mordell-Weil group of an elliptic curve.

Given our base point O, the point $P \in V$ corresponds to the degree-zero divisor $(P) - (O)$ in $\operatorname{Jac} V$; so, to compute $R = P + Q$, we would like to find a point $R \in V$ with $(R) - (O) \equiv (P) + (Q) - 2(O)$ in $\operatorname{Jac} V$ – that is, with $(P) + (Q) - (O) = (R) + \operatorname{Div} f$ for some function f on V. Let D be the degree-one divisor $(P) + (Q) - (O)$. The Riemann-Roch theorem tells us that, for V of genus one, the space of functions f on V such that $\operatorname{Div} f + D$ is positive is one-dimensional; the `RiemannRochSpace` function in `magma` can construct an explicit generator g for this space. Now, $\operatorname{Div} g + D$ is a positive degree-one divisor, so must be of the form (R) for some point $R \in V$; this R is the desired sum.

The normal line-and-extra-intersection method of adding points on an elliptic curve E corresponds to the Riemann-Roch approach with the point at infinity as O. Write the group of points on a curve $\mathcal{C}$, taking the point P as the origin, with co-ordinates from a field L, as $[\mathcal{C}(L), P, \oplus_P]$: $\oplus_P$ is the addition law. We will need to examine the interplay of different addition laws in later work; there is a law rather like associativity:

Lemma 1.3.1. Writing $\oplus_0$ and $\oplus_\alpha$ for $\oplus_{P_0}$ and $\oplus_{P_\alpha}$, to avoid redundant subscripts, we have

$$P \oplus_0 (Q \oplus_\alpha R) = (P \oplus_0 Q) \oplus_\alpha R$$

.

Proof. For this proof, I write $[P]$ for the divisor consisting of the point P; the usual notation uses parentheses, but these are very confusing in a context where we also use parentheses to bracket terms.

$$\begin{aligned}[P \oplus_0 (Q \oplus_\alpha R)] &= [P] + [Q \oplus_\alpha R] - [P_0] \\ &= [P] + [Q] + [R] - [P_\alpha] - [P_0] \\ &= [P] + [Q] - [P_0] + [R] - [P_\alpha] \\ &= [P \oplus_0 Q] + [R] - [P_\alpha] \\ &= [(P \oplus_0 Q) \oplus_\alpha R]\end{aligned}$$

□

On any curve $\mathcal{C} : y^2 = f(x)$, where

$$f(x) = \sum_{i=0}^{4} a_i x^i \in K[x]$$

is a quartic polynomial with leading coefficient $a = a_4$, we have a singularity at infinity, consisting of a pair of points defined over $K(\sqrt{a})$. Moving into projective co-ordinates and desingularising by writing $x = X/Z, y = XY/Z^2$, the pair of points are sent to $P_{\infty^+} = (1 : \sqrt{a} : 0)$ and $P_{\infty^-} = (1 : -\sqrt{a} : 0)$. Define $D_\infty = (P_{\infty^+}) + (P_{\infty^-})$ as the divisor of points at infinity; a function $x - x_0$ on $\mathcal{C}$ has divisor

$$((x_0, \sqrt{f(x_0)})) + ((x_0, -\sqrt{f(x_0)})) - D_\infty,$$

and, if we have points $P = (x, y)$ and $P' = (x, -y)$, divisors of the form $(P) + (P')$ will all be equivalent to one another and to D_∞. The map $P \to P'$ defines negation in $[\mathcal{C}, P_{\infty^+}, \oplus_{\infty^+}]$, in a very similar way to the standard definition for elliptic curves. Note that all these divisors are K-rational, since they are sums $P + P^\sigma$ for $\sigma \in \mathrm{Gal}(K[\sqrt{\lambda}]/K)$ for some λ.

1.4 Generalities about intersections of two quadrics

1.4.1 Some notation

Recall from [51] that a four-covering is given by an intersection of two quadrics in $\mathbb{P}^3$. In characteristic not equal to two, a quadric in $\mathbb{P}^3$ may be written as $\mathbf{x}M\mathbf{x}^{\mathrm{T}} = 0$ where $\mathbf{x} = (x_1, x_2, x_3, x_4)$ is the set of variables and M is a symmetric 4×4 matrix. So, in nearly all that follows, we represent the four-covering $\mathbf{x}A\mathbf{x}^{\mathrm{T}} = \mathbf{x}B\mathbf{x}^{\mathrm{T}} = 0$ by the pair $[A, B]$ of symmetric matrices with integer coefficients. For such a pair, define $\sigma(x, z) = \det(Ax + Bz)$ and $\Delta = \mathrm{disc}\,\sigma$; if

Δ is non-zero, $[A, B]$ represents a two-covering of the two-covering $y^2 = \sigma(x, z)$. The discriminant condition ensures that this covering curve is non-singular.

Given this, we use "four-covering" and "intersection of two $\mathbb{P}^3$ quadrics" synonymously in all that follows. If we wish to talk about a four-covering of a specific elliptic curve, we call it a four-descendent of that curve. When describing a four-descent, the elliptic curve at the top will be called E, the set of two-coverings derived from the two-descent on it will be called $\mathfrak{D}_2$, and any individual two-covering we work with will be called $\mathcal{C}$; the set of four-descendents will be $\mathfrak{D}_4$, and its elements $\mathcal{H}$.

We also use $\mathfrak{D}_4$ to refer to the whole space of four-coverings $[A, B]$. $\mathrm{GL}_2(\mathbb{Q})$ and $\mathrm{GL}_4(\mathbb{Q})$ both act on $\mathfrak{D}_4$; $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Q})$ sends $[A, B]$ to $[aA + bB, cA + dB]$, whilst $T \in \mathrm{GL}_4(\mathbb{Q})$ sends $[A, B]$ to $[TAT^{\mathrm{T}}, TBT^{\mathrm{T}}]$. These actions clearly commute: $T(Ax + By)T^{\mathrm{T}} = TAT^{\mathrm{T}}x + TBT^{\mathrm{T}}y$. There is also a trivial action of $\mathbb{Q}$ by multiplying A and B by the same non-zero constant λ, which clearly commutes with the other actions.

However, we do not have a faithful action of $\mathbb{Q} \times \mathrm{GL}_2(\mathbb{Q}) \times \mathrm{GL}_4(\mathbb{Q})$; elements like

$$\left(a^{-1}b^{-2},\ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix},\ \mathrm{diag}(b, b, b, b)\right)$$

act trivially. Taking a quotient by such elements, we are led to the definition

Definition 1.4.1. Let $A_{\mathfrak{D}_4}$ be the set of triples (λ, M_2, M_4) with $\lambda \in \mathbb{Q}^\times$, M_2 a non-singular 2×2 matrix with integral coefficients, and M_4 a non-singular 4×4 matrix with integral coefficients, such that $\det M_2$ is square-free and $\det M_4$ fourth-power free.

Then $A_{\mathfrak{D}_4}$ acts faithfully on $\mathfrak{D}_4$; making the quotient action explicit, we combine elements of $A_{\mathfrak{D}_4}$ by multiplying in $\mathbb{Q} \times \mathrm{GL}_2(\mathbb{Q}) \times \mathrm{GL}_4(\mathbb{Q})$ and then pulling out square scalar factors of $\det M_2$ and fourth-power scalar factors of $\det M_4$ into λ.

$\det(\lambda A + \mu B)$ is of course a homogeneous quartic form, so we can look at its I and J invariants: these may come in useful in the reduction step to provide further conditions for when a curve is impossible to reduce further.

When it is clear we are working at a particular prime p, we write $\tilde{A}$ to indicate the matrix obtained by reducing every element of the matrix A modulo p.

We say that an integral four-covering is minimal at p if $\mathrm{val}_p(\Delta([A, B]))$ is minimal among all integral $[A', B']$ obtained from $[A, B]$ by the action of some element of $A_{\mathfrak{D}_4}$. Notice that, since the action of (λ, M_2, M_4) on $[A, B]$ multiplies $\Delta([A, B])$ by $\lambda^{24} \det(M_2)^{12} \det(M_4)^{12}$, a non-minimal four-covering must have

$\mathrm{val}_p(\Delta([A,B])) \geq 12$.

1.4.2 Constructing intersections with given points

Given a point $P = (x_1, x_2, x_3, x_4)$ in $\mathbb{P}^3(\mathbb{Q})$, to find a quadric

$$Q(z_1, z_2, z_3, z_4) = \sum_{i=1}^{4} \sum_{j=1}^{i} a_{ij} z_i z_j$$

with $Q(P) = 0$ is a matter of solving a single linear equation in the a_{ij}, and there is a dimension-nine subspace of the ten-dimensional a_{ij} space consisting of such quadrics. So constructing an intersection of two quadrics through a given point is a matter of picking two linearly-independent points from such a subspace. This is very useful for providing test data, especially for minimisation where the main theorem applies only to curves with a point over $\mathbb{Q}_p$: assuring a point over $\mathbb{Q}$ guarantees this.

Indeed, given any **eight** rational points $P_1 \ldots P_8$, it is usually possible to find a two-dimensional subspace of the a_{ij} consisting of quadrics passing through all of them; the pair $\mathcal{H} = [A, B]$ of generators for this subspace is then the right shape for a four-covering, and the maps of appendix B give an elliptic curve E and a map $\Psi : \mathcal{H} \to E$. Surprisingly often, computing the determinant of the height-pairing matrix indicates that $\Psi(P_1), \ldots, \Psi(P_8)$ are independent, so E has rank at least eight. Unfortunately, E also nearly always has coefficients so enormous that there is no practical way of deducing anything more about it: its conductor is large enough to make computation of the analytic rank unthinkable, its discriminant so large that neither method of 2-descent is practical, and the X co-ordinates of $\Psi(P_i)$ have naïve heights large enough that a direct search for more points on E is hopeless.

1.4.3 The invariants

For a single homogenous four-variable quadric, given as $\mathbf{x}M\mathbf{x}^{\mathrm{T}} = 0$ for M a symmetric 4×4 matrix, there is a single invariant of degree 4 for the action of $\mathrm{SL}_4(\mathbb{Z})$, which equals $\det M$. Naturally, for a pair of such quadrics $\mathbf{x}M\mathbf{x}^{\mathrm{T}} = 0$, $\mathbf{x}N\mathbf{x}^{\mathrm{T}}$, under the action of $\mathrm{SL}_4(\mathbb{Z}) \times \mathrm{SL}_4(\mathbb{Z})$, we get $\det M$ and $\det N$ as the degree-4 invariants.

Now, if we include also the action of $\mathrm{SL}_2(\mathbb{Z})$ by linear combinations of the matrices M and N, we will get a smaller set of invariants. Computing these directly using the whole 20-coefficient model given by the matrices M and N would require prohibitive amounts of memory and CPU time. However, to survive the $\mathrm{SL}_4(\mathbb{Z})$-part of the action, they will have to be in the ring $\mathbb{Z}[\det M, \det N]$.

In section 3 of the paper [51], using material from [73], it is stated that the basic invariants of a pair $[A, B]$ of homogenous four-variable quadrics are the σ_i defined by $\det(Ax + By) = \sum_{i=0}^{4} \sigma_i x^i y^{4-i}$. Also in that section are given five covariant functions from $\P^3(\mathbb{Q})$ to $\mathbb{Q}$, and a syzygy between them which, when we simplify it by requiring that the point $\mathbf{x}$ lies on the intersection of quadrics, becomes something of the form $\det(AU(\mathbf{x}) + BV(\mathbf{x})) = W(\mathbf{x})^2$ – that is, it provides a mapping from points $\mathbf{x}$ on the intersection of quadrics to points $x = U(\mathbf{x}), y = W(\mathbf{x}), z = V(\mathbf{x})$ on the associated two-covering $y^2 = \det(Ax + Bz)$. The covariants are detailed in appendix B.

1.5 Obtaining S_n: two approaches to descent

There are two approaches to descent. We can work via algebraic number theory, showing that the n-coverings of a given elliptic curve are parametrized by the finitely-many elements with appropriate properties of some algebra and, listing those elements, constructing all the n-coverings. Alternatively, we can notice that an n-covering is given by finitely many parameters, prove using invariant theory that the set of inequivalent reduced, minimal n-coverings of a particular elliptic curve is contained within a bounded region of the parameter space, and search so as to find them all. These approaches are called 'direct' and 'indirect' descent respectively in [30]; I call them 'algebra-based' and 'invariant-theory-based', which are more unwieldy terms but more descriptive.

The reduction and minimisation theories of the coverings are obviously essential for the invariant-theory-based method. They are also required for the algebra-based method, since the constructed n-coverings are often very far from minimal and very far from reduced, and so minimisation and reduction are needed to make the coefficients small enough for it to be feasible to find points on the coverings by exhaustive search.

The invariant-theory-based approach for 2-descents was introduced by Birch and Swinnerton-Dyer in the early sixties in [5]; but with the computers and algorithms of that epoch, it was infeasible to perform the explicit calculations in the class and unit groups of number fields required for the algebra-based approach. Whilst the advent of Buchmann's algorithm [11], as well-described in [15] and as implemented in such packages as `pari` and `magma`, has made the algebra-based approach more practical, and $n = 3$ and higher descents feasible, it was still true at the time of writing that nearly all descents performed are 2-descents, and nearly all of those are invariant-theory-based 2-descents performed with Cremona's program `mwrank` [19], an implementation of the idea in [5] which has been refined over more than a decade and which, by clever sieving techniques, manages to work very quickly on curves with discriminant

up to 10^{14} or so. Less refined versions of Cremona's algorithms are used in `magma`'s rank-determination routines.

For the invariant-theory 2-descent, we note that every $\mathcal{C} : y^2 = f(x)$ with f quartic has an associated elliptic curve $\mathrm{Jac}\,\mathcal{C} : y^2 = x^3 - 27Ix - 27J$ where I and J are the invariants of f (see appendix B for the details). A careful 2- and 3-adic analysis, introduced in [5] with sketched proofs, described in great detail both over $\mathbb{Q}$ and over quadratic number fields in [58], and significantly refined in [71], tells us that, for appropriate definitions of 'minimal', there are only finitely many I and J values (one or two for $\mathbb{Q}$, between one and four for a quadratic number field) which can arise for a minimal quartic associated to a given elliptic curve – the refinement in [71] allows the use of fewer (I, J) pairs for some curves than [5] would have needed. We can then use the reduction theory for that field to find bounds on the possible coefficients of a minimal quartic with given I and J.

The method works over any field for which the bounds can be constructed; the reduction-theory bounds for the coefficients of a quartic with given I and J over $\mathbb{Q}$ date back to Hermite, though they were improved by Birch and Swinnerton-Dyer in [5], and further improved by Cremona in [23]. Real quadratic fields of class number 1 were handled in [58], and section 4.5.2 of this document contains the reduction-theory construction of the bounds for $\mathbb{Z}[i]$.

The algebra-based approach to two-descent over number fields is outlined in section 2.2.1: a good reference is Simon's [66], and he has an implementation, relying on Cohen's work [16] on computing the class and unit groups of number fields defined by relative extensions, as implemented in `pari` [2]. That paper uses a small amount of minimisation and no reduction, and indeed states "I do not claim that these choices will always lead to a Q' with small coefficients".

The increasing number of coefficients defining an n-covering make the invariant-theory approach increasingly impractical. For $n = 3$, the paper [30] constructs a search region for the coefficients of a ternary cubic form with given S and T-invariants which is of volume $|\Delta|^{13/6}$ (where $\Delta = T^2 + 64S^3$), and moreover fails to prove that there are only finitely many pairs of invariants to consider for a given elliptic curve. Their bounds seem fairly weak, since they bound all the coefficients simultaneously rather than allowing the bounds for one coefficient to depend on the value of all earlier ones; simply calculating the last two coefficients by solving the equations giving the S and T invariants would reduce the volume to $|\Delta|^{4/3}$. For $n > 3$ I have seen no references to an invariant-theory approach.

The algebra-based approach for $n = 3$ was made practical in 2003 after several years of work by Stoll, Cremona, O'Neil, Fisher and others (see [26] and [27]).

For prime n, Schaefer, Stoll and others have reduced this task to a distinctly non-trivial problem in algebraic number theory [56]: to begin we have to compute the class and unit group of the p-division field, which is a field extension normally of degree $p^2 - 1$. In practice I do not believe that a general 5- or higher descent has ever been performed, though Tom Fisher has handled the case where two curves are related by an n-isogeny and one of them has a rational n-torsion point, by an approach needing no field extensions.

1.6 Some notation for algebras

We will quite often want to take a square-free monic polynomial $f \in K[x]$ and consider the algebra $L = K[x]/(f)$. If f is irreducible, this is a number field, and we have the standard norm function $N_{L/K}$. If not, it is a direct sum $L_1 \oplus \ldots \oplus L_r$ of number fields whose degrees sum to $\deg f$; we write its elements as $[x_1, \ldots, x_r]$ with $x_i \in L_i$, and define

$$N_{L/K}([x_1, \ldots, x_n]) = \prod_{i=1}^{n} N_{L_i/K}(x_i).$$

If f is irreducible, let α be one of its roots in L; if not, for each i let α_i be one of its roots in L_i. We will consider the embedding $K \to L$ which sends x to $[x - \alpha_1, \ldots, x - \alpha_r]$, and write it as $x - \alpha$ even if L is not just a number field generated by α; for example, if $f(x) = (x^2 - 3)(x - 4)(x - 5)$, we have $L = \mathbb{Q}[\sqrt{3}] \oplus \mathbb{Q} \oplus \mathbb{Q}$, with $\alpha_1 = \sqrt{3}$, $\alpha_2 = 4$, $\alpha_3 = 5$, and $6 - \alpha$ would be $[6 - \sqrt{3}, 2, 1]$. Since we often want the image of $x - \alpha$ in $L^\times/(L^\times)^2$, we will need to modify some components if $x = \alpha_i$; the precise approach needed is given in the theorems of section 2.2.4, but the idea is to replace $\alpha_i - \alpha_i$ with $f'(\alpha_i)$ if it should happen to occur.

If $K = \mathbb{Q}$, we will sometimes write L_p for $\mathbb{Q}_p[x]/(f)$; since f may be irreducible over $\mathbb{Q}$ and not over $\mathbb{Q}_p$, this may be a direct sum of extension fields of $\mathbb{Q}_p$, written $L_{\mathfrak{p}_1} \oplus \ldots \oplus L_{\mathfrak{p}_r}$ with $L_{\mathfrak{p}_i}$ generated by $\alpha_{\mathfrak{p}_i}$, even when L is a number field. If L were not a number field, the L_i each decompose separately into a direct sum of extension fields, so there is a natural map from each L_i into some subset of the $L_{\mathfrak{p}_j}$, and these subsets are disjoint as i runs from 1 to r. Composing these natural maps gives us a natural map $L \to L_p$. A similar construction holds where $K = K_{\mathfrak{p}}$, a local field not equal to $\mathbb{Q}_p$, and in this case we call the localised algebra $L_{\mathfrak{p}}$.

If we have a number field N and a set of primes $S \in \operatorname{Spec} \mathcal{O}_N$, it is possible to construct a finite subgroup $N(S, 2) < N^\times/(N^\times)^2$ consisting of all elements whose valuations at the prime ideals outside S are even; indeed, `magma` has an

in-built function `pSelmerGroup` to compute generators for $N(S,2)$. If N has class number 1, we have $N(S,2)$ generated by the generators for the unit group of N and the elements generating the prime ideals of S; otherwise the situation is slightly more complicated. Hence computing $N(S,2)$ requires knowledge of the class and unit groups of N.

If instead we have our algebra L and an element $D \in K$, we can define

$$S_i = \{\mathfrak{p} \in \operatorname{Spec} \mathcal{O}_{L_i} : \mathfrak{p} | D\mathcal{O}_{L_i}\}$$

and write

$$L(D;2) = \bigoplus_{i=1}^{r} L_i(S_i, 2);$$

the semi-colon indicates that D is to be an element of K rather than a set of places of K.

1.7 Checking local solubility

The result of an n-descent on an elliptic curve E, is a finite set of inequivalent n-coverings $C_1 \dots C_r$, each equipped with a map $j_{C_i} : C_i \to E$. Ideally we want to know which of these curves have points over K and correspond to elements of $E(K)/nE(K)$; at the very least, we want to find the Selmer group S_n consisting of curves with points everywhere locally.

There are two approaches to this local solubility in the literature. When the invariant-theory two-descent is used, the two-coverings are produced directly, and then Birch and Swinnerton-Dyer [5], and later Cremona [19] and Siksek [62], have approaches which check directly whether $f(x,z) = y^2$ has a solution in $\mathbb{Q}_p$. For $n = 4$, the paper [51] uses a similar approach for intersections of two quadrics, but this requires effort of at least $O(p^3)$ to show that such an intersection has no p-adic point.

The other approach, which was mentioned to me by Stoll, takes advantage of the association between n-coverings and elements of the algebra L of the previous section. For the two-descent case, for instance, we have for each place $\mathfrak{p}$ of K the diagram

$$\begin{array}{ccc} E(K)/2E(K) & \xhookrightarrow{\mu} & L^\times/(L^\times)^2 \\ \downarrow & & \downarrow \\ E(K_\mathfrak{p})/2E(K_\mathfrak{p}) & \xhookrightarrow{\mu_\mathfrak{p}} & L_\mathfrak{p}^\times/(L_\mathfrak{p}^\times)^2 \end{array}$$

where L and $L_\mathfrak{p}$ are an algebra of the type described in section 1.6 and its localisation at $\mathfrak{p}$, and the μ and $\mu_\mathfrak{p}$ maps are the modified "$x - \alpha$" maps of that

section.

Generally, it is not too difficult to find points on $E(K_\mathfrak{p})$, and it is possible to compute in advance (by the method of section 2.4.1) the size of the image $R_\mathfrak{p} = \mu_\mathfrak{p}(E(K_\mathfrak{p}))$ in $L_\mathfrak{p}$, which turns out to be fairly small. So you can generate points on $E(K_\mathfrak{p})$, compute their images in $L_\mathfrak{p}$, and continue until you have filled out $R_\mathfrak{p}$ entirely. We can then reject any element of L whose image in $L_\mathfrak{p}$ does not land in $R_\mathfrak{p}$.

A similar procedure, though with slightly different μ and slightly different quotients on the left-hand side of the diagram, is the local-solubility process used for the second descent presented in the next chapter, and much of the work in sections 2.2 through 2.4 revolves around computing the exact size of $\mu_p(\mathcal{C}(\mathbb{Q}_p))$; the relevant diagram appears at the start of section 2.4.

Chapter 2

Four-descent

2.1 Outline of the argument

We begin with a two-covering $\mathcal{C} : y^2 = f(x)$, where $f \in K[x]$ is a quartic polynomial; these two-coverings will nearly always have been obtained as a two-descendent of some elliptic curve E. We will assume that f is irreducible over K; this will be the case if $E(K)[2]$ is trivial. The aim of all our computation will be to find $x \in \mathbb{Q}$ with $f(x)$ a square, and hence a point on $\mathcal{C}$ which lifts by the map of appendix A to one on E. We follow fairly closely the procedure in [51]; I will point out where the procedure here is refined from theirs.

We find an algebraic condition, such that the possible solutions to $y^2 = f(x)$ are divided into a finite number of classes, each corresponding to a coset

$$\varepsilon = \phi(x) = (x - \alpha)\left(K^\times (L^\times)^2\right)$$

of the number field L defined in section 1.6, where α is the root of f in L defined there. Let $M = L^\times / K^\times (L^\times)^2$ be the group of such cosets. The set of these ε is called $\mathfrak{D}_4^{\text{alg}}$, and to each ε we can associate an intersection $\mathcal{H}_\varepsilon$ of two quadrics over $\mathbb{P}^3(K)$.

The 4-Selmer group S_4 – which contains $E(K)$ together with any elements of order 2 or 4 in $Ш_K(E)$ – consists of those intersections of quadrics which have points everywhere locally. The procedure in [51] handled local solvability without using L, by using Hensel's lemma on the $\mathcal{H}_\varepsilon$; this often broke down. We instead use the local-images technique of section 1.7, with some modifications described in section 2.4.

Let $\mathfrak{D}_4$ be the set of everywhere locally solvable $\mathcal{H}_\varepsilon$. The elements of $\mathfrak{D}_4$ correspond to elements $\mathcal{H} \in S_4$ for which $2\mathcal{H} = \mathcal{C}$. Their number will be zero if $\mathcal{C} \in S_2$ is not twice an element of S_4, which happens if $\mathcal{C}$ represents an element

of Ш of exact order two. If there are no elements of order four or more in Ш$[2^\infty]$, the size of $\mathfrak{D}_4$ will be $2^{\operatorname{rank} E-1}$: the -1 arises because each element of $\mathfrak{D}_4$ corresponds to **two** elements of S_4, since there is scope to choose the sign of y arbitrarily in $j_{\mathcal{H}} : \mathfrak{D}_4 \to (x, y) \in \mathcal{C}$. We perform minimisation and reduction at this point, aiming to obtain quadrics with small discriminant and small coefficients; these are newly-developed procedures not used in [51].

Nor did [51] describe any efficient procedure for finding points on the $\mathcal{H}$. In this work, by parametrizing one quadric and substituting into the other, we convert an intersection of quadrics into a single ternary quartic F, and seek a point on that using a very efficient two-dimensional sieving procedure. A point on F then lifts to one on each of the quadrics in $\mathcal{H}$, and then to one on $y^2 = f(x, z)$, and further up to one on the elliptic curve: for each of these liftings, the x co-ordinate is given by a quadratic function of the co-ordinates on the previous model. However, the map $\psi : F \to \mathcal{H}$ has degree one, and so does not contribute a factor two to the height of points. Instead, we find, writing $\psi([x, y, z]) = [a, b, c, d]$, that $\mathrm{GCD}(a, b, c, d)$ is large, and dividing it out gives a point on $\mathcal{H}$ with co-ordinates about the size of those of the point on F.

2.2 The algebraic stage

2.2.1 Recap of the two-descent for elliptic curves

Let E be an elliptic curve

$$E : y^2 = f(x) = x^3 + a_4 x + a_6$$

defined over a number field K; we assume that f is irreducible over K. Let L be the number field $K(x)/(f(x))$, and let α be the root of f in L. Define the homomorphism $\mu : E(K) \to L^\times/(L^\times)^2$ by

$$\mu(x, y) = \begin{cases} (x - \alpha)(L^\times)^2 & x \neq \alpha \\ f'(\alpha)(L^\times)^2 & x = \alpha \end{cases}.$$

Since μ is a group homomorphism into $(L^\times)^2$, its kernel clearly **contains** $2E(K)$: we can prove (eg lemma 15.2 of [13]) that $\ker \mu = 2E(K)$, so we have

$$E(K)/2E(K) \xhookrightarrow{\mu} L^\times/(L^\times)^2 \ .$$

If $(x, y) \in E(K)$ then $N_L(x - \alpha) = f(x) = y^2$, so the points of $E(K)$ correspond to elements of square norm in $L^\times/(L^\times)^2$. In fact, for K a number field, we can guarantee – this result is crucial in proving the weak Mordell-Weil theorem –

that they lie in a finite subset $L_2(S,2)$ of a finite set $L(S,2)$ of the type described in section 1.6:

$$L_2(S,2) = \left\{ t \in L(S,2) \mid N_L(t) \in (K^\times)^2 \right\}.$$

The standard proof of the weak Mordell-Weil theorem – for example Proposition 1.5b in chapter 8 of [63] – takes

$$S = \{\mathfrak{p} : \mathfrak{p} | 2\Delta\mathcal{O}_K\}$$

where Δ is the discriminant of the polynomial f, though Simon [66] has shown that we need only consider primes dividing $f'(\alpha)\mathcal{O}_K$ and whose square divides $\Delta\mathcal{O}_K$.

Whichever S we choose, `magma` can compute generators for $L(S,2)$, and it is a simple matter of linear algebra to compute from them generators for $L_2(S,2)$. This group $L_2(S,2)$ has the 2-Selmer group of $E(K)$ as a subgroup; however, it also contains elements which arise from two-coverings that are not everywhere locally solvable. We remove these by the method of section 1.7: for each place $\mathfrak{p}$ of K that supports disc f, we construct the image $R_\mathfrak{p}$ in the localised algebra $L_\mathfrak{p}$ of $E(K_\mathfrak{p})$, and discard any elements of $L_2(S,2)$ whose images in $L_\mathfrak{p}$ lie outside $R_\mathfrak{p}$.

The result of this filtration is a set L_2' of elements of $L^\times$; the only remaining obstruction is solubility at the infinite prime, which could be handled here, but which we instead handle while we convert L_2' into a set of explicit two-coverings $\mathcal{C} : y^2 = g(x)$, g quartic, representatives of the 2-Selmer group of $E(K)$.

The conversion from elements of $L^\times$ to explicit two-coverings constructs the two-coverings as the obstructions to writing $\ell \in L^\times$ in the form $(x-\alpha)r^2$ for some $r \in L^\times$ – if ℓ is of that form, the x is the X-coordinate of some point on E.

Given an element $\ell = \ell_0 + \ell_1\alpha + \ell_2\alpha^2 \in L_2(S,2)$, write a general r as $r = r_0 + r_1\alpha + r_2\alpha^2$, and multiply out; considering the coefficients of $1, \alpha, \alpha^2$ in the result gives us three homogenous ternary quadratic equations

$$Q_0(r_0, r_1, r_2) = x \tag{2.1}$$
$$Q_1(r_0, r_1, r_2) = -1 \tag{2.2}$$
$$Q_2(r_0, r_1, r_2) = 0 \tag{2.3}$$

so finding x requires us to find some simultaneous solution to (2.2) and (2.3), and substitute back into (2.1). Since we know that our ℓ corresponds to a two-covering soluble at all primes $\mathfrak{p}$, the only way in which no solution can exist

is if Q_2 is positive or negative definite, or Q_1 positive definite; these situations correspond to ℓ representing a two-covering not soluble at the infinite prime.

We note that there exist many algorithms which, given a a homogenous ternary quadratic form, will determine a parametric solution

$$r_0 = f_0(x, y), r_1 = f_1(x, y), r_2 = f_2(x, y)$$

with f_i themselves binary quadratic forms, or to show that no solution exists. The earliest such algorithms were given by Legendre, but there has been a recent push to reduce the amount of factorisation required, leading to the algorithms of Cremona-Rusin [24] and Simon [67]. The latter requires only the factorisation of the determinant of the 3×3 matrix defining Q_2, which is very convenient since this determinant is roughly $N_{L/\mathbb{Q}}(\ell)$, whilst the coefficients can be enormous in the case of complicated fundamental units.

If there is no solution, we discard the element of $L_2(S, 2)$ and continue to the next. Otherwise, substituting back into (2.2) gives a quartic equation $g(x, y) = -1$, and, since multiplying r by any element of $\mathbb{Q}$ would give just as valid a solution, we could deal instead with the equation $g(x, y) = -z^2$. Relabelling, we find we have associated a two-covering $\mathcal{C}_\ell : y^2 = f(x, z)$ (f quartic) with each of the elements $\ell \in L_2(S, 2)$ that has equation (2.3) solvable; this set of two-coverings contains one representative for each element of the 2-Selmer group of $E(K)$, and in particular counting them gives us the size of $S_2(E_K)$. Notice that we have thrown away the original ℓ at this stage; [66] shows that the two-coverings we obtain by this circuitous process are indeed two-coverings of our original elliptic curve E.

We minimise and reduce these two-coverings, since (particularly when the unit group of L has large generators) they may have very large coefficients and many spurious primes dividing their discriminants. This done, we search for K-rational points on all of them; if we find one on some $\mathcal{C}_\ell$ we can apply the syzygy map $j_{\mathcal{C}_\ell \to E}$ to it to obtain a point in $E(K)$, and restrict the rest of our searching to the non-trivial quotient of $\mathfrak{D}_2 / \langle \mathcal{C}_\ell \rangle$. If not, we record the two-covering as a fit subject for four-descent.

2.2.2 The action of E on elements of S_2

Let E be an elliptic curve defined over some field K (with group operation $+_E$ and base point 0_E) such that $E(K)[2]$ is trivial. Let $\mathcal{C} : y^2 = f(x)$ ($f \in K[x]$ with $\deg f = 4$) be one of its two-descendents; we learn from [22] that all non-trivial two-descendents of a curve without 2-torsion are of the form $y^2 = f(x)$ with f irreducible. The projective closure of $\mathcal{C}$ has one singular point, at infinity; this is a node whose branches correspond to the points $P_{\infty^\pm}$ of section 1.3.1.

As usual, we use the same letter $\mathcal{C}$ to denote a smooth model for the projective completion of the affine curve $y^2 = f(x)$. $\mathcal{C}$ has genus 1.

The material presented in this subsection is to a large extent a recap of material from section 10.3 of [63], but I collect it here for ease of reference.

For any extension K_1 of K (including $K_1 = K$), if we have a point $P \in \mathcal{C}(K_1)$, we can obtain an isomorphism (defined over K_1) $\theta_P : \mathcal{C} \to E$ with

$$\theta_P(C_1 \oplus_P C_2) = \theta_P(C_1) +_E \theta_P(C_2)$$

and $\theta_P(P) = 0_E$; the construction is described very explicitly in [18] and given in appendix A.1. If P and Q are distinct points on $\mathcal{C}$, the maps θ_P and θ_Q are related by

$$\theta_Q(R) = \theta_P(R) - \theta_P(Q). \tag{2.4}$$

The definition of "principal homogeneous space" indicates that there is an action of E on $\mathcal{C}$: given $P \in E$ and $Q \in \mathcal{C}$, let D be some point in $\mathcal{C}(\overline{K})$ and define an action

Definition 2.2.1. $P \boxplus Q = \theta_D^{-1}(P +_E \theta_D(Q))$

This action is independent of the choice of D : we have by (2.4)

$$\theta_{D'}(R) = \theta_D(R) - \theta_D(D'), \quad \theta_{D'}^{-1}(P) = \theta_D^{-1}(P + \theta_D(D'))$$

so $\theta_{D'}^{-1}(P +_E \theta_{D'}(Q)) = \theta_D^{-1}(P +_E \theta_D(Q))$.

There is an obvious associated operation

Definition 2.2.2. If $C_1, C_2 \in \mathcal{C}$, define

$$C_1 \boxminus C_2 = \theta_D(C_1) -_E \theta_D(C_2).$$

This is an element of E, and clearly by (2.4) independent of D.

Lemma 2.2.1. $\boxplus$ and $\boxminus$, defined as above, are K-rational maps.

Proof. We need to show that, for all $\sigma \in \mathrm{Gal}(\overline{K}/K)$, $P^\sigma \boxplus Q^\sigma = (P \boxplus Q)^\sigma$, and likewise for $\boxminus$.

From proposition 4.3 of Cremona's paper [22], we have that, for each $\sigma \in \mathrm{Gal}(\overline{K}/K)$, there is a $T_\sigma \in E[2]$ with $\theta_D^\sigma - \theta_D = T_\sigma$. Since $(\theta_D(R))^\sigma - \theta_D(R^\sigma) = \theta_D^\sigma(R^\sigma) - \theta_D(R^\sigma)$, it is also equal to T_σ, and so $(\theta_D(R))^\sigma = \theta_D(R^\sigma) +_E T_\sigma$.

Let $R = P \boxplus Q$, so $\theta_D(R) = P +_E \theta_D(Q)$. Then $\theta_D^\sigma(R^\sigma) = P^\sigma +_E \theta_D^\sigma(Q^\sigma)$, since the addition operation on E is defined over K and remains unchanged under the action of σ.

Hence

$$\theta_D(R^\sigma) +_E T_\sigma = P^\sigma +_E \theta_D(Q^\sigma) +_E T_\sigma$$

and, subtracting T_σ from both sides and applying θ_D^{-1}, we have

$$R^\sigma = \theta_D^{-1}(P^\sigma +_E \theta_D(Q^\sigma)).$$

So indeed $(P \boxplus Q)^\sigma = P^\sigma \boxplus Q^\sigma$.

To prove the result for $\boxminus$ is simpler:

$$\begin{aligned}(C_1 \boxminus C_2)^\sigma &= (\theta_D(C_1) -_E \theta_D(C_2))^\sigma \\ &= \theta_D^\sigma(C_1^\sigma) -_E \theta_D^\sigma(C_2^\sigma) \\ &= \theta_D(C_1^\sigma) +_E T_\sigma -_E (\theta_D(C_2^\sigma) +_E T_\sigma) \\ &= \theta_D(C_1^\sigma) - \theta_D(C_2^\sigma) \\ &= C_1^\sigma \boxminus C_2^\sigma\end{aligned}$$

□

Since $(Q_2 \boxminus Q_1) \boxplus Q_1 = Q_2$, and $P_1 \boxplus Q = P_2 \boxplus Q \iff P_1 = P_2$, it is clear that $E(\overline{K})$ acts simply transitively on $\mathcal{C}(\overline{K})$; thus, given any base point $Q_0 \in \mathcal{C}(\overline{K})$, we have a bijection $\mathcal{C}(\overline{K}) \longleftrightarrow E(\overline{K})$ by $P \boxplus Q_0 \longleftrightarrow P$.

Let H be a subgroup of $E(\overline{K})$ – for example, $E(K)$. The points of $\mathcal{C}(\overline{K})$ clearly split up into H-orbits of the form

$$H \boxplus Q_0 = \{P \boxplus Q_0 : P \in H\}.$$

Since $\boxplus$ is defined over K, all the points in an $E(K)$-orbit of $\mathcal{C}(\overline{K})$ will be defined over K iff one of them is, and so $\mathcal{C}(K)$ is a union of complete $E(K)$-orbits. Since $\boxminus$ is also defined over K, and $C_1 \boxminus C_2 \in E(K)$ if C_1 and C_2 are in $\mathcal{C}(K)$, $\mathcal{C}(K)$ must either be empty or consist of a single $E(K)$-orbit. In either case, $E(K)$ acts simply transitively on $\mathcal{C}(K)$, and, given a subgroup $H_1 < E(K)$, we can split $\mathcal{C}(K)$ into orbits within which $C_1 \boxminus C_2 \in H_1$.

Lemma 2.2.2. Let H_1 be a subgroup of $E(K)$ of finite index, and let $\|\mathcal{C}(K)/H_1\|$ be the number of H_1-orbits on $\mathcal{C}(K)$. If $\mathcal{C}(K)$ is non-empty, we have

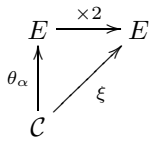
$$\|\mathcal{C}(K)/H_1\| = \|E(K)/H_1\| = [E(K) : H_1].$$

Proof. $\mathcal{C}(K)$ is a single $E(K)$-orbit, and so will be a union of H_1-orbits; the number of such orbits is precisely the index of H_1 in $E(K)$. □

2.2.3 What if we adjoin a root?

Let $\mathcal{C} : y^2 = f(x)$, with f an irreducible quartic polynomial defined over K, be a two-covering of the elliptic curve E. If we adjoin a root α of f to K, getting

the field K_α, we clearly have a point $P_\alpha = (\alpha, 0) \in \mathcal{C}(K_\alpha)$. Let $\theta_\alpha = \theta_{P_\alpha}$ be obtained by the construction of the previous subsection, and let $\xi : \mathcal{C} \to E$ be the syzygy map (defined over K, and called $j_{\mathcal{C}\to E}$ in section 1.3) described in appendix A.2. We observe, if necessary by computer algebra (or see [22]), that $\xi(P) = 2\theta_\alpha(P)$, giving the diagram



If $P = (x, y) \in \mathcal{C}$, write $P' = (x, -y)$; so $P'_\alpha = P_\alpha$. Recall from section 1.3.1 that divisors of the form $(P) + (P')$ are all equivalent to one another; hence they are all equivalent to $2(P_\alpha)$, and we have $P \oplus_\alpha P' = P_\alpha$ (the identity of the group $[\mathcal{C}(\overline{K}), P_\alpha, \oplus_\alpha]$) for all P; the three non-trivial points of order 2 on $[\mathcal{C}(\overline{K}), P_\alpha, \oplus_\alpha]$ are the $(\beta, 0)$, where β runs through the roots of f not equal to α.

If $\mathcal{C}(K) \neq \emptyset$, we can pick $P_0 \in \mathcal{C}(K)$, and consider the group $[\mathcal{C}(\overline{K}), P_0, \oplus_0]$ whose operation is defined over K. In this group, we have $P \oplus_0 P' = P'_0$ for all $P \in \mathcal{C}(\overline{K})$; in particular, $[2]_0 P_\alpha = P'_0$.

Lemma 2.2.3. Let $2\mathcal{C}(K)$ be the set

$$\{Q \oplus_0 Q : Q \in \mathcal{C}(K)\}.$$

Then $P'_0 \in 2\mathcal{C}(K)$ iff $g(x)$ has a root in K.

Proof. If $g(x)$ has a root $\gamma \in K$, then we can take $Q = (\gamma, 0)$; then $Q' = Q$ and $Q \oplus_0 Q' = P'_0$. For the other direction, we have

$$Q \oplus_0 Q = P'_0 = Q \oplus_0 Q',$$

and so $Q' = Q$; therefore the y co-ordinate of Q is zero, and $Q = (x, 0) \in \mathcal{C}(K)$ whence $g(x) = 0$. □

Lemma 2.2.4. If C_0 is any point in $\mathcal{C}(K)$, then $\theta_\alpha(\mathcal{C}(K)) = E(K) +_E Q_0$ where $Q_0 = \theta_\alpha(C_0)$.

Proof. Let $C_1 \in \mathcal{C}(K)$; we have $P = C_1 \boxminus C_0 \in E(K)$. Then $\theta_\alpha(C_1) = \theta_\alpha(P \boxplus C_0) = P +_E \theta_\alpha(C_0)$ by definition of $\boxplus$. □

In the remainder of this section, we want to find properties of $\mathcal{C}(K)$ which we can deduce from known properties of $E(K)$ and $\mathcal{C}(K_\alpha)$, without having to

use an explicit point in $\mathcal{C}(K)$ since the whole purpose of this work is to let us use the four-descent to **find** such a point.

So: let the subgroup G of $[\mathcal{C}(K_\alpha), P_\alpha, \oplus_\alpha]$ be defined as

$$G = \theta_\alpha^{-1}(E(K)).$$

Lemma 2.2.5. Suppose $\mathcal{C}(K)$ is non-empty, and contains the point P_0. Then $\mathcal{C}(K) = G \oplus_\alpha P_0$, the coset of G in $[\mathcal{C}(K_\alpha), P_\alpha, \oplus_\alpha]$ containing P_0.

Proof. Recall from lemma 2.2.4 that $\theta_\alpha(\mathcal{C}(K)) = E(K) +_E Q_0$, where $Q_0 = \theta_\alpha(P_0)$. So, if $P_1 \in \mathcal{C}(K)$, we have

$$\begin{aligned} \theta_\alpha(P_1) - \theta_\alpha(P_0) \in E(K) &\implies \theta_\alpha(P_1 \ominus_\alpha P_0) \in E(K) \\ &\implies P_1 \ominus_\alpha P_0 \in G \\ &\implies P_1 \in G \oplus_\alpha P_0 \end{aligned}$$

Conversely, if $P_1 \in \mathcal{C}(\overline{K})$, we have

$$\begin{aligned} P_1 \in G \oplus_\alpha P_0 &\implies \theta_\alpha(P_1) - \theta_\alpha(P_0) \in E(K) \\ &\implies \theta(P_1) \in Q_0 \oplus \theta(E(K) \\ &\implies P_1 \in \theta^{-1}(Q_0 \oplus E(K)) = \mathcal{C}(K) \end{aligned}$$

□

Lemma 2.2.6. The map $\psi_{P_0} : [\mathcal{C}(K), P_0, \oplus_0] \to E(K)$ defined by $P \to \theta_\alpha(P \ominus_\alpha P_0) = \theta_\alpha(P) -_E Q_0$ (where $Q_0 = \theta_\alpha(P_0)$) is a group isomorphism from $[\mathcal{C}(K), P_0, \oplus_0]$ to $E(K)$.

Proof. We know that the map $G \to \mathcal{C}(K)$ given by $P \to P \oplus_\alpha P_0$ is a group isomorphism from G to $[\mathcal{C}(K), P_0, \oplus_0]$, by working directly with the divisors as in section 1.3.1:

$$\begin{aligned} ((P_1 \oplus_\alpha P_0) \oplus_0 (P_2 \oplus_\alpha P_0)) &\sim (P_1 \oplus_\alpha P_0) + (P_2 \oplus_\alpha P_0) - (P_0) \\ &\sim ((P_1) + (P_0) - (P_\alpha)) + ((P_2) + (P_0) - (P_\alpha)) - (P_0) \\ &\sim (P_1) + (P_2) + (P_0) - 2(P_\alpha) \\ &\sim (P_1 \oplus_\alpha P_2) + (P_0) - (P_\alpha) \\ &= ((P_1 \oplus_\alpha P_2) \oplus_\alpha P_0) \end{aligned}$$

Under this isomorphism, we have $[\mathcal{C}(\overline{K}), P_\alpha] \simeq [\mathcal{C}(\overline{K}), P_0]$. Restricting this map to G gives us, by lemma 2.2.5, an isomorphism $G \simeq [\mathcal{C}(K), P_0]$; composing with θ_α, we get $[\mathcal{C}(K), P_0] \simeq E(K)$ via the map claimed in the lemma. □

Lemma 2.2.7. The homogenous-space action of $P \in E(K)$ on $Q \in \mathcal{C}(K)$ defined in definition 2.2.1 is given by

$$P \boxplus Q = Q \oplus_0 \psi_{P_0}^{-1}(P)$$

Proof. We have $\psi_{P_0}(R) = \theta_\alpha(R \ominus_\alpha P_0)$, so $\psi_{P_0}^{-1}(R) = \theta_\alpha^{-1}(R) \oplus_\alpha P_0$. Hence

$$\begin{aligned} Q \oplus_0 \psi_{P_0}^{-1}(P) &= Q \oplus_0 \left[\theta_\alpha^{-1}(P) \oplus_\alpha P_0\right] \\ &= \left[Q \oplus_0 \theta_\alpha^{-1}(P)\right] \oplus_\alpha P_0 \end{aligned}$$

But

$$\theta_0(Q \oplus_0 \left[\theta_\alpha^{-1}(P) \oplus_\alpha P_0\right]) = \theta_0(Q) +_E \theta_0(\theta_\alpha^{-1}(P) \oplus_\alpha P_0);$$

and by equation (2.4), the latter term is equal to P. And $\theta_0(P \boxplus Q) = P \oplus_0 \theta_0(Q)$ by definition 2.2.1; so $\theta_0(P \boxplus Q) = \theta_0(Q \oplus_0 \psi_{P_0}^{-1}(P))$ and, since θ_0 is an isomorphism, we can apply its inverse to both sides and get the desired result. □

If H is the subgroup of $E(K)$ generated by $2E(K)$ and $\theta_0(P_0^-)$, it is immediate that

$$\left\| \frac{[\mathcal{C}(K), P_0, \oplus_0]}{\langle 2\mathcal{C}(K), P_0^- \rangle} \right\| = \|\mathcal{C}(K)/H\| = \|E(K)/H\|$$

where $\|\mathcal{C}(K)/H\|$ is a count of H-orbits whilst $\|E(K)/H\|$ is the size of a quotient subgroup. We now want to know when P_0^- lies in $2\mathcal{C}(K)$, to tell whether this subgroup H is equal to $2E(K)$ or is twice the size. We do this by an implicit invocation of lemma 2.2.3.

Lemma 2.2.8. Let $\mathcal{C} : y^2 = f(x) = ax^4 + bx^3 + cx^2 + dx + e$ be a two-covering defined over K. Let $g(x) = x^3 - 3Ix + J$ be the associated cubic of $f(x)$, and let $g_1(x) = -27g(-\frac{x}{3})$ so $E : y^2 = g_1(x)$ is the elliptic curve associated to $\mathcal{C}$ by the syzygy map.

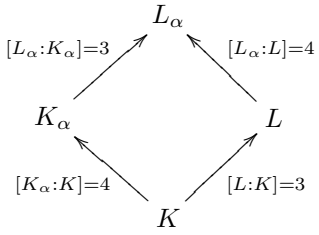
Let K_α be obtained by adjoining one root of f to K. Then the embedding

$$E(K)/2E(K) \longrightarrow E(K_\alpha)/2E(K_\alpha)$$

is injective if $\mathcal{C}(K) = \emptyset$, and has a kernel of size two (generated by $\xi(Q)$ where $Q \in \mathcal{C}(K)$) otherwise.

That is, $E(K) \cap 2E(K_\alpha) = 2E(K)$ if $\mathcal{C}(K) = \emptyset$, and otherwise is equal to $2E(K) + \langle \xi(Q) \rangle$

Proof. $P \in E(K)$ represents an element of the kernel of the embedding iff $P \in E(K) \cap 2E(K_\alpha)$. The proof uses the two-descent machinery of section 2.2.1; let $L = K(\beta)$ where β is a root of the cubic $g_1(x)$ (and so the root of $g(x)$ is $\beta' = -\beta/3$), and let $L_\alpha = L(\alpha)$. Recall from [22] the field diagram



where there is a unique intermediate quadratic field $L < M < L_\alpha$, $M = L(\sqrt{w})$, where w will be defined below.

Lemma 2.2.9. The map $L^\times/(L^\times)^2 \to L_\alpha^\times/(L_\alpha^\times)^2$ has a kernel of order precisely two

Proof. We have the unique intermediate quadratic field M; since it is a quadratic extension, we have $M = L(\sqrt{w})$ for some $w \in L^\times/(L^\times)^2$. So $w \in L$ is a non-trivial element of the kernel.

Let $\kappa = \sqrt{w}$, and write an arbitrary element of M as $a + b\kappa$.

$$\begin{aligned}(a+b\kappa)^2 &= a^2 + 2ab\kappa + b^2\kappa^2 \\ &= (a^2 + b^2 w) + 2ab\kappa\end{aligned}$$

so, if we require $(a+b\kappa)^2 \in L^\times$, we need $ab = 0$. If $b = 0$ then $(a+b\kappa)^2 = a^2 \in (L^\times)^2$; if $a = 0$ then $(a+b\kappa)^2 = b^2 w \in w(L^\times)^2$. So w is unique. □

We have the diagram

$$\begin{array}{ccc} E(K)/2E(K) & \xhookrightarrow{(x,y)\to x-\beta} & L^\times/(L^\times)^2 \\ \downarrow & & \downarrow \\ E(K_\alpha)/2E(K_\alpha) & \xhookrightarrow{(x,y)\to x-\beta} & L_\alpha^\times/(L_\alpha^\times)^2 \end{array}$$

By lemma 2.2.9, the right-hand map has a kernel of order exactly two, with the non-trivial element represented by, say, $w \in L^\times$. The left-hand map thus has a kernel of order one or two, depending on whether w is or is not in the image of $E(K)$.

But, combining that lemma and the first part of proposition 3.2 of [22], we have that the element w generating M is given absolutely explicitly as $w = \frac{1}{3}(4a\beta' + 3b^2 - 8ac)$, and that this element lies in the image of $E(K)$ iff $\mathcal{C}(K)$ is non-empty. □

Lemma 2.2.10. Let $P_0 = (x, y), y \neq 0$ be any point on a two-covering $\mathcal{C}(K)$ of the elliptic curve E. From P_0 we obtain a second point $P_0^- = (x, -y)$, a group law $\oplus_0$ and a map $\theta_0 : \mathcal{C} \to E$; let P_α be the point $(\alpha, 0)$ that we know and love, giving us another group law $\oplus_\alpha$ and a map $\theta_\alpha : \mathcal{C} \to E$, and recall the syzygy map $\xi : \mathcal{C} \to E$ with $\xi = [2]\theta_\alpha$.

Then the subgroup H of $E(K_\alpha)$ defined as $2E(K) + \left\langle \theta_0(P_0^-) \right\rangle$ is equal to $E(K) \cap 2E(K_\alpha)$; in particular, it does not depend on P_0.

Proof. We have $P_0 \oplus_\alpha P_0^- = P_\alpha$, and so $\theta_\alpha(P_0^-) = -\theta_\alpha(P_0)$. And so

$$\begin{aligned}
\theta_0(P_0^-) &= \theta_\alpha(P_0^- \ominus_\alpha P_0) \\
&= \theta_\alpha((\ominus_\alpha P_0) \ominus_\alpha P_0) \\
&= -2\theta_\alpha(P_0) \\
&= -\xi(P_0) \in E(K)
\end{aligned}$$

So $H = 2E(K) + \left\langle \theta_0(P_0^-) \right\rangle$ is the same as $2E(K) + \langle \xi(P_0) \rangle$; by lemma 2.2.8, we know that $E(K) \cap 2E(K_\alpha) = 2E(K) + \xi(P_0)$, and so we have shown $H = E(K) \cap 2E(K_\alpha)$.

□

2.2.4 The Main Theorem for two-coverings

Recall the following theorem of Cassels ([14], pp 35–39):

Theorem 2.2.1. Consider the genus-1 curve $\mathcal{C} : Y^2 = f(X)$, with f a **monic** quartic polynomial defined over a field K with non-zero discriminant. $\mathcal{C}$ has two points $P_{\infty^\pm}$ at infinity. Let $G = [\mathcal{C}(K), P_{\infty^+}, \oplus_{\infty^+}]$ be the Mordell-Weil group using P_{∞^+} as the origin. Let L be the algebra $K[X]/(f(X))$, and let M denote the quotient $L^\times / K^\times (L^\times)^2$.

L is a direct sum of r number fields L_i, where L_i is generated by α_i; $L_i^\times / (L_i^\times)^2$ is a direct sum of quotients $M_i = L_i^\times / (L_i^\times)^2$ arising from such number fields, and $M = \bigoplus M_i / K^\times$

Define $\mu_i : G \to M_i$ by

$$\begin{aligned} \mu_i(x,y) &= x - \alpha_i && (y \neq 0) \\ \mu_i(P_{\infty^\pm}) &= 1 \\ \mu_i(\alpha_i, 0) &= g'(\alpha_i) \end{aligned}$$

and define $\mu : G \to M$ by $g \to [\mu_1(g), \ldots, \mu_r(g)]$; this is the modified "$x - \alpha$" map referred to in section 1.6.

Then μ is a group homomorphism, whose kernel is the subgroup generated by $2\mathcal{C}(K)$ and P_{∞^-}; by the same argument as section 2.2.1, $\operatorname{im}\mu \subset N$, where $N < M$ is the subgroup of elements of M with $N_{M/K} \in (K^\times)^2$.

As an obvious corollary, we have that the image of μ is a subgroup of M of order $[\mathcal{C}(K) : 2\mathcal{C}(K) + \langle P_{\infty^-} \rangle]$.

Note that, if $\mathcal{C} : Y^2 = f(X)$ where the coefficient of X^4 in $f(X)$ is a_1^2 ($a_1 \in K$), then $\mathcal{C} \cong \mathcal{C}_1$ where $\mathcal{C}_1 : Y^2 = f_1(X)$ and $f_1(X) = f(X)/a_1^2$ by the map $(x,y) \to (x, y/a_1)$. The α_i will be the same for $L = K[x]/(f(x))$ and $L_1 = K[x]/(f_1(x))$, so the theorem holds for these more general curves.

Next, we consider a two-covering $\mathcal{C}$ given by a quartic without any special condition on the leading coefficient, but where we assume knowledge of a point on $\mathcal{C}(K)$:

Theorem 2.2.2. Let $\mathcal{C} : y^2 = f(x)$ where $f \in K[x]$ is a quartic; let $P_0 = (x_0, y_0) \in \mathcal{C}(K)$ so we consider the Mordell-Weil group $[\mathcal{C}(K), P_0, \oplus_0]$. As before, let $L = K[x]/(f(x))$ and $M = L^\times/K^\times(L^\times)^2$, with $L = \bigoplus L_i$ and α_i generating L_i.

Define $\mu_i : \mathcal{C}(K) \to L_i$ at the point $P = (x,y)$ by

$$\mu_i(P) = \begin{cases} 1 & P = P_0, P_0' \\ x_0 - \alpha_i & P = P_{\infty^\pm} \\ f'(\alpha_i) & P = (\alpha_i, 0) \\ \frac{x-\alpha_i}{x_0-\alpha_i} & \text{otherwise} \end{cases}$$

($P_{\infty^\pm}$ will only be in $\mathcal{C}(K)$ if f happened to have leading coefficient a square), and define $\mu : G \to M$ as $g \to [\mu_1(g), \ldots, \mu_r(g)]$ as in the previous theorem.

Then μ is a group homomorphism from $[\mathcal{C}(K), P_0, \oplus_0]$ to M, whose image lies in the subgroup

$$N = \left\{ \varepsilon \in M | N_{A/K}(\varepsilon) \in (K^\times)^2 \right\},$$

whose kernel is the subgroup of $[\mathcal{C}(K), P_0, \oplus_0]$ generated by $2\mathcal{C}(K)$ and $P_0{}'$, and

whose image in N has size $[\mathcal{C}(K) : 2\mathcal{C}(K) + \langle P_0' \rangle]$.

Proof. Essentially, we construct a birational map ϕ from $\mathcal{C}$ to a model $\mathcal{C}_2$ satisfying the conditions of theorem 2.2.1, and compose the μ of the original theorem with the ϕ to get the μ of the new theorem.

We have $P_0 = (x_0, y_0) \in \mathcal{C}(K)$. Let $g(t) = t^4 f(\frac{1}{t} + x_0)$, defining a polynomial with square leading coefficient, and let $\mathcal{C}_2 : y^2 = g(t)$. Let $Q_{\infty^\pm}$ be the two points at infinity on $\mathcal{C}_2$.

Now consider the map $\phi : \mathcal{C} \to \mathcal{C}_2$ given by

$$\phi(x, y) = \left((x - x_0)^{-1}, (x - x_0)^{-2} y\right); \quad \phi^{-1}(x_2, y_2) = \left(x_2^{-1} + x_0, x_2^{-2} y_2\right);$$

this is constructed by a translation to make the x co-ordinate of P_0 equal to zero, which makes the constant term a square, followed by an inversion to move the square term to the x^4 position. It is a bijection between $\mathcal{C}(K)$ and $\mathcal{C}_2(K)$, except that $\phi(P_0) = Q_{\infty^+}$ and $\phi(P_0') = Q_{\infty^-}$; indeed, it is a birational isomorphism

$$[\mathcal{C}, P_0, \oplus_0] \longleftrightarrow [\mathcal{C}_2, P_{\infty^+}, \oplus_2]$$

defined over K and correctly handling the group law.

We now show that the definition of μ in theorem 2.2.2 is the composition of $\phi : \mathcal{C} \to \mathcal{C}_2$ with the map $\mu_2 : \mathcal{C}_2(K) \to L$ obtained when we apply theorem 2.2.1 for the curve $\mathcal{C}_2$, and hence the image and kernel of μ will be the images under ϕ of the image and kernel of μ_2 for the other curve.

Away from the special cases, we have

$$\phi(x, y) = \left(\frac{1}{x - x_0}, \frac{y}{(x - x_0)^2}\right).$$

The generic roots $\alpha_i^{(2)}$ of the quartic defining $\mathcal{C}_2$ will be $(\alpha_i - x_0)^{-1}$ in the L of the current theorem, so the image $\mu_2(\phi(x, y))$ is

$$\frac{1}{x - x_0} - \frac{1}{\alpha_i - x_0} = \frac{\alpha_i - x}{(x - x_0)(\alpha_i - x_0)} \equiv \frac{x - \alpha_i}{x_0 - \alpha_i}$$

since $x - x_0 \in K^\times$.

For the special cases, we have

1. The points P_0 and P_0' map to the points at infinity on $\mathcal{C}_2$, which are sent to 1 by μ_2
2. The points at infinity $(0, \pm\sqrt{a})$ are sent to

$$0 - \alpha_i^{(2)} = \frac{1}{x_0 - \alpha_i} \equiv x_0 - \alpha_i \pmod{(L^\times)^2}$$

3. If $g(\alpha_i) = 0$, then $P_1 = (\alpha_i, 0) \in \mathcal{C}$ maps to $Q_1 = (\alpha_i^{(2)}, 0) \in \mathcal{C}_2$, and $\mu_2(Q_1) = g'(\alpha_i^{(2)})$, but, since $g(t) = t^4 f(t^{-1} + x_0)$, we have

$$g'(t) = 4t^3 f(t^{-1} + x_0) - t^2 f'(t^{-1} + x_0).$$

And so

$$g'(\alpha_i^{(2)}) = -(\alpha_i^{(2)})^2 f'(\alpha_i) \equiv f'(\alpha) \pmod{K^\times (L^\times)^2}.$$

The result about the kernel of μ follows at once from theorem 2.2.1; for the image, we have $N_{L/K}(x - \alpha) = y^2/a$, $N_{L/K}(x_0 - \alpha) = y_0^2/a$, and so $N_{L/K}(\mu(x, y)) = (y/y_0)^2 \in (K^\times)^2$. If the points at infinity are rational, the leading coefficient a is itself in $(K^\times)^2$, say $a = a_0^2$; so

$$N(\mu(P_{\infty^\pm})) = N(x_0 - \alpha) = y_0^2/a = y_0^2/a_0^2 \in (K^\times)^2.$$

□

The final task is to remove the dependency on the point P_0. We can already show using the results of subsection 2.2.3 that the *size* of the image $\mu(\mathcal{C}(K))$ does not depend on P_0; this size is $[\mathcal{C}(K) : 2\mathcal{C}(K) + \langle P_0' \rangle]$, which by lemma 2.2.3 is equal to

$$v[(\mathcal{C}(K), P_0) : 2(\mathcal{C}(K), P_0)]$$

with

$$v = \begin{cases} 1 & f(x) \text{ has a root in } K \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

and, applying lemma 2.2.6, the order of the image is equal to $[E(K) : 2E(K) + \theta_\alpha(P_0)]$, which we know by lemma 2.2.8 to equal $v[E(K) : 2E(K)]$. Write $Q_0 = \theta_\alpha(P_0)$.

Summarising, at the moment we have a subgroup $X < N$, and a bijection

$$\frac{\mathcal{C}(K)}{2E(K) + \langle Q_0 \rangle} \longrightarrow X$$

given as

$$(x, y) \to \mu((x, y)) = \frac{x - \alpha}{x_0 - \alpha}.$$

We now multiply through by $x_0 - \alpha$ to get

Theorem 2.2.3 (The Main Theorem). Let $\mathcal{C} : ay^2 = f(x) = x^4 + bx^3 + cx^2 + dx + e$ be a two-covering, defined over a field K, of the elliptic curve $E : y^2 = x^3 - 27I(f)x - 27J(f)$.

Let L be the algebra $K[x]/(f(x))$, let M be the quotient $L^\times/K^\times(L^\times)^2$. Let L be written as a direct sum $\bigoplus_{i=1}^r L_i$ with the root in L_i of f being α_i. Let $\mu_i : \mathcal{C}(K) \to L_i^\times$ be given by $\mu_i((x,y)) = x - \alpha_i$ if $x \neq \alpha_i$ and $\mu_i((x,y)) = 1$ otherwise, and let $\mu : \mathcal{C}(K) \to L^\times$ be given as $\mu(P) = [\mu_1(P), \ldots, \mu_r(P)]$.

Then μ gives a bijection

$$\frac{\mathcal{C}(K)}{2E(K) + \langle Q_0 \rangle} \longleftrightarrow X'$$

where X' is a coset of X in M, consisting of elements whose norm in $(K^\times)^2$ is equal to a.

Proof. The statement of this theorem is just that of theorem 2.2.2 with μ replaced by $\mu' = (x_0 - \alpha)\mu$. So, if X is the image subgroup obtained above, X' will be the coset $(x_0 - \alpha)X$. The norm of $x_0 - \alpha$ is equal to a modulo squares; the elements of X had square norm, so all the elements of X' will have norm equal to a modulo squares. □

This Main Theorem is given without proof in [51]; the proof given here is new.

2.3 Putting it together – explicit four-descent over $\mathbb{Q}$

In the remainder of this section, we consider four-descent only over $\mathbb{Q}$, and only for elliptic curves without rational 2-torsion; for such curves, all two-descendents will be of the form $y^2 = g(x)$ with g irreducible. Given such a two-covering $\mathcal{C} : y^2 = g(x)$, we start by changing variables so we are dealing with the equation $ay^2 = f(x)$, where f is a monic quartic whose coefficient of x^3 is equal to zero: a is the coefficient of x^4 in $g(x)$. Let $\Delta = \operatorname{disc} f$.

We construct the algebra $L = \mathbb{Q}[x]/\mathbb{Q}[x]f(x)$ described in 1.6; since f is irreducible over $\mathbb{Q}$, this is simply the number field $K(\alpha)$ obtained by adjoining one root. We have

$$N_{L/K}(x - \alpha) = f(x)$$

and so the task of finding points on $\mathcal{C}$ is going to be a subset of the task of finding elements of L with norms equal to a times a square.

Lemma 2 of [51] informs us that, if we have x, z with $f(x/z) = ay^2$, and consider the decomposition of the ideals $(x - \theta_i z)\mathcal{O}_{L_i}$ of the number fields as $\mathfrak{a}_i\mathfrak{b}_i^2$ where $\mathfrak{a}_i$ is a squarefree ideal, then those prime ideals of L_i dividing $\mathfrak{a}_i$ and not lying above 2 must divide either $a\mathcal{O}_{L_i}$ or $\Delta\mathcal{O}_{L_i}$. In the language of 1.6, this

tells us that

$$(x - \theta_i z)(L^\times)^2 \in L(2a\Delta; 2).$$

The Main Theorem tells us that elements of $(L^\times)^2$ differing by an factor in $\mathbb{Q}(L^\times)^2$ are equivalent, and so we should be working in $L^\times/\mathbb{Q}^\times(L^\times)^2$. So, having computed generators for $L(2a\Delta; 2)$ we find the image e_p in it of each of the rational primes $p_1, \ldots, p_n$ dividing $2a\Delta$, and take the quotient

$$L' = L(2a\Delta; 2)/\left\langle e_{p_1}, \ldots, e_{p_n} \right\rangle.$$

You may want to think of the object we are quotienting by as $\mathbb{Q}(2a\Delta; 2)$.

The Dirichlet S-unit theorem tells us that $L(2a\Delta; 2)$ will be a 2-group with

$$\sum_{p|2a\Delta} \|\{\mathfrak{p} : \mathfrak{p}|p\}\| + r + c - 1$$

generators, where r and c are the number of real and complex embeddings of K; the definition of L' loses one generator per rational prime dividing $2a\Delta$, and so it will have

$$n_{L'} = r + c - 1 + \sum_{p|2a\Delta} (\|\{\mathfrak{p} : \mathfrak{p}|p\}\| - 1)$$

generators. We store these generators in a list called `RB`.

We now use the fact that the norm of an element of $(L^\times)^2$ corresponding to a point on $\mathcal{C}$ will be a times a square. Consider the basis $\langle \ell_1, \ldots, \ell_n \rangle$ that we have constructed for L'. By factorising the norms of the ℓ_i as $N(\ell_i) = \prod p_j^{e_{ij}}$ and performing linear algebra on the matrix $E : E_{ij} = e_{ij} \mod 2$ over $\mathbb{F}_2$, we obtain a basis for the group L_2 of elements of L' of square norm, and separately a single element ξ of L' of norm ab^2 where a is the leading coefficient of the polynomial. We store the basis elements and ξ as elements of $\mathbb{F}_2^{N_{L'}}$ indicating which elements of `RB` should be taken into a product: the generators of `RB` can be large enough that we want to delay computation of explicit products for as long as possible. Let λ be the element in $\mathbb{F}_2^{N_{L'}}$ representing ξ, and M_S be the matrix whose rows are the elements in $\mathbb{F}_2^{N_{L'}}$ representing the generators of L_2.

If we have no element with norm in $a(\mathbb{Q}^\times)^2$, then there can be no four-descendents and we have shown that the two-covering $\mathcal{C}$ represented an element of Ш[2]. Otherwise, write

$$\mathfrak{D}_4^{\text{alg}}(f) = \left\{\varepsilon \in L'(S, 2) : \exists k \in K : N_{L/K}(\varepsilon) = ak^2\right\}$$

for the set of "algebraic four-descendents" obtained by this process for the two-covering $y^2 = f(x)$, and represent it by the triple $(\texttt{RB}, M_S, \lambda)$.

2.3.1 From algebraic four-descendents to four-coverings

To convert algebraic four-descendents into four-coverings of the shape in section 1.4.1, consider an element $\varepsilon \in L$ with $N_{L/K}\varepsilon = ab^2$. Points on $\mathcal{C}$ correspond to elements of the form $a + b\theta$ satisfying this norm condition, so we seek some $\ell \in L$ such that

$$\varepsilon\ell^2 = a + b\theta.$$

Writing ℓ with respect to a power basis as $\ell = \sum_{i=0}^{3} \ell_i\theta^i$, we see that

$$\varepsilon\ell^2 = \sum_{i=0}^{3} q_i\theta^i,$$

where each q_i is a homogenous quadratic in the four ℓ_i; requiring $q_2 = q_3 = 0$ then gives us immediately an intersection of two $\mathbb{P}^3$ quadrics. The coefficients of $\ell_i\ell_j$ for $i \neq j$ will be even by construction; so, defining A and B by $q_2 = \mathbf{x}^{\mathrm{T}}A\mathbf{x}$ and $q_3 = \mathbf{x}^{\mathrm{T}}B\mathbf{x}$, we get the four-covering $[A, B]$. And Merriman, Siksek and Smart proved in [51] that this procedure indeed produces a four-covering whose associated quartic is equivalent to the $f(x)$ that the process started with; that, after all this work, we have in fact performed a four-descent on the expected curve.

Theorem 2.3.1 (Banded structure). The matrices A and B arising from this construction are both of the 'banded' form

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_2 & a_3 & a_4 & a_5 \\ a_3 & a_4 & a_5 & a_6 \\ a_4 & a_5 & a_6 & a_7 \end{pmatrix}.$$

Proof. Using computer algebra, expand the θ^2 and θ^3 terms of a totally general product $(\sum_{i=0}^{3} b_i\theta^i)^2(\sum_{i=0}^{3} c_i\theta^i)$, over the totally general ring $\mathbb{Q}(\theta)/(\theta^4 + \sum_{j=0}^{3} d_i\theta^i)$, write them as quadratic forms in the b_i, convert to the matrix notation and observe the above structure. □

I do not take advantage of this structure at the moment – it would make the calculation of A and B very slightly more efficient at the price of using significantly more ugly code – and it disappears as we minimise and reduce the four-coverings.

It is not worth converting the ε into explicit four-coverings until local solvability has been checked, if only because the triple $(\texttt{RB}, M_S, \lambda)$ where M_S has n rows is considerably more convenient to store than 2^n pairs of matrices with possibly-vast coefficients.

2.3.2 Optimising the leading coefficient

Given a specific two-covering $y^2 = f_0(x, z)$, we are at liberty to perform a change of variables and consider the equivalent two-covering

$$y^2 = f_1(x, z) = f_0(\alpha x + \beta z, \gamma x + \delta z)$$

for any parameters $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$ with $\alpha\delta - \beta\gamma = 1$. The quartics f_1 and f_0 generate isomorphic number fields and have the same discriminant, but the leading coefficient of f_1 is equal to $f(\alpha, \gamma)$, whose prime factorisation need have nothing in common with that of the leading coefficient of f_0.

If solvability over $\mathbb{R}$ is our interest, we note that any two-covering possessing real points must have an equivalent two-covering with negative leading coefficient, since real points indicate a change in sign of $f_0(x, 1)$ and thus $f_0(x, 1)$ is sometimes negative; so we can pick a rational $\frac{\alpha}{\gamma}$ with $f_0(\alpha, \gamma) < 0$, construct an element of $\mathrm{SL}_2(\mathbb{Z})$ with $(\alpha\ \gamma)^{\mathrm{T}}$ as the first column, and transform by this. Quartics with negative leading coefficient are positive only in a finite interval; surprisingly often[1] it's possible to transform by an element of $\mathrm{SL}_2(\mathbb{Z})$ with small entries so as to make this finite interval very narrow, which assists the straightforward sieving search for points with $f(x, z) = y^2$ by allowing very few x (often zero, since $x \in \mathbb{Z}$) to be considered at each z level.

Since the set S of primes used in the computation of the Selmer group can be written as $S = S_\Delta \cup S_a$, where S_Δ contains primes dividing 2 or the discriminant, and S_a primes dividing the leading coefficient, we can make it smaller only by changing the leading coefficient. This was more critical when checking local solvability was the slowest part of the four-descent, but even now a smaller S makes the rest of the computations more convenient, particularly if it allows us to avoid a bad prime for which the computation of the local image by the method of section 2.4.4 takes too long. It also makes the certificate of membership in ${}_2Ш$ shorter if we can give an equivalent two-covering with no algebraic four-descendents.

To investigate the effect of replacing f by an equivalent quartic, I considered a number of two-coverings $y^2 = f(x)$ known to represent elements of Ш[2] for a variety of elliptic curves, prepared a set of transformed two-coverings $y^2 = f_1(x) \ldots y^2 = f_{100}(x)$ by applying random elements of $\mathrm{SL}_2(\mathbb{Z})$ (picked to have coefficients in $-50 \ldots 50$), and examining the distribution of $\left\| \mathfrak{D}_4^{\mathrm{alg}}(f_i) \right\|$.

This experiment gave evidence to justify the definitions

Definition 2.3.1. A two-covering $\mathcal{C} : y^2 = f(x)$ is *resolvable* if there exists an

[1] In experiments with several dozen two-coverings for curves with $\|Ш[2]\| = 4$ and small conductor, I was always able to arrange an interval of positivity narrower than 0.01 using an element of $\mathrm{SL}_2(\mathbb{Z})$ with entries of absolute value less than fifteen

element $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $\mathfrak{D}_4^{\mathrm{alg}}(M \cdot f)$ is empty.

Definition 2.3.2. Let $y^2 = f(x)$ be a two-covering. Then

$$\min_{M \in \mathrm{SL}_2(\mathbb{Z})} \left\| \mathfrak{D}_4^{\mathrm{alg}}(M \cdot f) \right\|$$

is a well-defined function of f, which I will call the *irresolvability*; if the irresolvability is zero, the two-covering is resolvable, and therefore insoluble.

The term "resolvable" is used because we are checking whether we can resolve the question of the solubility of $\mathcal{C}$ at the algebraic stage without checking local images.

In my sample of 753 two-coverings (three per elliptic curve from each of the 251 curves of $N < 12000$, rank zero and $Ш_{\mathrm{anal}} = 4$ from [20]), the irresolvabilities appeared to be distributed as follows – I am not sure why $\left\| \mathfrak{D}_4^{\mathrm{alg}}(M \cdot f) \right\| = 1$ is never observed:

'Irresolvability'	Number of times observed	Count of lowest-observed value
0	69	20–52
2	261	2–62
4	318	6–76
8	97	11–41
16	7	15–36
32	1	28

Note that these are simply the smallest *observed* value of $\left\| \mathfrak{D}_4^{\mathrm{alg}}(M \cdot f) \right\|$ – I have no way beyond repeated experiment of estimating the irresolvability for a given two-covering, though for the cases where the lowest-observed value was observed less than 10% of the time, I re-calculated the distribution using ten times as many random $\mathrm{SL}_2(\mathbb{Z})$ matrices (picked to have coefficients in $-200 \ldots 200$ this time) and, whilst χ^2 tests suggested that the distribution did change when the range of coefficients for the matrices was widened, the lowest-observed value didn't change.

We observe that, more than 43% of the time, the irresolvability is less than the size of $Ш_{\mathrm{anal}}$, meaning that, if we had a method to determine irresolvability and replace a two-covering by an equivalent one with smaller $\mathfrak{D}_4^{\mathrm{alg}}$, it would frequently serve as an improvement.

Siksek, in [61], also considers $M \cdot f$ for many M; if $f(\alpha, \beta) = \gamma\delta^2$ with γ small, he derives by quadratic-reciprocity arguments a congruence condition on $\alpha X + \beta Y$ with $f(X, Y)$ square. By a "duelling congruences" argument, he can sometimes prove, particularly if $f(x, 1)$ is positive only on a very narrow subinterval of $\mathbb{R}$, that a two-covering can have no rational points; if a proof does

not materialise, he can construct finitely many lattices in $\mathbb{Z}^2$, each of reasonably large discriminant, such that a point (X, Y) with $f(X, Y) = Z^2$ must lie on one of the lattices.

This procedure represents an entirely different approach from descent to the goal of proving that a two-covering has no rational points, and non-trivial Ш need not be an obstacle to its success (in the paper, Siksek shows that the three two-coverings for the elliptic curve of lowest conductor with non-trivial Ш are all insoluble).

However, I did not find the procedure very generally applicable; most two-coverings that I looked at do not have $f(\alpha, \beta) = \gamma\delta^2$ for γ small when I searched for $|\alpha|, |\beta| < 1000$, and large γ do not produce helpful congruences – the congruence conditions eliminate half the residues mod γ or mod 4γ, meaning that we obtain roughly γ separate lattices to consider. Moreover, the duelling-congruence argument appears to work well only when the interval over which $f(x, 1)$ is positive is very narrow; of course, f can often be transformed so that it has this property. It would be interesting to consider using Elkies' method [33] to find points with $|f(\alpha, \beta)|$ small, and then search among those points for ones with $f(\alpha, \beta) = kr^2$ for small $|k|$.

2.4 Checking p-adic solvability by local images

We wish to find an efficient way of working out which elements of $\mathfrak{D}_4^{\mathrm{alg}}$ have points everywhere locally, and so actually correspond to elements of $S_4(E)$. We use the technique of chapter 1.7; the result of section 2.2.4 indicates that, for the two-covering $\mathcal{C} : y^2 = f(x)$, I should consider the diagram

$$\begin{array}{ccc} \mathcal{C}(\mathbb{Q}) & \xrightarrow{\mu} & L^\times/\mathbb{Q}^\times(L^\times)^2 \\ \downarrow & & \downarrow \\ \mathcal{C}(\mathbb{Q}_p) & \xrightarrow{\mu_p} & L_p^\times/\mathbb{Q}_p^\times(L_p^\times)^2 \end{array}$$

and the relevant primes are the infinite prime, 2, and the primes dividing $\Delta = \operatorname{disc} f$. The infinite prime is handled in section 2.4.3; in sections 2.4.1 and 2.4.2 we compute the size of $\mu_p(\mathcal{C}(\mathbb{Q}_p))$, which turns out to have at most 8 elements, and in section 2.4.4 we discuss how to compute the image itself given its size.

The input to this routine is the representation for $\mathfrak{D}_4^{\mathrm{alg}}$ defined in section 2.3 – that is, a triple $(\texttt{RB}, M_S, \lambda)$ where RB contains n elements of $L^\times/K^\times(L^\times)^2$ generating L', M_S is an $m \times n$ matrix indicating which products of elements of RB have square norm and have passed all the local tests tried so far, and $\lambda \in \mathbb{F}_2^n$ corresponding to a product of elements of RB of norm ab^2 for a the leading

coefficient of the quartic defining $\mathcal{C}$.

For each (possibly-infinite) prime p, we produce a linear map $f : \texttt{RB} \to V$ and a coset of V which is the image

$$\{f(T) : T \in L' \text{ could correspond to an } \mathcal{H} \text{ soluble at } p\}.$$

We then use the following lemma to update λ and M_S to represent the elements of $\mathfrak{D}_4^{\text{alg}}$ which are locally soluble at all the primes considered so far, and, unless we have shown there are none, we proceed to the next prime.

Lemma 2.4.1 (Explicit intersection of a coset and the image of a coset under a linear map). Suppose we have a linear map $f : \mathbf{x} \to \mathbf{x}M_1$ from a vector space $V_1 = \mathbb{F}_2^a$ to a vector space $V_2 = \mathbb{F}_2^b$, a source coset $C_1 \subset V_1$ of the form $\lambda + S$ and a target coset $C_2 \subset V_2$ of the form $\mu + T$, where T is a d-dimensional subspace of V_2, and S a c-dimensional subspace of V_1. Let M_T and M_S be the matrices (of sizes $d \times b$ and $c \times a$ respectively) giving bases for these subspaces. Let

$$M_T^* = \left(\ker({M_T}^{\mathrm{T}})\right)^{\mathrm{T}}$$

be the dual matrix of M_T.

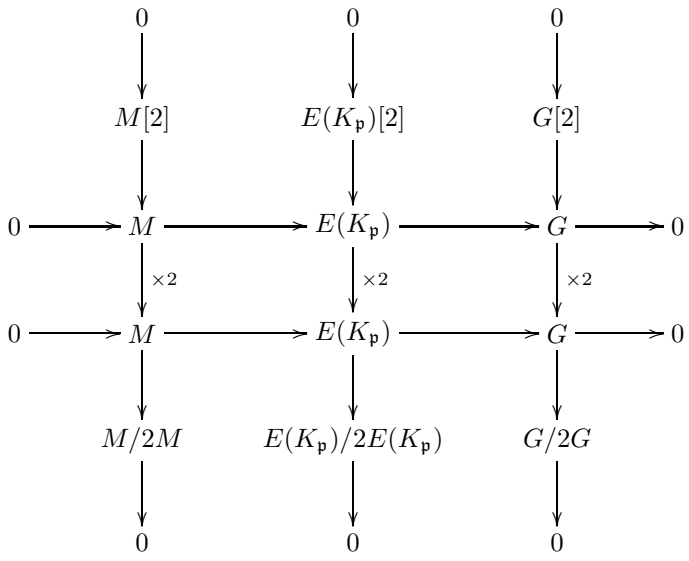
Then, if there are no solutions to $\mathbf{x}M_S M_1 M_T^* = (\mu + \lambda M_1)M_T^*$, we have $C_2 \cap f(C_1) = \emptyset$. If there are solutions, they will lie in a coset $\nu + G$ of F_2^c; let M_G be the $r \times c$ matrix giving a basis for G. The subset of C_1 consisting of points $\mathbf{x}$ with $f(\mathbf{x}) \in C_2$ is the coset $\lambda' + S'$ of F_2^a, where $\lambda' = \lambda + \nu M_S$, and $M_S' = M_G M_S$ is an $r \times a$ matrix whose rows generate S'.

2.4.1 The value of $[E(\mathbb{Q}_p) : 2E(\mathbb{Q}_p)]$

Lutz discovered that an elliptic curve defined over a local field $K_{\mathfrak{p}}$ has a subgroup M of finite index isomorphic to $\mathcal{O}_{K_{\mathfrak{p}}}$. If we consider the exact sequence

$$0 \longrightarrow M \longrightarrow E(K_{\mathfrak{p}}) \longrightarrow G \longrightarrow 0$$

with G the finite quotient group $E(K_\mathfrak{p})/M$, combine it with the multiplication-by-two map on each of its terms to get



and then apply the snake lemma, we obtain the long exact sequence of finite groups

$$0 \to M[2] \to E(K_\mathfrak{p})[2] \to G[2] \to M/2M \to E(K_\mathfrak{p})/2E(K_\mathfrak{p}) \to G/2G \to 0.$$

$M \simeq \mathcal{O}_{K_\mathfrak{p}}$ is torsion-free and so $\|M[2]\| = 1$; G is finite and so $\|G[2]\| = \|G\|/2G$. Working through the alternating product of sizes, we have

$$\|E(K_\mathfrak{p})/2E(K_\mathfrak{p})\| = [M : 2M] \, \|E(K_\mathfrak{p})[2]\| .$$

This result is described as "classical" in [8]. As obvious corollaries, we have

Lemma 2.4.2.

$$\|E(K_\mathfrak{p})/2E(K_\mathfrak{p})\| = \left\|\mathcal{O}_{K_\mathfrak{p}}/2\mathcal{O}_{K_\mathfrak{p}}\right\| \|E(K_\mathfrak{p})[2]\|$$

and, specialising further to $K_\mathfrak{p} = \mathbb{Q}_p$, $M = \mathbb{Z}_p$,

Lemma 2.4.3. Let E be an elliptic curve defined over $\mathbb{Q}$, given by $y^2 = f(x)$. Let ℓ be one plus the number of linear factors of f over $\mathbb{Q}_p$. Then

$$\|E(\mathbb{Q}_p)/2E(\mathbb{Q}_p)\| = \begin{cases} 2\ell & p = 2 \\ \ell & p \neq 2 \end{cases}.$$

2.4.2 The size of $\operatorname{im}\mathcal{C}(\mathbb{Q}_p)$ in $L_p^\times/\mathbb{Q}_p^\times(L_p^\times)^2$

Recall lemma 2.2.3:

Lemma 2.4.4. Let $\mathcal{C} : y^2 = f(x)$ be a two-covering. Then $P_{\infty^-} \in 2[\mathcal{C}(K), P_{\infty^+}]$ if and only if f has a root in K.

Theorem 2.4.1. Let $\mathcal{C} : y^2 = f(x)$ be a two-covering of the elliptic curve $E : y^2 = g(x)$. Let R_p be the image of $\mathcal{C}(\mathbb{Q}_p)$ in $L_p^\times/\mathbb{Q}_p^\times(L_p^\times)^2$ under the map $(x, y) \to x - \alpha$ of theorem 2.2.3. Then the size N_p of R_p is $v[E(\mathbb{Q}_p) : 2E(\mathbb{Q}_p)]$, where $v = 1$ if $f(x)$ has a root in $\mathbb{Q}_p$, and $v = \frac{1}{2}$ otherwise, where the value of $[E(\mathbb{Q}_p) : 2E(\mathbb{Q}_p)]$ is given explicitly in lemma 2.4.3.

Proof. Let α be a root of f in $\overline{\mathbb{Q}_p}$, and let $K_p = \mathbb{Q}_p(\alpha)$. From theorem 2.2.3, we have

$$R_p \simeq \frac{\mathcal{C}(\mathbb{Q}_p)}{E(\mathbb{Q}_p) \cap 2E(K_p)},$$

and by lemma 2.2.10 and the result before it, this coset space has size equal to $[E(\mathbb{Q}_p) : E(\mathbb{Q}_p) \cap 2E(K_p)]$.

Since

$$2E(\mathbb{Q}_p) \subset E(\mathbb{Q}_p) \cap 2E(K_p) \subseteq E(\mathbb{Q}_p)$$

we have

$$\begin{aligned}[E(\mathbb{Q}_p) : 2E(\mathbb{Q}_p)] &= [E(\mathbb{Q}_p) : E(\mathbb{Q}_p) \cap 2E(K_p)] \times [E(\mathbb{Q}_p) \cap 2E(K_p) : 2E(\mathbb{Q}_p)] \\ &= N_p \times [E(\mathbb{Q}_p) \cap 2E(K_p) : 2E(\mathbb{Q}_p)]\end{aligned}$$

Now, lemma 2.2.8 tells us that the rightmost term is 1 if $f(x)$ has a root in $\mathbb{Q}_p$, and 2 otherwise; rearranging terms gives the claimed result.

□

Note that N_p can never be larger than eight, and $N_p = 8$ is not particularly common. We require $p = 2$, and need a quartic, irreducible over $\mathbb{Q}$, which splits into linear factors over $\mathbb{Q}_2$, and whose associated cubic is also irreducible over $\mathbb{Q}$ but splits into linear factors over $\mathbb{Q}_2$;

$$y^2 = 2x^4 - 3x^3 - 15x^2 + 38x - 16$$

is such a two-covering. The rarity arises because neither of the necessary splitting patterns is a particularly common one for random polynomials over $\mathbb{Q}_2$ of the relevant degrees.

2.4.3 Checking at the infinite prime

Every element of $\mathbb{C}$ is a square, but the negative elements of $\mathbb{R}$ are not squares; this gives a test for local solvability at the infinite prime, which is essential since there do exist two-coverings, representing elements of Ш[2], which have four-descendents solvable at all but the infinite prime.

For a non-trivial two-covering, the quartic f in $ay^2 = f(x)$ will have zero, two or four distinct real roots; if it has no real roots we cannot gain anything at this stage.

So, assume f has n real roots, $r_1 < \ldots < r_n$. We have n embeddings $E_i : L \to \mathbb{R}$ given by sending θ to r_i; given an $\varepsilon \in L^\times/(L^\times)^2$ which we think may be an element of $\mathfrak{D}_4^{\mathrm{alg}}$, consider the signs of $E_i(\varepsilon)$. Certainly the product of these signs will have the same sign as a if $\varepsilon \in \mathfrak{D}_4^{\mathrm{alg}}$: this is the only condition we can use if $n = 2$.

If $n = 4$ and f is irreducible (so $L = K(\theta)$ for $f(\theta) = 0$), note that, if $\varepsilon \in \mathfrak{D}_4^{\mathrm{alg}}$ and therefore can be written as $\varepsilon = (x + y\theta)z^2$, then the signs of $E_i(\varepsilon)$ will be those of $x + yr_i$, since $E_i(z^2)$ will always be a square, hence positive. Since the r_i are in ascending order, the $x + yr_i$ will be monotonically increasing or decreasing. So the sign can change at most once – that is, if a is positive, the signs for an element of the right form must be $++++$ or $++--$ or $--++$ or $----$, and if a is negative, they must be $+++-$ or $---+$ or $+---$ or $-+++$.

Let M_1 be the $n \times 4$ matrix over $\mathbb{F}_2$ giving the signs of the four embeddings for all the elements of `RB`; we map the ± 1 to $\mathbb{F}_2$ by sending $+1 \to 0, -1 \to 1$. The set of permissible signs for a given leading coefficient is a coset $\mu + T$ of a two-dimensional subspace T of ${\mathbb{F}_2}^4$: let M_T be the 2×4 matrix giving a basis for this subspace, and apply lemma 2.4.1.

To be absolutely explicit,

$$\mu = \begin{cases} (0\,0\,0\,0) & a > 0 \\ (0\,0\,0\,1) & a < 0 \end{cases}; \quad M_T = \begin{cases} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix} & a > 0 \\ \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} & a < 0 \end{cases}$$

2.4.4 Checking at finite primes

To check solubility at a finite prime, we compute the image R_p of $\mathcal{C}(\mathbb{Q}_p)$ in the quotient $L_p/\mathbb{Q}_p(L_p)^2$, where L_p is the localised algebra of section 1.6, and then see which elements of $\mathfrak{D}_4^{\mathrm{alg}}$ lie in this image under the localisation map of that section. We know the size of the image by theorem 2.4.1, and we have the following lemma:

Lemma 2.4.5. The set $\mathbb{Q}_p^2$ has measure ρ in $\mathbb{Q}_p$, where ρ is $1/6$ if $p = 2$ and $\frac{p}{2p+2}$ otherwise.

Proof. A non-zero square in $\mathbb{Q}_p$ is an element w with even valuation v, such that $p^{-v}w$ is a square unit. By standard results involving quadratic residues, the density of squares among the units of $\mathbb{Z}_p$ is one half, unless $p = 2$, where it is one quarter. A random element of $\mathbb{Z}_p$ will have valuation 0 with probability $1 - p^{-1}$, 1 with probability $p^{-1} - p^{-2}$, and so on. So the probability of even valuation is $1-p^{-1}+p^{-2}-p^{-3}\ldots$, or $\frac{p}{p+1}$. Multiplying these probabilities gives the indicated result. □

The lemma suggests that we can expect to find points on $\mathcal{C}(\mathbb{Q}_p)$ fairly easily, at least if we can assume that $f(x) : x \in \mathbb{Q}_p$ is a randomly-distributed collection of elements of $\mathbb{Q}_p$. This is not a safe assumption, but experiment indicates that running through elements of $p^{-t}\mathbb{Z}_p$ for various small t does normally produce enough points to fill out R_p very quickly.

We have a convenient `magma` function `MakeModSquares`, which takes a number field K and a prime ideal $\mathfrak{p}$ of K, and gives an $\mathbb{F}_2$ vector space $G_\mathfrak{p} \simeq K_\mathfrak{p}^\times/(K_\mathfrak{p}^\times)^2$ of dimension $d_\mathfrak{p}$ and a map $\phi_\mathfrak{p} : K^\times \to G$. Using this function with $K = \mathbb{Q}[x]/(f(x))$ allows us to construct $G_p \simeq L_p^\times/(L_p^\times)^2$ where

$$L_p^\times = \bigoplus_{\mathfrak{p}_i|p} L_{\mathfrak{p}_i}^\times$$

by letting G_p be an $\mathbb{F}_2$ vector space of dimension $\sum_{\mathfrak{p}|p} d_\mathfrak{p}$, and (since the base algebra L is equal to the number field K because f was irreducible over $\mathbb{Q}$) we can construct $\psi_1 : L^\times \to G_p$ as the concatenation of the $\mathbb{F}_2$-vectors $\phi_\mathfrak{p}$ for all $\mathfrak{p}$ dividing p.

However, the image of the Cassels map μ actually lives in $L_p^\times/\mathbb{Q}_p^\times(L_p^\times)^2$; this group is isomorphic to the quotient of G_p by the image in it of $\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$, so we use `magma`'s quotient facilities to quotient by the group $\langle\psi_1(t_i)\rangle$ for $T = \{t_i\}$ a generating set of $\mathbb{Q}_p^\times/(\mathbb{Q}_p^\times)^2$ (we take $T = \{2, 3, 5\}$ for $p = 2$, and $T = \{p, q\}$, where q is the smallest positive integer not a quadratic residue mod p, otherwise).

Let H be this quotient vector space, and define its dimension as k; let $\tau : G_p \to H$ be the natural quotient map, and $\psi(x) = \tau(\psi_1(x))$. With this tool, and our prior knowledge of N_p, we can build the set R_p fairly easily. Since f did not have any factors over $\mathbb{Q}$, we can be sure that any $x \in \mathbb{Q}$ will not have $x = \alpha_{\mathfrak{p}_i}$ for any of the $L_{\mathfrak{p}_i}$, so the Cassels map $\mu : \mathcal{C}(\mathbb{Q}_p) \to L_p^\times$ is simply the generalised $x \to [x - \alpha_1, \ldots, x - \alpha_r]$ of section 1.6, which with the definitions above is just $\psi(x - \alpha)$ where α is the root of f in K.

To construct the set, we simply test, for values of $x \in \mathbb{Q}$ starting at 0 and going up by some δ at each step, whether $f(x)$ is a non-zero p-adic square; if

so, compute $\psi(x-\alpha)$, see if it is in the set, and add it if it is not. Stop once we have N_p elements in the set.

An obvious refinement is to take advantage of the fact that R_p is of the form $\mu + T$ where μ is some unknown element and T is an exponent-2 subgroup of known degree $e = \log_2 N_p$). So, the first element of R_p that we encounter we use as μ; we then maintain a list of generators, initially empty. When we find another element b, we compute $b - \mu$ and check if it can be written as a sum of some subset of the generators; if not, we add it as a generator. When we have found e generators, stop.

The unrefined process will terminate iff every orbit of $\mathcal{C}(\mathbb{Q}_p)$ under the action of $2E(\mathbb{Q}_p)$ contains an element (x, y) with x/δ integral, though we have no guarantee how long it may take. The refined process requires only that a generating set of (x, y) with $x/\delta \in \mathbb{Z}$ exists.

It is not clear how to pick a value of δ which will always work. After some experiment, I use a two-stage process: start with `R=250`, $\delta = 1$, `dellim = 4`. Search $[0, Rp)$ in steps of size 1, $[Rp, 2Rp)$ in steps of size p^{-1}, $[2Rp, 3Rp)$ in steps of size p^{-2}, and so on until the steps have reached size $p^{-\mathrm{dellim}}$. Then increase R by a factor p^4, increase `dellim` by one, and start again at $[0, Rp)$.

The need for this elaborate process can be seen in examples like

$$f(x) = 2x^4 + 4x^3 + 237x^2 + 620x + 6772$$

where the x with $f(x)$ a 2-adic square are all of the form $x = -3238 + O(2^{15})$, and the x found to generate the two classes are 29530 and 62298; it takes something like 35 seconds to find these. Over several thousand runs of the four-descent machinery, it emerges that the problematic prime is almost always 2, although the most troublesome curve I encountered was

$$y^2 = 5x^4 + 4x^3 - 4306x^2 + 67452x - 299959$$

which behaves unpleasantly over $\mathbb{Q}_{11}$.

Once we have generated μ and T, we let M_1 be the $n \times k$ matrix whose rows are the images $\psi(\varepsilon)$ in H of the elements of `RB`, and apply lemma 2.4.1 to (M_1, μ, M_T).

2.5 Minimisation of four-coverings

It turns out that minimisation of four-coverings requires only one more step than the minimisation of two-coverings which has been routine since Birch and Swinnerton-Dyer's paper, and which is very completely described in [71].

The critical, new step is

Theorem 2.5.1. Given a prime p, and an integral four-covering $[A, B]$ possessing a point over $\mathbb{P}^3(\mathbb{Q}_p)$ and with $p^2 | \det(Ax+By)$, there exists an integral four-covering $[A', B']$ equivalent to $[A, B]$ with $\det(A'x + B'y) = p^{-2} \det(Ax + By)$ or $p^{-2} \det(Ay + Bx)$. Also, there is an effective method to construct this.

Note that, for $p \geq 5$, the condition of [71] is that $\mathrm{val}_p I < 4$ or $\mathrm{val}_p J < 6$, where I and J are the covariants of the quartic $Ax + By$; as a result, p^{12} does not divide $\det q$. For $p = 2, 3$ the minimisation is more complicated, and it is possible that $p^{12} | \det q$ for a p-minimal quartic, though not that $p^{24} | \det q$.

Given this theorem, we define an integral four-covering $[A, B]$ as p-minimised if $\det(Ax + By)$ is a p-minimal quartic. Being p-minimised is not as strong a condition as being p-minimal in the sense of section 1.4.1, but the conditions are equivalent for $p \geq 5$, since transformations $(\alpha, I_2, I_4) \in A_{\mathfrak{D}_4}$ change $\mathrm{val}_p \Delta([A, B])$ by a multiple of 12, and a p-minimised $[A, B]$ will have $\mathrm{val}_p \Delta([A, B]) < 12$. Moreover, we have

Theorem 2.5.2. If $[A, B]$ is p-minimal, it is p-minimised.

Proof. Suppose not, so there exists a p-minimal four-covering with $\det(Ax+By)$ not minimal at p. The algorithm in [71] can easily be converted to take a quartic $f(x, y)$ and return

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}) : \det M = p^r$$

and e with $2e > r$ and $p^{-2e} f(ax+by, cx+dy)$ minimal: in any case where one of the variables is to be divided by some p^f, we instead multiply the other variable by p^f and add $2f$ to e. Do this to $f = \det(Ax + By)$, then replace $[A, B]$ with $[aA+bB, cA+dB]$, which multiplies each coefficient in $\det(Ax+By)$ by p^r. We can now apply theorem 2.5.1 e times to bring the multiplier back down.

After this process, we have an element of $A_{\mathfrak{D}_4}$ whose action reduces the power of p dividing $\det(Ax + By)$; which is a contradiction since $[A, B]$ was formerly assumed p-minimal. □

2.5.1 Constructive proof of theorem 2.5.1

To prove theorem 2.5.1 involves considering a number of cases, in each of which we construct a member of $A_{\mathfrak{D}_4}$ which brings the determinant down.

Let $[A, B]$ be an integral four-covering, let $q = \det(Ax+By)$ be the associated quartic, and let p be a prime with p^2 dividing the content of q. Let $\tilde{M}$ be the

image of the matrix $Ax + By$ in $M_4(\mathbb{F}_p[x, y])$, and r_{gen} be the rank of $\tilde{M}$ over $\mathbb{F}_p[x, y]$: let $\tilde{A}$ and $\tilde{B}$ be the images of A and B in $M_4(\mathbb{F}_p)$.

We cannot have $r_{\text{gen}} = 4$, since $\det \tilde{M} \equiv q \pmod{p}$, and by hypothesis $q \equiv 0 \pmod{p}$; so $\tilde{M}$ has zero determinant and must be singular.

Lemma 2.5.1. If $r_{\text{gen}} < 4$, then there can be at most r_{gen} matrices in the pencil with rank less than r_{gen}, and none with rank $> r_{\text{gen}}$.

Proof. If a matrix has rank s, all the $s+1 \times s+1$ minors must have zero determinant; also, at least one $s \times s$ minor must have non-zero determinant. Now, $\tilde{M}$ is of rank r_{gen}, so has at least one $r_{\text{gen}} \times r_{\text{gen}}$ submatrix whose determinant is a *non-zero* sum of products of precisely r_{gen} linear terms – that is, a homogenous degree-r_{gen} polynomial $f(x, y)$.

The determinant of the corresponding submatrix in $\lambda A + \mu B$ is then $f(\lambda, \mu)$; for the rank to be lower than r_{gen}, this must be zero, which can happen in at most r_{gen} ways.

Suppose $M = \lambda\tilde{A} + \mu\tilde{B}$ had rank r with $r > r_{\text{gen}}$. Then there is an $r \times r$ minor of M with non-zero determinant. But the corresponding $r \times r$ minor of $\tilde{M}$ has determinant identically zero, and M is obtained by specialising $x = \lambda, y = \mu$ in $\tilde{M}$. So this cannot happen. □

Lemma 2.5.2. Given an m-dimensional subspace of $\mathbb{F}_p^n$ spanned by the m linearly independent vectors $\langle V_1, \ldots, V_m \rangle$, there exists a matrix in $\mathrm{SL}_n(\mathbb{Z})$ which transforms the subspace into one spanned by vectors whose images are the first m standard basis vectors mod p.

Proof. Note that a matrix with the desired property must exist in $\mathrm{SL}_n(\mathbb{F}_p)$, and use the result of lemma 1.38 of [59], namely that the map $\mathrm{SL}_n(\mathbb{Z}) \to \mathrm{SL}_n(\mathbb{Z}/N\mathbb{Z})$ is surjective for all $n \geq 1$, $N \geq 1$. □

It turns out that an essential operation for one of the cases of minimisation is the *flip*:

Definition 2.5.1. Let A and B be two symmetric matrices in $M_4(\mathbb{Z})$ such that the top left 2×2 submatrices mod p are the zero matrix.

Let A' and B' be obtained from A and B respectively by dividing the entries in the top left 2×2 submatrix by p, and multiplying those in the bottom right 2×2 submatrix by p; then $[A', B']$ is the flip of $[A, B]$, and is integral.

Lemma 2.5.3 (Flipping Lemma). $[A, B]$ and $[A', B']$ are equivalent four-coverings, and, if $[A, B]$ has the point $\mathbf{x} = (x_1, x_2, x_3, x_4)$, then $[A', B']$ has the point $\mathbf{x}' = (px_1, px_2, x_3, x_4)$.

Proof. The flipping operation is equivalent to multiplying both matrices by p, and then performing a change of variables to replace x_1 and x_2 by px_1 and px_2 respectively; both these transforms send four-coverings to equivalent ones, and the latter clearly acts on points in the way claimed in the lemma.

The same argument makes it clear that $\det(Ax+By) = \det(A'x+B'y)$. □

Nearly all the subcases of the proof will involve transforming $[A, B]$ so that the kernel of some matrix lies along the standard basis vectors; to save space, make the following definition:

Definition 2.5.2. Let $e_1 = (1,0,0,0), e_2 = (0,1,0,0), e_3 = (0,0,1,0), e_4 = (0,0,0,1)$ be the standard basis vectors for $\mathbb{F}_p^4$.

Proof of theorem 2.5.1. The proof proceeds case by case. The arguments almost all involve careful considerations of the powers of p dividing the entries of a matrix; we write, for example, $a_{11}^{(2)}$ for $a_{11}p^{-2}$, once we have demonstrated that p^2 divides a_{11} exactly.

2.5.2 $r_{\text{gen}} = 0$

If $r_{\text{gen}} = 0$ then $\tilde{M}$ is the zero matrix; so $\tilde{A} = \tilde{B} = 0$. So p^4 must divide D, and $[A', B'] = [p^{-1}A, p^{-1}B]$ will be an equivalent four-covering with $D' = p^{-4}D$.

2.5.3 $r_{\text{gen}} = 1$

Lemma 2.5.4. If $r_{\text{gen}} = 1$, then every matrix in the pencil modulo p is of the form $\lambda\tilde{M}$ for $\lambda \in \mathbb{F}_p$ and $\tilde{M}$ some matrix of rank 1 with entries from $\mathbb{F}_p$ ($\lambda = 0$ is permitted).

Proof. $\tilde{C} = \tilde{A}x + \tilde{B}y$ is a matrix of linear forms; say $\tilde{C} = (c_{ij})$. Because $\tilde{C}$ is of rank 1, we have $c_{ii}c_{jj} = c_{ij}^2$, so in particular if $c_{ii} = 0$ the whole ith row and column are zero. If c_{ii} and c_{jj} are not both zero, then $c_{ii}c_{jj} = c_{ij}^2$, and by unique factorisation in $\mathbb{F}_p[x, y]$ we have $c_{ii} = c_{ij} = c_{jj}$ up to a scalar multiple. That is, all the non-zero elements of $\tilde{C}$ must be equal up to a scalar multiple. □

Thus there is a common rank-three kernel for every non-zero matrix in the pencil; change variables so this kernel is spanned by $\langle e_1, e_2, e_3\rangle$, and then write $x_4' = p^{-1}x_4$, which has the effect of replacing $[A, B]$ with $[PAP^T, PBP^T]$ where $P = \text{diag}(1,1,1,p)$. This causes every matrix in the pencil to be congruent to the zero matrix mod p, whilst multiplying the associated quartic q by p^2. But now we can divide both A and B by p, which divides q by p^4 and has removed the desired power p^2.

2.5.4 $r_{\text{gen}} = 2$

In this case, there must be a non-trivial common kernel; if there were not, we could change variables such that $\ker \tilde{A} = \langle e_1, e_2 \rangle$ and $\ker \tilde{B} = \langle e_3, e_4 \rangle$. But then $\tilde{A}$ occupies the top 2×2 quadrant, $\tilde{B}$ occupies the bottom 2×2 quadrant, and their sum would have rank four.

We have the following lemma:

Lemma 2.5.5 (A consequence of local solvability). Let $[A, B]$ be a four-covering with a point $\mathbf{z} = (z_1\ z_2\ z_3\ z_4)$ over $¶^3(\mathbb{Q}_p)$, such that

$$M = Ax + By = \begin{pmatrix} pC_1 & pX \\ pX^{\mathrm{T}} & C_2 \end{pmatrix} = \begin{pmatrix} p(C_{11}x + C_{12}y) & pX \\ pX^{\mathrm{T}} & C_{21}x + C_{22}y \end{pmatrix}.$$

Let $\mathbf{z}$ be written such that at least one of its components is a p-adic unit.

Then, for at least one of $C = C_1$ and $C = C_2$, there is an element $\mathbf{u}$ of $\mathbb{F}_p^2$, not equal to $(0\ 0)$, with $\mathbf{u}C\mathbf{u}^{\mathrm{T}} = 0$.

Proof. Suppose not, so $\mathbf{u}C_i\mathbf{u}^{\mathrm{T}} \equiv 0 \pmod{p} \implies \mathbf{u} \equiv (0\ 0) \pmod{p}$.

Write $\mathbf{z}_1 = (z_3\ z_4)$ and $\mathbf{z}_2 = (z_1\ z_2)$; then we have $\mathbf{z}M\mathbf{z}^{\mathrm{T}} \equiv \mathbf{z}_1 C_2 {\mathbf{z}_1}^{\mathrm{T}} \pmod{p}$.

And so $\mathbf{z}_1 \equiv (0\ 0) \pmod{p}$; say $\mathbf{z}_1 = (pz_3'\ pz_4')$.

But consider $[A', B']$, the flip of $[A, B]$. We know by 2.5.3 that $\mathbf{z}' = (pz_1\ pz_2\ pz_3'\ pz_4')$ is a point on $[A', B']$; and, since we're working in projective space, so must $\mathbf{z}'' = (z_1\ z_2\ z_3'\ z_4')$ be.

However, $\mathbf{z}''M{\mathbf{z}''}^{\mathrm{T}} \equiv \mathbf{z}_2 C_1 {\mathbf{z}_2}^{\mathrm{T}} \pmod{p}$, which, since we assumed that quadratic form also only had trivial points mod p, implies that $\mathbf{z}_2 \equiv (0\ 0) \pmod{p}$. Hence p divides all four elements of $\mathbf{z}$, and we have a contradiction of the assumption that at least one of those elements is a unit.

So there must be a non-zero solution $\mathbf{u}$ to at least one of the quadratics $\mathbf{u}C_i\mathbf{u}^{\mathrm{T}} = 0$.

□

Case 1: $\tilde{A}$ and $\tilde{B}$ have a common two-dimensional kernel

Change variables so that the common kernel is $\langle e_1, e_2 \rangle$. In these coordinates, $xA + yB$ will be of the shape

$$\begin{pmatrix} pC_1 & pX \\ pX^{\mathrm{T}} & C_2 \end{pmatrix}$$

that we saw in lemma 2.5.5.

If C_2 represents zero non-trivially – that is, when we write $C_2 = xD + yE$, we find $\exists \mathbf{x} = (x_1\ x_2) \neq (0\ 0) \pmod{p}$ with $\mathbf{x}D\mathbf{x}^{\mathrm{T}} = \mathbf{x}E\mathbf{x}^{\mathrm{T}} \equiv 0 \pmod{p}$ – then we can consider an element

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

such that $\alpha \equiv x_1 \pmod{p}, \beta \equiv x_2 \pmod{p}$, and transform $[A, B]$ by the matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \alpha & \beta \\ 0 & 0 & \gamma & \delta \end{pmatrix}.$$

This produces a 3×3 submatrix of zeroes in the top left-hand corner of $\tilde{A}$ and $\tilde{B}$, and then transforming by $(p^{-1}, I_2, \mathrm{diag}([1, 1, 1, p]))$ will replace $[A, B]$ by an $[A', B']$ with $\det(A'x + B'y) = p^{-2} \det(Ax + By)$.

To check whether C_2 represents zero non-trivially, we look for a common root to the quadratic polynomials $\mathbf{x}M_i\mathbf{x}^{\mathrm{T}}$ in $\mathbb{F}_p[x, y]$, where M_i are the bottom 2×2 submatrices of A_1 and B_1. This fails at the prime 2, since $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is non-zero but the associated quadratic is identically zero; however, in that case $[0, 1]$ is a common point.

If C_2 represents zero only trivially, perform a flip on $[A_1, B_1]$ to get $[A_2, B_2]$ such that $xA_2 + yB_2$ is of the shape

$$\begin{pmatrix} C_1 & pX \\ pX^{\mathrm{T}} & pC_2 \end{pmatrix}.$$

Since $[A, B]$ is assumed solvable at p, by lemma 2.5.5 we must have C_1 representing zero non-trivially, and we proceed as above.

Case 2 : $\dim(\ker \tilde{A} \cap \ker \tilde{B}) = 1$

Change variables so that the common kernel is $\langle e_1 \rangle$, the additional generator of $\ker A$ is e_2 and that of $\ker B$ is e_3. So we have

$$A = \begin{pmatrix} pa_{11}^{(1)} & pa_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ pa_{12}^{(1)} & pa_{22}^{(1)} & pa_{23}^{(1)} & pa_{24}^{(1)} \\ pa_{13}^{(1)} & pa_{23}^{(1)} & a_{33} & a_{34} \\ pa_{14}^{(1)} & pa_{24}^{(1)} & a_{34} & a_{44} \end{pmatrix}, B = \begin{pmatrix} pb_{11}^{(1)} & pb_{12}^{(1)} & pb_{13}^{(1)} & pb_{14}^{(1)} \\ pb_{12}^{(1)} & b_{22} & pb_{23}^{(1)} & b_{24} \\ pb_{13}^{(1)} & pb_{23}^{(1)} & pb_{33}^{(1)} & pb_{34}^{(1)} \\ pb_{14}^{(1)} & b_{24} & pb_{34}^{(1)} & b_{44} \end{pmatrix}$$

where the submatrices

$$A_1 = \begin{pmatrix} a_{33} & a_{34} \\ a_{34} & a_{44} \end{pmatrix}, B_1 = \begin{pmatrix} b_{22} & b_{24} \\ b_{24} & b_{44} \end{pmatrix}$$

have determinant $\neq 0 \pmod{p}$.

Naturally $p|\det(Ax+By)$ since the whole first row and column are divisible by p. *Ex hypothesi*, $p^2|\det(Ax+By)$; modulo p^2, we have $\det(Ax+By) \equiv pg(x,y)$ where

$$\begin{aligned} g(x,y) = & \quad a_{11}^{(1)} b_{22} \det A_1 x^3 y + \left(a_{11}^{(1)} a_{33} \det B_1 + b_{11}^{(1)} b_{22} \det A_1 \right) x^2 y^2 \\ & + a_{33} b_{11}^{(1)} \det B_1 x y^3 \\ = & \quad xy \left(a_{11}^{(1)} x + b_{11}^{(1)} y)(b_{22} \det A_1 x + a_{33} \det B_1 y \right) \end{aligned}$$

We have unique factorisation in $\mathbb{F}_p[x,y]$, so one of the two bracketed terms must be zero; thus either $a_{11}^{(1)} = b_{11}^{(1)} = 0$ or $b_{22} = a_{33} = 0$.

If $a_{11}^{(1)} \equiv b_{11}^{(1)} \equiv 0$ then we can apply $\left(p^{-2}, I_2, \mathrm{diag}(1,p,p,p)\right)$ and remove a factor p^2. If $b_{22} = a_{33} = 0$ then we can apply $\left(p^{-1}, I_2, \mathrm{diag}(1,1,1,p)\right)$, in aggregate removing a factor p^2.

2.5.5 $r_{\mathrm{gen}} = 3$

If the generic rank is three and there is no common kernel, then at least one of $\tilde{A}$ and $\tilde{B}$ must have rank three, since otherwise we could transform to make $\ker \tilde{A} = \langle e_1, e_2 \rangle$ and $\ker \tilde{B} = \langle e_3, e_4 \rangle$ and $\tilde{A} + \tilde{B}$ would have rank four.

Swapping A and B has the effect of reversing the order of the coefficients of the associated quartic; however, the definition of minimality in Appendix A of [71] indicates that this will have no effect on its minimality. So we can assume that $\tilde{A}$ has rank three; $\tilde{B}$ then has rank two or three, since $\tilde{B}$ of rank one could be written with $\ker \tilde{B} = \langle e_1, e_2, e_3 \rangle$, forcing $\ker \tilde{A} = \langle e_4 \rangle$ and then $\tilde{A} + \tilde{B}$ has rank four. So we can decompose into three cases:

Case 1: $r_{\mathrm{gen}} = 3$**,** $\ker \tilde{A} \cap \ker \tilde{B} \neq \{\mathbf{0}\}$

Transform so that the common kernel is $\langle e_4 \rangle$. We have

$$C = \tilde{A}x + \tilde{B}y \equiv \begin{pmatrix} & & & 0 \\ & C_1 & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

where the entries of C_1 are linear forms in integers, and $\det C_1 \not\equiv 0 \pmod p$ because $\operatorname{rank} C = 3$. That is,

$$Ax + By = \begin{pmatrix} & & & pd_1 \\ & C_1 & & pd_2 \\ & & & pd_3 \\ pd_1 & pd_2 & pd_3 & pd_4 \end{pmatrix}$$

Now, the associated quartic $q = \det(Ax + By)$ is congruent to $pd_4 \det(C_1) \pmod{p^2}$. And we know $q \equiv 0 \pmod{p^2}$, so we have $d_4 \equiv 0 \mod p$, and applying $\left(p^{-2}, Id_2, \operatorname{diag}(p, p, p, 1)\right)$ will remove a factor p^2 from q.

Case 2: $r_{\text{gen}} = 3$**,** $\operatorname{rank} \tilde{A} = \operatorname{rank} \tilde{B} = 3$**, no common kernel**

Perform a unimodular transformation to make $\ker \tilde{A} = \langle e_1 \rangle$ and $\ker \tilde{B} = \langle e_2 \rangle$; so

$$A = \begin{pmatrix} pa_{11}^{(1)} & pa_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ pa_{12}^{(1)} & a_{22} & a_{23} & a_{24} \\ pa_{13}^{(1)} & a_{23} & a_{33} & a_{34} \\ pa_{14}^{(1)} & a_{24} & a_{34} & a_{44} \end{pmatrix}, B = \begin{pmatrix} b_{11} & pb_{12}^{(1)} & b_{13} & b_{14} \\ pb_{12}^{(1)} & pb_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ b_{13} & pb_{23}^{(1)} & b_{33} & b_{34} \\ b_{14} & pb_{24}^{(1)} & b_{34} & b_{44} \end{pmatrix}.$$

Define submatrices

$$A_1 = \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ a_{23} & a_{33} & a_{34} \\ a_{24} & a_{34} & a_{44} \end{pmatrix}, B_1 = \begin{pmatrix} b_{11} & b_{13} & b_{14} \\ b_{13} & b_{33} & b_{34} \\ b_{14} & b_{34} & b_{44} \end{pmatrix}.$$

Since the ranks of $\tilde{A}$ and $\tilde{B}$ are both 3, $\det A_1 \not\equiv 0 \pmod p$ and likewise for B_1.

As always, we have $\det(Ax + By) \equiv 0 \pmod{p^2}$, and so all its coefficients are $0 \pmod{p^2}$; modulo p, the coefficient of x^3y in $q \mod p$ is $b_{11} \det A_1$ and the coefficient of xy^3 is $a_{22} \det B_1$, so we must have $a_{22} \equiv b_{11} \equiv 0 \pmod p$.

With those conditions, we have

$$\det Ax + By = \left(a_{23}b_{14} - a_{24}b_{13}\right)^2 x^2y^2 \pmod p,$$

so $a_{23}b_{14} \equiv a_{24}b_{13} \pmod p$, and the vectors $\begin{pmatrix} a_{23} \\ a_{24} \end{pmatrix}$ and $\begin{pmatrix} b_{13} \\ b_{14} \end{pmatrix}$ are proportional. They are not both the zero vector, because that would make $\det A_1 \equiv \det B_1 \equiv 0$, so, by a unimodular transformation affecting only rows 3 and 4, we can arrange $a_{24} \equiv b_{14} \equiv 0$.

So, without loss of generality, the matrices look like

$$A = \begin{pmatrix} pa_{11}^{(1)} & pa_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ pa_{12}^{(1)} & pa_{22}^{(1)} & a_{23} & pa_{24}^{(1)} \\ pa_{13}^{(1)} & a_{23} & a_{33} & a_{34} \\ pa_{14}^{(1)} & pa_{24}^{(1)} & a_{34} & a_{44} \end{pmatrix}, B = \begin{pmatrix} pb_{11}^{(1)} & pb_{12}^{(1)} & b_{13} & pb_{14}^{(1)} \\ pb_{12}^{(1)} & pb_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ b_{13} & pb_{23}^{(1)} & b_{33} & b_{34} \\ pb_{14}^{(1)} & pb_{24}^{(1)} & b_{34} & b_{44} \end{pmatrix}.$$

We now perform a "flip" operation; multiply x_3 and x_4 by p and divide the whole matrix by p. After this operation, the matrices look like

$$A' = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ a_{12}^{(1)} & a_{22}^{(1)} & a_{23} & pa_{24}^{(1)} \\ pa_{13}^{(1)} & a_{23} & pa_{33} & pa_{34} \\ pa_{14}^{(1)} & pa_{24}^{(1)} & pa_{34} & pa_{44} \end{pmatrix}, B' = \begin{pmatrix} b_{11}^{(1)} & b_{12}^{(1)} & b_{13} & pb_{14}^{(1)} \\ b_{12}^{(1)} & b_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ b_{13} & pb_{23}^{(1)} & pb_{33} & pb_{34} \\ pb_{14}^{(1)} & pb_{24}^{(1)} & pb_{34} & pb_{44} \end{pmatrix}.$$

By the flipping lemma, $\det(Ax+By) = \det(A'x+B'y)$, so $p^2 | \det(A'x+B'y)$. Now, $\langle e_4 \rangle$ is a common kernel of the two matrices; hence, we are in a situation with a common kernel, and we have already shown that in such a situation we can remove a factor p^2.

Case 3: $\operatorname{rank} \tilde{A} = 3$**,** $\operatorname{rank} \tilde{B} = 2$**, no common kernel**

Change basis such that $\ker \tilde{A} = \langle e_1 \rangle$ and $\ker \tilde{B} = \langle e_2, e_3 \rangle$.

We have

$$A = \begin{pmatrix} pa_{11}^{(1)} & pa_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ pa_{12}^{(1)} & a_{22} & a_{23} & a_{24} \\ pa_{13}^{(1)} & a_{23} & a_{33} & a_{34} \\ pa_{14}^{(1)} & a_{24} & a_{34} & a_{44} \end{pmatrix}, B = \begin{pmatrix} b_{11} & pb_{12}^{(1)} & pb_{13}^{(1)} & b_{14} \\ pb_{12}^{(1)} & pb_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ pb_{13}^{(1)} & pb_{23}^{(1)} & pb_{33}^{(1)} & pb_{34}^{(1)} \\ b_{14} & pb_{24}^{(1)} & pb_{34}^{(1)} & b_{44} \end{pmatrix};$$

define A_1 and B_1 as

$$A_1 = \begin{pmatrix} a_{22} & a_{23} & a_{24} \\ a_{23} & a_{33} & a_{34} \\ a_{24} & a_{34} & a_{44} \end{pmatrix}, B_1 = \begin{pmatrix} b_{11} & b_{14} \\ b_{14} & b_{44} \end{pmatrix}.$$

The given conditions on the ranks of A and B mean that neither $\det A_1$ nor $\det B_1$ is divisible by p. As always, $\det(Ax + By) \equiv 0 \pmod{p}^2$; modulo p, we have

$$\det(Ax + By) = (\det B_1(a_{22}a_{33} - a_{23}^2))x^2y^2 + b_{11} \det A_1 x^3 y \equiv 0.$$

So $b_{11} \equiv 0 \pmod{p}$; also, the determinant of the matrix

$$\begin{pmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{pmatrix}$$

is zero mod p, so by a suitable change of variables we have $a_{22} \equiv a_{23} \equiv 0 \pmod{p}$.

Hence, without loss of generality, we have

$$A = \begin{pmatrix} pa_{11}^{(1)} & pa_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ pa_{12}^{(1)} & pa_{22}^{(1)} & pa_{23}^{(1)} & a_{24} \\ pa_{13}^{(1)} & pa_{23}^{(1)} & a_{33} & a_{34} \\ pa_{14}^{(1)} & a_{24} & a_{34} & a_{44} \end{pmatrix}, B = \begin{pmatrix} pb_{11}^{(1)} & pb_{12}^{(1)} & pb_{13}^{(1)} & b_{14} \\ pb_{12}^{(1)} & pb_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ pb_{13}^{(1)} & pb_{23}^{(1)} & pb_{33}^{(1)} & pb_{34}^{(1)} \\ b_{14} & pb_{24}^{(1)} & pb_{34}^{(1)} & b_{44} \end{pmatrix}.$$

Perform a flip operation, and we have

$$A' = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & pa_{13}^{(1)} & pa_{14}^{(1)} \\ a_{12}^{(1)} & a_{22}^{(1)} & pa_{23}^{(1)} & a_{24} \\ pa_{13}^{(1)} & pa_{23}^{(1)} & pa_{33} & pa_{34} \\ pa_{14}^{(1)} & a_{24} & pa_{34} & pa_{44} \end{pmatrix}, B' = \begin{pmatrix} b_{11}^{(1)} & b_{12}^{(1)} & pb_{13}^{(1)} & b_{14} \\ b_{12}^{(1)} & b_{22}^{(1)} & pb_{23}^{(1)} & pb_{24}^{(1)} \\ pb_{13}^{(1)} & pb_{23}^{(1)} & p^2b_{33}^{(1)} & p^2b_{34}^{(1)} \\ b_{14} & pb_{24}^{(1)} & p^2b_{34}^{(1)} & pb_{44} \end{pmatrix}$$

which clearly have $\langle e_3 \rangle$ as a common kernel; so we can apply previous sections to remove a factor p^2 from $q([A', B'])$.

□

A practical implementation of minimisation for four-coverings simply follows the shape of this proof slavishly.

2.6 Reducing four-coverings

2.6.1 Two forms of naïve reduction

We have actions of GL_2 and GL_4 on four-coverings; they commute, so we reduce under the two actions separately. To handle the action of GL_2, we want the generating matrices A and B to have the smallest possible L^2 norms: we achieve this by reducing the two-dimensional lattice in $\mathbb{Z}^{16}$ generated by (a_{ij}) and (b_{ij}), using an algorithm due to Gauss [38].

For the GL_4 reduction, until the work of Stoll mentioned below, I used a steepest-descent procedure to find a local minimum for $\sum a_{ij}^2 + b_{ij}^2$.

Define the matrix norm [2] $L^2(M) = \sum a_{ij}^2$, and extend it in the obvious way to pairs of matrices by $L^2([A,B]) = L^2(A) + L^2(B)$.

To perform one reduction step, consider the twelve 4×4 matrices with ones along the diagonal, and zeroes everywhere else except for a single x; computing $L^2(M \cdot [A,B])$ for such a matrix M gives a quartic polynomial $p(x)$.

$p(x)$ is positive definite because it is a sum of squares, so, for a given r, $p(x) - r$ will be negative in only a finite region (either one interval or two). By solving the quartic $p(x) - r = 0$ we can obtain this region; if it is short enough that it makes sense to check every integer within it, we do that and return the one giving the least $p(x)$. If not, we replace r by $p(t)$ for t chosen at random within the region, and repeat. The region must get shorter at each iteration so this process is finite.

This gives for each M the value of x that minimises $L^2\left(M\left(x\right) \cdot [A,B]\right)$; we look over the minimal values given by all twelve matrices and pick the M and x which give the smallest value. If it is not an improvement on $L^2([A,B])$ then stop, otherwise replace $[A,B]$ by $M(x) \cdot [A,B]$ and continue.

The convergence of this method is not at all good, since we are restricted to movements along a small set of directions, and the route to the local minimum need not lie conveniently with relation to this set; on many curves it ran overnight without completion. It would be straightforward to augment the set of directions used, by using matrices of the form

$$M \begin{pmatrix} 1 & x & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} M^{-1}$$

for arbitrary $M \in \mathrm{GL}_4(\mathbb{Z})$, but the method is in any case rendered obsolete by Stoll's work in the next section.

2.6.2 Stoll's reduction method

Stoll's reduction method [69], a side-result of his work on reduction of ternary cubics in [27], associates a real 4-variable Hermitian quadratic form C' to any four-covering, and then declares the four-covering reduced by transforming it by the matrix M such that $MC'M^{\mathrm{T}}$ is LLL-reduced.

To define C', note that over $\mathbb{P}(\overline{\mathbb{Q}})$ there will be four points $P_i = (\lambda_i : \mu_i)$ such that $\det(\lambda_i A + \mu_i B) = 0$. Each P_i corresponds to a singular quadric in our pencil[3]; let $\mathbf{v}_i$ be a vector in $\ker(\lambda_i A + \mu_i B)$, the vertex of this singular

[2] to be pedantic, this is the square of the normal L^2 matrix norm

[3] given that we forbid 2-torsion on the elliptic curve, and hence the quartic defining our

quadric. Note that $\mathbf{v}_i$ is only defined up to a scalar multiple.

Replacing $[A, B]$ by $[A', B'] = [MAM^{\mathrm{T}}, MBM^{\mathrm{T}}]$ where $M \in \mathrm{SL}_4(\mathbb{C})$ has rows the $\mathbf{v}_i$, we convert the four-covering to a standard form where the singular quadrics each have vertex $\mathbf{e}_i$ and are represented by diagonal matrices $M^{(i)} = \lambda_i A' + \mu_i B'$, with $M^{(i)}_{ii} = 0$.

The 'standard-position form' is then defined as

$$C = \mathrm{diag}\left(1, \sqrt{\left|\frac{M^{(3)}_{22} M^{(4)}_{22}}{M^{(3)}_{11} M^{(4)}_{11}}\right|}, \sqrt{\left|\frac{M^{(2)}_{33} M^{(4)}_{33}}{M^{(2)}_{11} M^{(4)}_{11}}\right|}, \sqrt{\left|\frac{M^{(2)}_{44} M^{(3)}_{44}}{M^{(2)}_{11} M^{(3)}_{11}}\right|}\right).$$

We transform back to the original co-ordinate frame by $C' = \overline{M^{-1}} C {M^{-1}}^{\mathrm{T}}$, which is the correct rule for transforming a Hermitian form: think of $C = \sum C_i |y_i|^2$. C' is then a symmetric positive-definite matrix of real numbers: we scale its entries so that $\sum_{i,j=1}^{4} c_{ij}^2 = 1$.

C' is a lattice associated with the four-covering $[A, B]$; we say that $[A, B]$ is reduced if C' is LLL-reduced. And, since the C' is essentially a covariant of the four-covering, we can construct (by using `magma`'s `LLLGram` function) the matrix $K \in \mathrm{SL}_4(\mathbb{Z})$ for which $KC'K^{\mathrm{T}}$ is LLL-reduced, and then replace $[A, B]$ by the reduced four-covering $[A', B'] = [KAK^{\mathrm{T}}, KBK^{\mathrm{T}}]$ as a reduced form of $[A, B]$.

2.6.3 Canonical forms for four-coverings

LLL reduction for a lattice need not give an absolutely unique element in the $\mathrm{SL}_4(\mathbb{Z})$-orbit containing that lattice; if a canonical form for a reduced four-covering is desired, we must also take account of the fact that the variables can be permuted, their signs can be changed, one or other or both of the matrices A and B can be negated.

Since the lattice involved is of rank 4, there is a truly-canonical reduction in which we list the vectors of the lattice in order of increasing length, and pick, four times, the shortest vector not a linear combination of vectors already picked. `magma` has various facilities built in for listing vectors in order of length, which is the hard part of this operation.

To deal with the other non-uniquenesses, we define a canonicalisation procedure: the canonical form of the four-descendent $[A, B]$

- has the diagonal entries of the matrix A in increasing order by absolute value, with ties broken by looking at the diagonal entries of B (that is, if $a_{ii} = a_{jj}$ with $i < j$, we want $b_{ii} \leq b_{jj}$).
- has the first non-zero diagonal entry of both A and B positive

two-covering is irreducible over $\mathbb{Q}$, these P_i will not lie in $\mathbb{P}(\mathbb{Q})$

- $A_{12}, A_{13}, A_{14} \geq 0$
- If $A_{12} = 0$ then $B_{12} \geq 0$, and likewise for A_{13} and A_{14}

These definitions are useful when writing down four-coverings, since they provide a single form for a reduced, minimised four-coverings, and hence four-coverings with those properties are equivalent iff they look the same.

2.6.4 Practical notes

The Stoll reduction performs an $\mathrm{SL}_4(\mathbb{Z})$ transform, and the LLL reduction performs an $\mathrm{SL}_2(\mathbb{Z})$ transform. Stoll-reducing an LLL-reduced form need not produce an LLL-reduced form, and vice versa. So we apply LLL and Stoll reductions alternately, keeping track of every form we see and stopping when we reach some form for the second time; Stoll reduction can sometimes substantially *increase* the L^2 norm of a four-covering (though usually `LLLReduce(StollReduce(x))` will have smaller L^2 norm than `x`), so we cannot simply apply transformations until the L^2-norm stops going down.

There is no point in keeping track of the transformations used; it would be useful only in the unlikely event that we might want to see what a point found on the reduced four-covering looked like on the original four-covering, and in that case I've found it easy to follow

$$\mathcal{H}_1 \xrightarrow{j_{\mathcal{H}_1 \to E}} E \xrightarrow{j^{-1}_{\mathcal{H}_2 \to E}} \mathcal{H}_2$$

where the inverse map is calculated by the method of section B.1.

The conversion of a four-covering with large elements to standard form requires calculations to an extremely high level of precision; the current implementation requests $12d + 1000$ significant figures of each root of the associated quartic, where d is the largest number of digits of any entry in either matrix defining the four-covering, and implementations working with less grotesque precision frequently failed to have MAM^{T} even close to a diagonal matrix. Thankfully, `magma` has fast enough ultra-high-precision arithmetic that this is not an insurmountable obstacle.

2.7 Finding rational points on a four-covering

At this stage in the four-descent procedure, we have obtained a set of size 2^{s+r-1} (where $2^s = \|Ш[2]\|$ and $r = \operatorname{rank} E$; see section 2.10.1 for the derivation of this number), whose elements are pairs of 4×4 matrices (A, B) with integer coefficients, representing everywhere-locally-solvable intersections of two quadrics.

Our goal becomes to find explicit points (w, x, y, z) with

$$f_1(w, x, y, z) = f_2(w, x, y, z) = 0.$$

There are two obvious approaches: a direct search, which is easy to write but slow, and an indirect approach which involves significantly more preparatory work, but makes the implementation of the sieve more straightforward since we can work on a variety defined by a single polynomial, and offers a significant constant factor of speed improvement.

For the direct search, having first checked whether there is a point with $x_1 = x_2 = 0$, we let $(x_1 : x_2)$ run through $\mathbb{P}^1(\mathbb{Q})$ (IE through pairs of coprime integers), substitute into both equations to get a pair of homogenous quadratics in x_3 and x_4, and use standard resultant-based methods to solve two simultaneous quadratics. The problem boils down to finding the roots in $\mathbb{Q}$ of a quartic equation, once for each choice of $(x_1 : x_2)$.

For the indirect approach, we pick some quadratic form Q from the pencil and find one point on it, and use lines through that point to construct a parametrization of the rational points on Q, of the form $P_Q = (\phi_1(t, u, v) \ \ldots \ \phi_4(t, u, v))$ with the ϕ_i homogenous quadratic forms.

Let $Q_2(x_1, x_2, x_3, x_4)$ be a quadratic form in $\langle f_1, f_2 \rangle$ independent of Q, and construct the homogenous ternary quartic $T(t_1, t_2, t_3) = (\phi_1(t_1, t_2, t_3) \ \ldots \ \phi_4(t_1, t_2, t_3))$.

We now perform a direct search on this ternary quartic: let $(t_1 : t_2)$ run through pairs of coprime integers and substitute in, which gives a quartic equation for t_3; if it has roots over $\mathbb{Q}$ then we have successfully found a point, which we map back to $\langle f_1, f_2 \rangle$ by evaluating the ϕ_i at (t_1, t_2, t_3).

Although ϕ is a parametrization by quadratic forms with potentially rather large coefficients (if the point we picked on Q had large co-ordinates), it is a degree-one map. This apparent contradiction resolves itself by the experience that, substituting in the co-ordinates of size around N of a point found on T, we get a point with co-ordinates of size around N^2 on Q, but these co-ordinates have a common factor of size around N.

Both attacks can be sped up by sieving procedures, where the sieves are derived by working backwards from the requirement that the quartics have roots in all $\mathbb{F}_p$. Constructing the sieve for the symmetric approach is not entirely straightforward, but the routines for checking solubility over $\mathbb{F}_p$ of a pair of quadratics were implemented as part of an earlier unsuccessful attempt to understand local solvability of four-coverings.

The asymmetric attack requires more setting-up – to find even one point on one of the quadratic forms can be difficult. The easiest approach is to take hyperplane sections, which then give plane conics on which we can use the

sophisticated routines for finding rational points developed by Cremona and Rusin [24] and implemented in `magma`.

There is one hyperplane corresponding to each element of $\mathbb{P}^3(\mathbb{Q})$, so we have many possible choices; in my current implementation I check the points arising from all 1120 choices of

$$(x_1 : x_2 : x_3 : x_4) : x_i \in \mathbb{Z}, |x_i| \leq 3, \mathrm{GCD}\,\{x_i\} = 1$$

and pick the rational point whose coordinates have smallest maximum absolute value. I do not have any argument that that is the best choice.

Setting up the conics for this approach, however, requires at each stage the factorisation of the coefficients of the ternary quadratic form Q_t describing the hyperplane section (or, at least, the factorisation of the determinant if we use the method of [67]). This is a severe problem if using unreduced four-coverings; with the much smaller coefficients of Stoll-reduced four-coverings the problem essentially vanishes.

$\mathrm{GL}_3(\mathbb{Q})$ acts on ternary quartics by change of variable, so as in previous cases there is the potential for minimisation and reduction. When the four-covering is reduced, the coefficients in the ternary quartic are small enough that there is not much scope for further reduction; I am sure that a reduction approach via a Hermitian covariant exists, but leave finding it to future work.

2.7.1 Parametrizing a $\mathbb{P}^3$ quadric via a rational point

We assume that the rational point has no zero coordinates, or at least that the non-zero coordinates are at variables where we can affinise. If our point is $\mathbf{p} = (p_1 : p_2 : p_3 : p_4)$, and our quadric Q is in the variables x_1, x_2, x_3, x_4, begin by picking some $i : p_i \neq 0$.

Produce an affine ternary quadratic $\tilde{Q}$ by setting $x_i = p_i$; we obtain a point $\mathbf{p}'$ by listing the $p_j : i \neq j$. Construct

$$\mathbf{x} = \mathbf{p}' + \begin{pmatrix} tu \\ tv \\ tw \end{pmatrix},$$

and substitute this into the equation for $\tilde{Q}$ to get a quadratic equation in t, one of whose roots will be zero.

The other root will be of the form $\ell(u, v, w)/q(u, v, w)$, where ℓ is a homogenous linear form and q a homogenous quadratic. So, substituting back into $\mathbf{x}$,

we get something of the form

$$[f_1(u,v,w)/f_4(u,v,w), f_2(u,v,w)/f_4(u,v,w), f_3(u,v,w)/f_4(u,v,w)]$$

with all the f_i homogenous quadratics, and $f_4 = q$. Reprojectivise, clear f_4 from the denominators, multiply through by the LCM of the denominators of all the coefficients, and we have a parametrization over $\mathbb{P}^3$ by homogenous quadratic elements of $\mathbb{Z}[u,v,w]$.

2.7.2 Some invariants of the ternary quartic

Note that $\mathrm{SL}_3(\mathbb{Z})$ is generated by the maps $M_1 : (x,y,z) \to (x+y,y,z)$, $M_2 : (x,y,z) \to (x+z,y,z)$, $M_3 : (x,y,z) \to (x,x+y,z)$ and $M_4 : (x,y,z) \to (x,y,x+z)$.

To prove this, we recall that $\mathrm{SL}_3(\mathbb{Z})$ must be generated by the elementary 3×3 matrices. These elementary matrices have ones along the diagonal and a single non-zero off-diagonal element, so are clearly powers of M_1 through M_4, together with $M_5 : (x,y,z) \to (x,y+z,z)$ and $M_6 : (x,y,z) \to (x,y,y+z)$. Now, observe that

$$M_5 = M_1^{-1} M_3 M_1^{-1} M_2 M_1 M_3^{-1} M_1$$

and

$$M_6 = M_1^{-1} M_3 M_1^{-1} M_4 M_1 M_3^{-1} M_1.$$

To motivate that observation, recall the computer-science trick of swapping A and B by the instruction sequence `A := A+B; B := A-B; A := A-B;` we are using this to swap x and y.

Consider a generic ternary quartic

$$\begin{aligned} q(x,y,z) = {} & ax^4 + bx^3y + cx^3z + dx^2y^2 + ex^2yz + fx^2z^2 + gxy^3 + hxy^2z \\ & + ixyz^2 + jxz^3 + ky^4 + ly^3z + my^2z^2 + nyz^3 + oz^4 \end{aligned}$$

where I will call the terms $x^\alpha y^\beta z^\gamma$ 'q-terms', and the $a \dots o$ 'ℓ-terms': the ℓ-terms label the q-terms, and we say for example that g labels xy^3. We call a term – a product of the ℓ-terms raised to some powers, of degree the sum of the relevant exponents – coherent of weight w if the product of the q-terms labelled by its ℓ-terms is $(xyz)^w$. For example, $cdfjl^2$ is a degree-six term which is coherent of weight 8, since $x^3zx^2y^2x^2z^2xz^3(y^3z)^2 = x^8y^8z^8$. Coherent terms exist only for degrees a multiple of 3; it's fairly clear (by considering the action of I, 2I, 3I ...) that any invariant must be a sum of coherent terms of equal degree.

By a procedure which is essentially equating coefficients — let $f = \sum \lambda_i t_i$, where t_i are the coherent degree-3 terms, and require $f(M_i \circ q) - f(q) = 0$ for $i = 1 \ldots 4$ to get a great number of expressions in the λ_i which must all be simultaneously zero — we find that

$$\begin{aligned}\Delta(q) = 144ako - 36aln + 12am^2 - 36bgo + 9bhn - 6bim + 9bjl + 9cgn - 6chm \\ + 9cil - 36cjk + 12d^2o - 6den + 4dfm - 6dhj + 2di^2 + 2e^2m \\ - 6efl + 9egj - ehi + 12f^2k - 6fgi + 2fh^2\end{aligned}$$

is invariant under the action of $\mathrm{SL}_3(\mathbb{Z})$. It is obvious that $\Delta(nq) = n^3\Delta(q)$, and easy to check that $\Delta(q(\alpha x, \beta y, \gamma z)) = (\alpha\beta\gamma)^4\Delta(q(x, y, z))$.

Working on coherent degree-six terms instead, we have a rather complicated sparse linear system to solve, and we end up with two independent invariants Δ^2 and Ψ. These invariants Δ and Ψ have in fact been known for a very long time – they appear in [54] – but it is interesting to derive them, and rather easier and more likely to be correct than typing in the page-long expressions from the reference book.

2.7.3 The `LinearMinimise` approach

Let f_1 be an integral homogenous ternary quartic; we wish to find a transformation $M \in \mathrm{SL}_3(\mathbb{Q})$ to make the valuations of the invariants $\Delta(M \cdot f_1)$ and $\Psi(M \cdot f_1)$ as small as possible while keeping $M \cdot f_1$ integral. Naturally we can do this one prime at a time, and, since the action by a matrix of determinant p^{-1} would send $\Delta \to p^{-3}\Delta$ and $\Psi \to p^{-6}\Psi$, we need consider only the (usually very few) p with $p^3|\Delta(f_1)$ and $p^6|\Psi(f_1)$.

For a given p, we consider the set P of points on f over $\mathbb{P}^2(\mathbb{F}_p)$. If we ever find that all the elements of P lie on the same line $\ell : \ell_x x + \ell_y y + \ell_z z = 0$ – which is a simple matter of linear algebra – we have an expression which holds for all the points of f over $\mathbb{P}^2(\mathbb{F}_p)$. So, if we lift ℓ_x, ℓ_y and ℓ_z to $\mathbb{Z}$, we have, for all points in $\mathbb{P}^2(\mathbb{Z})$, that $\ell_x x + \ell_y y + \ell_z z = pw$. We then eliminate x, y or z (naturally we require $\ell_x \neq 0$ to eliminate x and so on) in exchange for pw; for example, substituting $\ell_y y = pw - \ell_x x - \ell_z z$, we get a new homogenous ternary quartic f_2 in x, w, z, and a transformation matrix, which we call M_{12}, for recovering a point on f_1 from one on f_2.

We can remove any factor dividing all the coefficients of f_2, and repeat this process to get f_3, f_4 and so on, accumulating the transformation matrices to get M_{13}, M_{14} and so on. This repeated process can finish in one of two ways. Either the points of f_n over $\mathbb{F}_p$ fail all to lie on a line, in which case we report f_n

and M_{1n} and terminate, or we can find that M_{1n} has all its coefficients divisible by p. In this latter case, we have shown that f_1 has no non-trivial points in $\mathbb{P}^2(\mathbb{Z}_p)$, since any such point would arise from one on f_n, but the action of M_{1n} would send it to something with a common factor in all three coordinates, which is not permitted.

Of course, the latter case does not occur in the four-descent as implemented, since we have assured local solubility at the algebraic stage before even constructing the four-coverings. This step serves in the current implementation only to reduce the problems that might arise from an infelicitous choice of generator on the first $\mathbb{P}^3$ quadric.

2.8 Sieving for points on homogenous ternary polynomials

We are given a homogenous ternary polynomial $Q(x, y, z) = 0$, and seek an integral point on it. Fixing two of the variables gives a univariate polynomial in the third; we are happy if this has a solution over $\mathbb{Q}$ since we can simply scale to get one over $\mathbb{Z}$. To find solutions quickly, we apply a sieving process first.

This part of the algorithm is implemented as a separate C++ program `sievetq`, using Victor Shoup's `NTL` library [60] for the polynomial factorisation, with a small interface routine written in `magma` to translate the polynomial into the list of coefficients which `sievetq` requires, and translate the output of `sievetq` into a format that `magma` can handle.

For the polynomial to have a solution over $\mathbb{Q}$, it must have one over $\mathbb{F}_p$ for all p, and indeed over $\mathbb{Z}/p^n\mathbb{Z}$ for all prime powers. We divide the prime powers into two sets; the eighteen ones less than 64 are called 'small' and used in the first phase of the sieving, whilst the sixteen primes between 67 and 139 are used as filters.

We work simultaneously on the three polynomials $Q_1 = Q(x, y, z), Q_2 = Q(x, z, y), Q_3 = Q(z, y, x)$; this is to ensure that a point on Q will be found in time dependent on its second-largest co-ordinate, rather than its largest. For each of these polynomials P, and each small prime power q, we construct a $q \times q$ array of bits A_q (since $q \leq 64$ this array is stored as q 64-bit integers), with $A_q[x, y] = 1$ if there exists a $z \in [0, q)$ with $P(x, y, z) \equiv 0 \pmod{q}$, and 0 otherwise. We find such a z simply by running through all possibilities; for q this small, this is much quicker than using `NTL` to factorise the polynomial mod q.

Since P is homogenous, $P(ax, ay, az) \equiv 0 \iff P(x, y, z) \equiv 0$; we do not take advantage of this for q a prime power, but, if q is prime, we compute the

$y = 1$ row explicitly, then construct the other rows by $A_q[x, y] = A_q[xy^{-1}, 1]$ where the inverse is taken mod q. Clearly $A[x, 0] = A_q[1, 0]$ for $x \neq 1$, which allows the $y = 0$ row to be constructed very quickly.

For a filter prime p, we construct a one-dimensional array A_p of $p+2$ elements numbered $0 \ldots p + 1$; $A_p[x]$ records $\exists z : P(x, 1, z) \equiv 0 \pmod{p}$ for $x \in [0, p)$, $A_p[p]$ records $\exists z : P(1, 0, z) \equiv 0 \pmod{p}$, and $A_p[p+1]$ records $\exists z : P(0, 0, z) \equiv 0 \pmod{p}$. Constructing these sieves, for the three permuted polynomials and the thirty-four primes, takes a couple of seconds.

We sieve the plane in regions, picked of a size $r \times r$ such that the number of bits in a region fits in the 256- or 512-kilobyte level-two cache of the computer used; we start with a region centred at $(0, 0)$, and then work outwards in a pattern of squares. This allows a search to be conveniently extended incrementally, without re-searching regions already handled.

To sieve a region centred at a given point, we construct an $r \times r$ array of bits, initialised to 1, and then tile it with appropriately-aligned and -offset copies of the small-prime sieve arrays, combining them using the AND operation. After this, a bit will be set only if it is a possible solution modulo all the small prime powers. We read out the co-ordinates of the set bits, offset them to get x and y co-ordinates, and proceed to the filtering stage. The initial implementation missed out the filtering stage, proceeding directly to `NTL`'s polynomial factorisation, and accordingly spend nearly all its time factorising.

At the filtering stage, we consider candidate (x, y) pairs, and reduce them modulo each of the filter primes p. If the Y co-ordinate is zero we check $A_p[p]$ if $x \neq 0$ and $A_p[p + 1]$ otherwise; if not, we check $A_p[xy^{-1}]$. This requires a single modular inversion – done by a look-up table – per candidate, and is very quick; indeed, quicker than sieving by the filter primes would be, since the time taken is proportional to the number of candidates rather than the size of the $r \times r$ sieve array. It rejects a large number of useless candidates; the remainder are checked by polynomial factorisation, and, if success is declared, passed to an output routine.

The two-stage sieving process, using bitwise sieves for primes smaller than the size of a machine word and a second filtering step for larger primes, is also used in Stoll et al's `ratpoint` program for finding points on $y^2 = f(x)$.

To search, without success, the region $|x| < 10240, |y| < 10240$, for all three permuted polynomials, for an input ternary quartic, takes approximately 12.3 seconds on a P4/2400 computer; to search $|x|, |y| < 5 \times 10^4$ takes about fifteen minutes. Obviously, successful searches would be quicker since we stop at the first point found.

The same approach would work on homogenous ternary polynomials of any order, in particular the ternary cubics which would appear in any effort to

perform an explicit 3-descent; indeed, the ternary-cubic case can be handled by `sievetq` simply by filling the degree- four elements of the input quartic with zeroes, at a small performance penalty over customised code.

For bounds larger than about 10^4, my optimised C++ implementation of the sieve is inferior to my `magma` implementation of the much more clever algorithm due to Elkies in the next section.

2.9 p-adic Elkies search on a pair of quaternary quadratics

Elkies [33] gives an algorithm which, for very generally-defined functions

$$f(x_1, \ldots, x_m),$$

will find all the points over $\mathbb{Z}^m$ with $|x_i| \leq N$ and $|f(x_1, \ldots, x_m)| < \delta$ (for sufficiently-small δ) in time $O(N \log^{O(1)} N)$; in particular, it will find all the integral points on f with $|x_i| \leq N$.

I present below an implementation I devised for the quaternary-quadratic case which interests us, and I also have an implementation written by Elkies himself for working on ternary quartics. One expects better results from working on the quaternary quadratics, since the lattices have determinant $\Delta = p^5$ rather than $\Delta = p^3$, and the number of 'spurious' vectors of length ℓ is $O(\ell^2 \Delta^{-1})$.

The paper [33] proposes an algorithm using real approximations to the curve – surrounding the curve by a union of parallelepipeds – rather than the p-adic approach used here: that would be rather more difficult to implement because effective bounds on the curvature of the surfaces involved would be necessary to optimise the number and volume of parallelepipeds used while guaranteeing that they contain the curve.

The approach is rather reminiscent of the classical lattice arguments used in first-year number theory texts to show that every integer is a sum of four squares, though the final step is a construction via LLL reduction rather than an existence argument via Minkowski's theorem.

We start with a pair of homogeneous four-variable quadratics

$$f_i(x_1, x_2, x_3, x_4) = 0.$$

Write

$$d_i = \left[\frac{\partial f_i(1, x_2, x_3, x_4)}{\partial x_2}, \frac{\partial f_i(1, x_2, x_3, x_4)}{\partial x_3}, \frac{\partial f_i(1, x_2, x_3, x_4)}{\partial x_4} \right]$$

as the two partial derivatives.

Given a search bound N, we pick a prime p slightly greater than N, and run through the range $x = 1 \ldots p$. For each x, we list the solutions to

$$f_1(1, x, y, z) \equiv f_2(1, x, y, z) \equiv 0 \pmod{p}$$

by finding the roots of the quartics over $\mathbb{F}_p$ obtained by substituting $x_1 = 1, x_2 = x$ into pre-computed resultants $\mathrm{Res}_y(f_1, f_2)$ and $\mathrm{Res}_z(f_1, f_2)$: at absolute worst there will be sixteen, usually there are between zero and two.

For each of these solutions P, we perform a Hensel lift to a point P' with $f_1(P') \equiv f_2(P') \equiv 0 \pmod{p^2}$, by solving

$$\begin{aligned} d_1 \cdot [u, v, w] &\equiv -p^{-1} f_1(P_i) \pmod{p} \\ d_2 \cdot [u, v, w] &\equiv -p^{-1} f_2(P_i) \pmod{p} \end{aligned}$$

This is an under-determined system, so solutions are of the form

$$[x_1, x_2, x_3] + \lambda[y_1, y_2, y_3].$$

We can scale so that $y_1 = 0$ or $y_1 = 1$.

Let $P' = P + [0, px_1, px_2, px_3]$, let $D = [0, py_1, py_2, py_3]$ where we are considering the y_i as elements of $\mathbb{Z}$. If $y_1 = 1$, consider the lattice L generated by $P', D, [0, 0, p^2, 0]$ and $[0, 0, 0, p^2]$; if not, use $P', [0, p^2, 0, 0], D$ and $[0, 0, 0, p^2]$ as the generators.

Any vector $\mathbf{v} \in L$ will by construction have $f_1(\mathbf{v}) \equiv f_2(\mathbf{v}) \equiv 0 \pmod{p^2}$ and $\mathbf{v} \equiv \lambda P \pmod{p}$ for some λ; equally, any vector satisfying the two above conditions will lie in the lattice. In particular, any vector with $\mathbf{v} \equiv \lambda P \pmod{p}$ and $f_1(\mathbf{v}) = f_2(\mathbf{v}) = 0 \in \mathbb{Z}$ will do so.

We can ask `magma` for the vectors in the lattice with norm less than $4p^{5/2}$, using the `ShortVectors` function, and it has very efficient algorithms for calculating them; since the determinant of the matrix defining the lattice is p^5, we expect there to be $O(1)$ entries of norm that short. However, any vector $\mathbf{v} \in \mathbb{Z}^4$ with all its coordinates in $[-N^{5/4}, N^{5/4}]$ and with $f_1(\mathbf{v}) = f_2(\mathbf{v}) = 0$ will be among them. So we look along our list of short vectors for ones that are simultaneous points on f_1 and f_2, and return those: this will sometimes return points one of whose coordinates is a bit greater than N and the others enough smaller to compensate, but we don't mind obtaining **extra** rational points.

In experiments on a number of four-coverings, we find that an unsuccessful search with bound N takes very close to N milliseconds on my P4/2400 computer; as mentioned above, this breaks even with the sieving process at $N^{5/4} \approx 10^4$ or $N \approx 1600$.

There are two minor problems. The search method will not give points with $x_1 = 0$, which is not a serious defect since such points can be found by setting $x_1 = 0, x_2 = 1$ and solving the pair of quadratics. If, for some value of x_2, the derivative vectors d_1 and d_2 are linearly dependent mod p, the search method does not work for that x_2 value; but it will work for a different p.

And in practice one considers several different p, since, when a very simple point P exists, the short vectors will include $P, 2P, 3P, \ldots$; the `ShortVectors` function then spends effort to find these, and another loop then needs to spend effort to filter out points with $(x_1, x_2, x_3, x_4) \neq 1$. So it's worth starting with a small p and then increasing it until a point is found or a time limit exhausted; this also avoids the embarrassment of finding a point with very small coefficients towards the end of a search with a very large p. As usual for this kind of p-adic algorithm, it's not possible to use the results of a search with small p to speed up a search with larger p.

2.10 Working with several descendents

Recall that the four-descent process is a *second* two-descent; so the input to it is an elliptic curve E, a pair of integers $r_1 < r_2$ with $r_2 = \operatorname{rank} S_2(E)$, a set of r_1 independent points on E which we generally ignore, and a set $\mathfrak{D}_2$ of $2^{r_2 - r_1} - 1$ everywhere-locally-solvable two-coverings, one per non-trivial element of the quotient of $S_2(E)$ by the part generated by the known generators of $E(\mathbb{Q})$.

We compute four-descents on all the elements $T \in \mathfrak{D}_2$. If some T has no everywhere-locally-solvable four-descendents, we know it represented an element of Ш[2]; if not, it gives us some elements of $S_4(E)$.

Folklore has it that a two-descent with 'reasonable' bounds for point search on the two-coverings will miss out at most one generator; that is, that you do not expect an elliptic curve to have more than one enormous generator. The conjectured finiteness of Ш[2^∞] means that a two-descent on a curve with non-trivial Ш[2], finding all the generators for the Mordell-Weil group, will have an even difference between the Mordell-Weil and the 2-Selmer ranks. So it was expected that a two-descent which finds r_1 generators and a 2-Selmer rank of r_2 indicated that the curve has Mordell-Weil rank r_1 ($r_2 - r_1$ even) or $r_1 + 1$ ($r_2 - r_1$ odd), and Ш[2] = $\mathbb{Z}/2\mathbb{Z}^{r2-r1(+1)}$.

It was also believed that Ш[2] = Ш[2^∞] for almost all curves, though it is not clear what evidence existed to justify this. Until this work, the only way to look at Ш[2^∞] was to believe all the conjectures and compute $Ш_{\mathrm{anal}}$. This takes a long time for curves of even moderately large conductor, even if routines to evaluate L-functions and their derivatives are easily available – and they are not available in the default installation of any common computational number

theory package.

2.10.1 How large is $\mathfrak{D}_4$?

Suppose we have an elliptic curve E with $E(K)[2]$ trivial, and $E(K) \simeq \mathbb{Z}^r$; naturally $E/2E \simeq (\mathbb{Z}/2\mathbb{Z})^r$ and $E/4E \simeq (\mathbb{Z}/4\mathbb{Z})^r$. Suppose that the Tate-Shaferevich group of E has

$$Ш[2] \simeq (\mathbb{Z}/2\mathbb{Z})^s,\ Ш[4] \simeq (\mathbb{Z}/2\mathbb{Z})^{s_1} \times (\mathbb{Z}/4\mathbb{Z})^{s_2}$$

so $\|Ш[2]\| = 2^s$ and $\|Ш[4]\| = 2^{s_1}4^{s_2}$, with $s = s_1 + s_2$. We also have $\|S_2\| = 2^{r+s}$ and

$$\|S_4\| = 2^{s_1}4^{r+s_2} = \|S_2\|^2\, 2^{-s_1}.$$

We have the standard 2- and 4-descent diagrams

$$0 \longrightarrow E/2E \longrightarrow S_2 \longrightarrow Ш[2] \longrightarrow 0$$

$$0 \longrightarrow E/4E \longrightarrow S_4 \longrightarrow Ш[4] \longrightarrow 0$$

which combine to give the second-descent diagram (see for example [22])

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & E/2E & \xrightarrow{\times 2} & E/4E & \longrightarrow & E/2E & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & S_2 & \longrightarrow & S_4 & \longrightarrow & S_2 & \longrightarrow & \mathrm{coker} \longrightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \| \\
0 & \longrightarrow & Ш[2] & \hookrightarrow & Ш[4] & \xrightarrow{\times 2} & Ш[2] & \longrightarrow & \mathrm{coker} \longrightarrow 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

where the cokernel is $Ш[2]/2Ш[4]$, of size 2^{s_1}; its size could also be calculated by recalling that the alternating product of the orders of the groups along any row or column is 1.

So the image of S_4 in S_2 has index 2^{s_1}, indicating that 2^{r+s_2} of the $2^{r+s_1+s_2}$ two-descendents do lift to four-descendents. If $s_1 = 0$, all the contribution of Ш to the 2-Selmer group is the result of generators of order at least four, and the four-descent doesn't improve the bounds; if $s_2 = 0$, all the contribution of Ш to the 2-Selmer group is the result of generators of order exactly two, and

the only part of S_2 which lifts is the image of E.

If an element of S_2 lifts to S_4, its preimage will have order $\|S_2\| = 2^{r+s}$. Each four-covering curve gives two elements of S_4, so we will find 2^{s+r-1} four-descendents for each of the 2^{r+s_2} elements of S_2 which lift. These counts include the trivial element of S_2, which lifts but which in practice we do not observe.

For example, if E were a rank-one curve with Ш $\simeq (\mathbb{Z}/2\mathbb{Z}\times\mathbb{Z}/4\mathbb{Z})^2$, we have $r = 1, s = 4, s_1 = 2, s_2 = 2$; we obtain $2^5 - 1 = 31$ non-trivial two-descendents, of which seven lift, each giving 16 four-descendents. Exactly one of this total of 112 four-coverings will have a rational point which will lift to the generator of E, but we have to check all of them to find it; this is irritating if the correct four-covering is the 112th checked.

2.10.2 Some worked examples

Consider the curve $E : y^2 = x^3 - 758201$. E has trivial torsion over $\mathbb{Q}$, and `mwrank` tells us that the 2-Selmer rank is four, which would give us fifteen non-trivial two-descendents. However, a point (which lifts to $P_1 = (105, 632) \in E$) is found on one of the descendents $\mathcal{C}_1$; `mwrank` conveniently computes the quotient $S_2/\langle P_1\rangle$, which has eight elements, and gives us seven non-trivial two-coverings which serve as coset representatives from this quotient group.

According to the folklore in the introduction to this section, we expect to be able to find one more generator and to prove that Ш$[2^\infty] = \mathbb{Z}/2\mathbb{Z}^2$; that is, $r = 0, s_1 = 2, s_2 = 0$, which would give $\|E(\mathbb{Q})/4E(\mathbb{Q})\| = 4^2$ and $\|S\|_4 = 2^2 4^2 = 64$.

Six of these two-coverings are resolvable, in that we find $\mathfrak{D}_4^{\mathrm{alg}} = \emptyset$ even without checking local solvability. On the seventh, all the algebraic checks for local solvability succeed, and we find eight four-descendent curves; as noted before, each of these curves corresponds to two elements of S_4.

We construct the ternary quartics corresponding to each of these eight curves, and search for points up to median$(|u|, |v|, |w|) < 3 \times 2^{11}$ (since this is the smallest search region supported by `sievetq`; an unsuccessful search takes about ten seconds, a successful search is quicker since it stops as soon as a point is found). On two of the curves, we find points; these map back to

$$P_2 = (1045373303394937468897 21/20952564486^2, Y^+)$$

where Y^+ is a complicated positive rational number, and to $P_2' = P_2 + 2P_1$. If we add $-P_2$ and $-P_2'$ to this set, which we are allowed to do by picking the opposite sign for Y, we get a whole coset $(P_2 + 2E(\mathbb{Q})) + 4E(\mathbb{Q})$. LLL reduction of the height-pairing matrix gives $\{P_1, P_2 - P_1\}$ as a basis consisting of points

of minimal height;

$$P_2 - P_1 = (29555174951277593473 17/2127187943^2, Y_1^-)$$

is only marginally more attractive than our original P_2.

One two-descendent $\mathcal{C}_2$ lifted, to give eight four-descendents. But, had we not quotiented out by $\langle P_1 \rangle$ or ignored the trivial two-covering, we would have had four two-descendents $(\mathrm{triv}, \mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_1 + \mathcal{C}_2)$ capable of lifting. So, from section 2.10.1, we have $2^{s_1+s_2+r-1} = 8$ and $2^{r+s_2} = 4$, and by listing explicit generators we have $r \geq 2$. We can solve to get $s_2 = 0$ and $s_1 = 2$; so $E \simeq \mathbb{Z}^2$ and $Ш[2^\infty] \simeq (\mathbb{Z}/2\mathbb{Z})^2$.

Occasionally, and generally only when performing a four-descent on a curve with a point so small that it can be found trivially in some other way, we find several points on the same four-descendent: for example, on the pair of quadrics

$$2wy - 2xz + z^2 = 2wx + 2wy + 2y^2 + 2yz - z^2 = 0,$$

associated to the curve $y^2 = x^3 - 15x - 10$ with rank 1 and generator $P = [-1, 2]$, a search up to $N = 10^4$ finds points [0 1 0 0], [1 0 0 0], [2 1 0 2], [2 −5 −6 2], [11 − 3618 − 6], [396 79 − 102 − 216], [828 − 1445 − 1734 − 3672]; these correspond to $P, -3P, 5P, -7P, 9P, -11P, 13P$ respectively.

Looking at such a series of points lets us examine the relationship between $\max \log |x_i|$ where $\mathbf{x} = (x_1, x_2, x_3, x_4)$ lies on $\mathcal{H}$, and $H(j_{\mathcal{H}\to E}(\mathbf{x}))$: as we expect, once we consider sufficiently large multiples of P_0, we have

$$H(j_{\mathcal{H}\to E}(\mathbf{x})) \approx 8 \max \log |x_i|$$

but if we define

$$f(n) = \frac{\max \log |x_i|}{n^2 H(P_0)} \text{ where } j_{\mathcal{H}\to E}(\mathbf{x}) = nP_0$$

we find $f(n)$ tends to 8 from above; the height-improvement factor is greater, the smaller the naïve height of $\mathbf{x}$.

2.10.3 How long does this all take?

To get some idea of the timings involved, consider the curve $E : y^2 = x^3 + 17x + 3137294$. `mwrank` takes 8.5 seconds to give $0 \leq r \leq \operatorname{rank} S_2 = 3$ and to give seven non-trivial two-coverings. `d4` takes 256.4 seconds to find four four-descendents on one of those two-coverings and to demonstrate that the others are insoluble, and a further 15.1 seconds to find a point (on the last four-covering

checked) which lifts to one on E. Of the 256.4 seconds, 38.2 seconds were spent computing class groups, 40.2 computing unit groups, 31.9 computing generators for $L'(S,2)$, 19.8 checking local solvability, 3.3 minimising four-coverings (there were only four to minimise), and 120.0 reducing the four four-coverings; the remaining 3.0 seconds were probably spent in initialisation.

For another example, consider $E : y^2 = x^3 + 17x + 3140360$. Again we have $0 \le r \le \operatorname{rank} S_2 = 3$ and seven two-coverings $\mathcal{C}_1, \ldots, \mathcal{C}_7$, but this time all seven two-coverings give four descendents $\mathcal{H}_{i1}, \mathcal{H}_{i2} \ldots$ each – this was computed in 193.1 seconds, but without reducing the descendents. We then start reducing descendents and looking for points on them; success strikes with $\mathcal{H}_{13}$, which after reduction has a very simple point which lifts to a point $P \in E(\mathbb{Q})$ of height 49.091; using the techniques in appendix B.1, we find as expected that $j_{\mathcal{C}}^{-1}(P)$ is defined over $\mathbb{Q}$ only on $\mathcal{C}_1$, and $j_{\mathcal{H}_{1i}}^{-1}(P)$ is defined over $\mathbb{Q}$ only for $\mathcal{H}_{13}$.

So far, we do not know whether E is of rank 1, or of rank 3 with some peculiarly enormous extra generators, and to resolve this purely by descent means would require a third descent, the techniques for which are not yet known. But, since we expect the rank to be one, we can use the theorem of Kolyvagin that a curve with the sign of the functional equation negative will have rank one if $L'(E,1) \neq 0$; after several hours, an analytic-rank computation using the program described in section 3.1.2 reveals $L'(E,1) \approx 273.2 \neq 0$ and so $\operatorname{rank} E = 1$. Our four-descent has calculated $Ш[4] = (\mathbb{Z}/4\mathbb{Z})^2$, and if we believe the BS-D conjecture we have $Ш_{\text{anal}} \approx 16.01$ and $Ш = (\mathbb{Z}/4\mathbb{Z})^2$.

The implementation of four-descent is sufficiently routine that it is used on several thousand curves in section 3.5, though there is no space to give any detail about any of those calculations.

2.10.4 Ridiculously large generators

For an extreme example of how effective four-descent can be at finding otherwise-hopelessly-large generators, consider the curve

$$E : y^2 = x^3 + 17x + 312521.$$

`mwrank` finds a single two-covering

$$Q : y^2 = -x^4 + 414x^2 + 1048x + 14351$$

in a few seconds, and `d4` takes 25.6 seconds, mostly spent computing the unit group of $\mathbb{Q}(\theta)$, to find the single four-covering

$$D = \left[\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \begin{pmatrix} 4 & 5 & 3 & 8 \\ 5 & -4 & 9 & -1 \\ 3 & 9 & 8 & -4 \\ 8 & -1 & -4 & -7 \end{pmatrix} \right]$$

with associated two-covering $Q' : y^2 = -x^4 + 8x^3 + 390x^2 - 576x + 13895$ equivalent to Q.

A search up to bound 10^6 takes 320.7 seconds to find the point

$$[450570 \quad 256794 \quad -565939 \quad -10772]$$

on D; this lifts to

$$\left[\frac{-635126027431097}{61394254976653}, \frac{-765829042088018884851454009986}{613942549766532} \right]$$

on Q', and to a point $(X/Z^2, Y/Z^3)$ on E of height 141.2113 with

$$\begin{aligned} X =& 1283141882627346107077039328768626146238488215504507162178947 2 \\ Y =& 5569959541791574233359120168461868433347482017663846852325915 9 \\ & 6099707828113469174680412856 51 \\ Z =& 3829145210440094424257270049 93 \end{aligned}$$

This should be compared with the largest heights practically accessibly by the 2-descent process, where finding a generator of height 80 involved a computation of several days.

Chapter 3

Survey work on the distribution of ranks and of Tate-Shaferevich groups

In this chapter, we introduce a new tool and an implementation of a well-known technique, and use them to study the distribution of ranks and of Tate-Shaferevich groups across families of elliptic curves with various properties. The research is prompted by the simple question "what proportion of elliptic curves have rank two". It does not answer this question – it does not even decide what the correct measure on elliptic curves to use to define "proportion" would be – but it gives reasonably compatible results from large searches on many different families.

The new tool is `ecsieve`, which can find very quickly the elliptic curves in a one-parameter family with many small integral points. On the output curves, we investigate the distribution of ranks and conductors, with the aim of finding curves with small – hopefully smallest-possible – conductor for their rank. To give some idea of how skewed our samples are, we also look at the distribution of small integral points on a random sample of elliptic curves.

The implementation work provides a respectably efficient way of evaluating L-series and their derivatives; this lets us compare, on large sets of examples, the $Ш[2]$ obtained by four-descent and the $Ш_{\mathrm{anal}}$ of the Birch–Swinnerton-Dyer formula.

3.1 Tools and implementations

3.1.1 The `ecsieve` algorithm

The algorithm works for families of curves of the form $y^2 = xf(x) + K$ or $y^2+y = xf(x)+K$, for $f(x)$ any polynomial function; in the following work $f(x)$ will be quadratic, though on a request from Stoll in December 2000 I performed some limited calculations with larger-degree f to find genus-2 and genus-3 curves with numerous integer points, in the expectation that the Jacobians of such curves might have unusually large rank. I found no examples of higher rank than Stoll already knew.

The method relies on the fact that enumerating the squares, or the numbers of the form y^2+y (an integer c is of that form iff $4c+1$ is a square) in an interval of $\mathbb{Z}$ is a very quick operation – for an interval $[a, b]$, we compute $y_a = \lceil\sqrt{a}\rceil$ and $y_b = \left\lfloor\sqrt{b}\right\rfloor$, and run through $\left\{t^2 : t \in [y_a, y_b]\right\}$ if $y_a < y_b$. For a short interval, which might contain not a single square, this takes roughly as long as two tests for squareness of an integer (since a squareness test, though it may begin with p-adic filters, must include the computation of a square root at some point).

So, for a sieving run, we fix bounds B_1 and B_2, an interval I_x of x coordinates to check, and a threshold value T. Squares are positive, and all our f will be monic quadratics so x^3 will be large and negative for x large and negative, so I_x will usually be narrow to the left of zero and wide to the right; see the example intervals used in the trials.

The run-time is proportional to the length of I_x but with a rather small constant of proportionality. The run time is fairly insensitive to B_2, but the memory usage is $2B_2$ bytes. It is possible to reduce this memory usage by any factor desired, at the price of an equivalent increase in execution time.

We then run through $a_2 \in \{-1, 0, 1\}$ and $a_4 \in [-B_1, B_1]$. Then, for each $x \in I_x$, we compute $A = x^3 + a_2x^2 + a_4x$, construct the set

$$S = \{s : s = t^2, t \in \mathbb{Z}, |s - A| < B_2\},$$

and increment a counter $c[B]$ whenever $A + B$ lies in S. After considering all the x values, we look for counters holding values greater than or equal to T; the index of the counter then gives us a curve $y^2 = x^3 + a_2x^2 + a_4x + B$ with many integral points.

We also construct the set $S' = \{s : s = t^2 + t, t \in \mathbb{Z}, |s - A| < B_2\}$, and use the same sieving techniques to find curves $y^2 + y = x^3 + a_2x^2 + a_4x + B$ with many integral points.

The algorithm is implemented in C++, and takes approximately 20 CPU-seconds on a P4/2266 with DDR memory for a single a_4 value, all six possibilities

of a_2 and a_3, the I_x above, and $B_2 = 2^{26}$. The total number of curves sampled is $24B_1B_2$.

The performance is entirely limited by memory bandwidth; the counters that are incremented tend all to lie in different cache lines, and whilst they are accessed in increasing order in memory, they are distributed sufficiently randomly around the whole array that the cache does not help.

One implementation subtlety is that c is an array of bytes (to allow B_2 to be taken large), and so in principle, and in practice for singular curves when I_x is long, it might overflow; so if $c[B]$ ever reaches 254 we never increment it again. We check at the end that every curve with $c[B] = 254$ is singular, and do not bother announcing singular curves.

It does not seem possible to extend this algorithm to handle curves with $a_1 \neq 0$; that is, curves of the form $y^2 + xy + y = f(x)$. So, in my survey work, I encounter such a curve only when I find a curve isomorphic to it and with $a_1 = 0$: comparing my results with some unpublished work of Watkins [74] in section 3.2.8, it is clear that this was a more severe obstruction than I had expected.

3.1.2 Analytic ranks

Buhler, Gross and Zagier [12] note that, for an elliptic curve E of conductor N and rank $\leq u$ and of the same parity, we have the sum

$$L^{(u)}(1) = \sum_{n=1}^{\infty} \frac{a_n}{n} G_u(n\sqrt{N}),$$

where a_n are the coefficients of the modular form associated to E, computed as a multiplicative function after the a_p are calculated as $\|E(\mathbb{F}_p)\| = p + 1 - a_p$ for p of good reduction and by Tate's formula otherwise,

$$G_u(x) = P_u(-\log x) \sum_{j=1}^{\infty} (j^u j!)^{-1} x^j$$

and $P_u(x)$ is a polynomial with coefficients taken from the Taylor expansion of $\Gamma(1 + x)$. They introduce a recursive method which evaluates $\sum_{i=1}^{K} \theta(i) f(i)$ for a multiplicative function θ using only $O(\sqrt{K})$ storage; since the baby-step-giant-step method for computing a_p takes time $O(p^{1/4})$, we have an algorithm taking roughly $O(B^{5/4})$ time to compute the sum of the first B terms of the series.

Setting $B = \sqrt{N}$ gives accuracy of around six decimal places and reasonable calculation time; I implemented this algorithm in Pari, where it is usable on a

modern PC for curves with conductors up to about 10^{14}. I announced it on the Pari mailing list, and made my implementation available for download from my Web pages (`http://www.maths.nottingham.ac.uk/personal/pmxtow/BG.gp`).

Given a technique for computing L-series derivatives, we can non-rigorously calculate analytic ranks; we set $u = 0$ if the functional equation has sign $+1$ and $u = 1$ otherwise, compute $L^{(u)}(1)$, and calculate the pseudo-regulator

$$\xi = \frac{L^{(u)}(1)|E_{\mathrm{tors}}|^2}{\Omega \prod c_p}.$$

If this has absolute value less than 10^{-4} – picked to be smaller than the smallest regulator I have ever observed, which is 0.008914, for the rank-one conductor-3990 curve

$$y^2 + xy + y = x^3 + x^2 - 125615x + 61201397$$

with generator $(7107, -602054)$ from Cremona's tables [20] – we assume it is identically zero, increase u by two, and try computing $L^{(u)}(1)$ again. Otherwise, we assert that u is the analytic rank, and state that R_EШ is equal to the pseudo-regulator ξ.

3.2 Curves with many integral points

3.2.1 Experiments performed

Two substantial runs of `ecsieve` were performed: `small`, where $B_1 = 2^{16}$, $B_2 = 2^{24}$, $I_x = I_1 = [-1000, 20000]$, $T = 24$, and `large`, where $B_1 = 2^{17}, B_2 = 2^{26}$, $I_x = I_2 = [-2^{15}, 2^{20}]$, $T = 40$. `large` took roughly four weeks to compute, using two P4/2266 computers and two 833MHz Alpha processors.

`small` sampled about 2.6×10^{13} curves, producing 5.9×10^6 with point counts exceeding the threshold. Using `pari`, I computed the conductor and the sign of the functional equation for all the 877398 curves with ≥ 31 points; using `mwrank`, I computed the rank and generators for all 70767 pairwise non-isomorphic curves with ≥ 41 points. All these curves had rank ≥ 4, and on each of them a full set of generators was found.

`large` sampled 2.1×10^{14} curves, of which 1551130 had ≥ 40 points. I computed conductors and signs for all of these, and ranks and generators for the 9832 with ≥ 68 points (all had rank ≥ 5, and for each of them a full set of generators was found); this latter calculation itself took about two CPU-weeks.

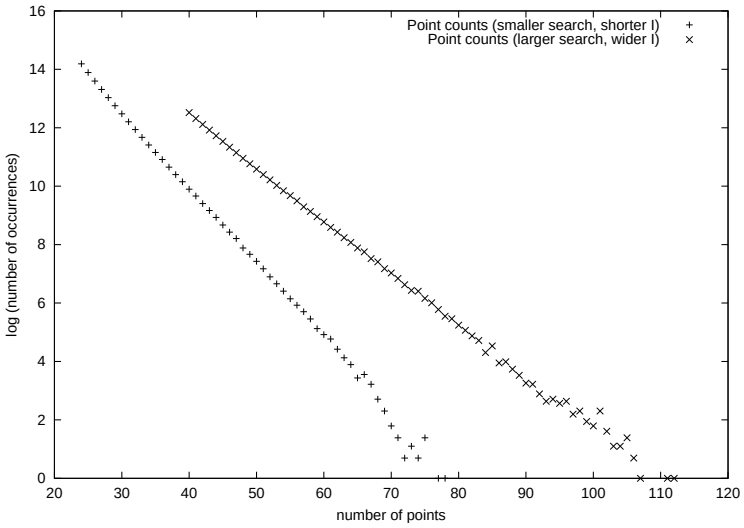


Figure 3.1: Proportion of curves from the two samples that had a given point count. The two lines are not honestly comparable, since I_x differs between the samples.

3.2.2 How are the point counts distributed?

Figure 3.1 gives the proportion of curves from each of the two samples that had a given point count; you will observe that the distribution is close to exponential.

These results are very much a large-deviation sample, so I also computed point-counts, for the shorter interval I_1, for a random sample of a million curves from the region investigated in `small`, obtaining the results shown in figure 3.2. The error bars here are derived by assuming that 'having j points' is an event with some probability p_j; our million observations give a binomially-distributed random variable, which is approximated as normal with mean Np_j and standard deviation $Np_j(1-p_j)$; that is, for given N and p, there is a 95% probability that

$$S_N \in \left(Np_j - 1.96\sqrt{Np_j(1-p_j)}, Np_j + 1.96\sqrt{Np_j(1-p_j)} \right).$$

Observe that by far the most common count of points is 0.

3.2.3 What effect does changing I_x have?

The searches `small` and `large` were performed on overlapping regions but with different I_x intervals. I therefore computed, for each of the 71378 curves of known rank from `small`, how many extra points would have been obtained had we used the wider intervals, and plotted this to give figure 3.3.

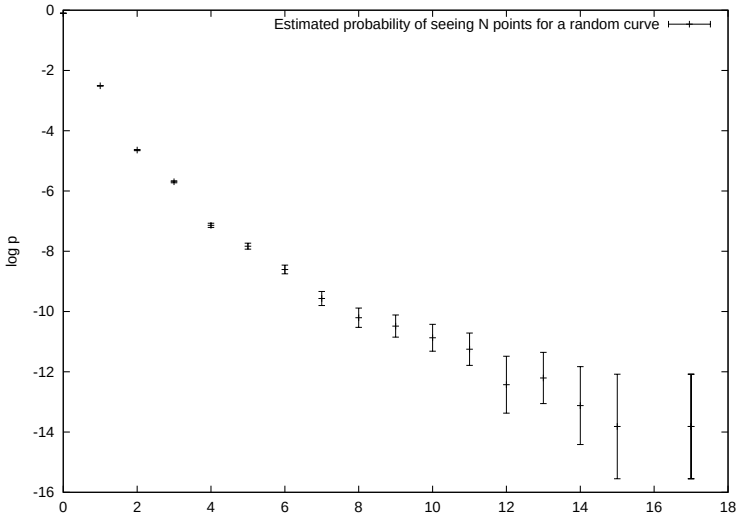


Figure 3.2: Estimated probability of observing N points on an elliptic curve with a_2, a_3, a_4, a_6 randomly selected from the region investigated in study `small`. Note that the $N = 0$ point is in the top left corner, almost indistinguishable from the axis markers

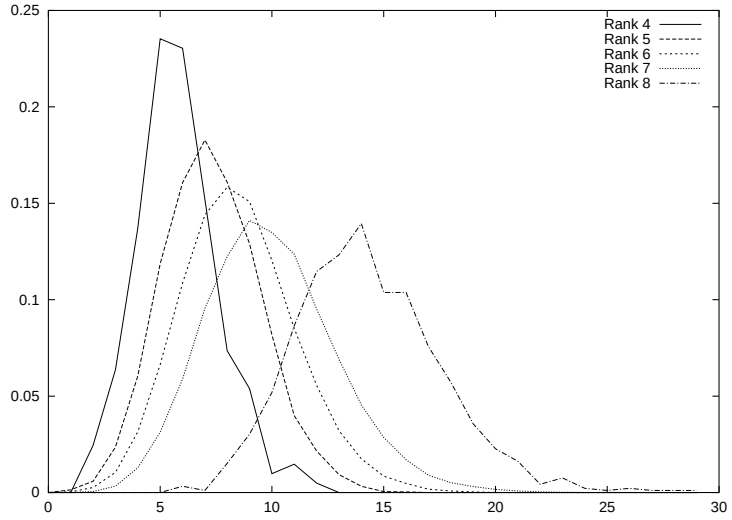


Figure 3.3: Probability density functions for the number of points with x coordinate in $[2 \times 10^4, 2^{20}]$, on the curves with ≥ 41 points in $[-10^3, 2 \times 10^4]$, split by rank.

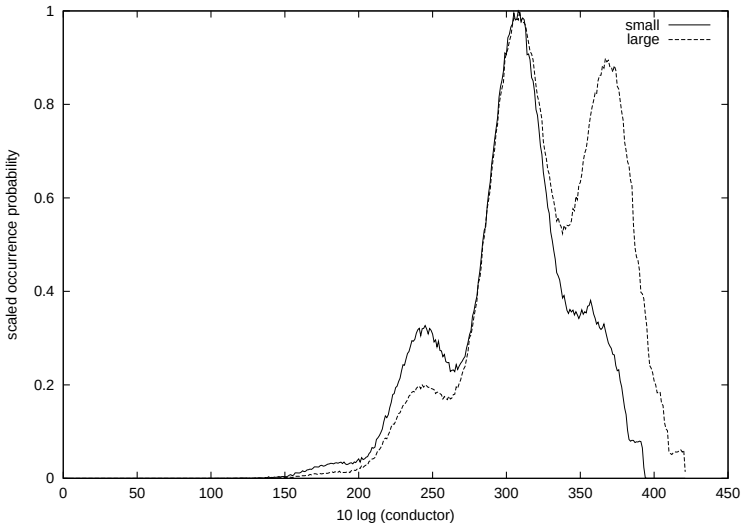


Figure 3.4: Probability density functions for $\log N$, for the small and large samples

The additional-point count seems to be roughly Poisson distributed, with a parameter that increases as the rank goes up, though the curves overlap sufficiently that it is not possible reliably to predict rank by observing point-count augmentation.

3.2.4 How are conductors distributed for curves with many integral points?

Figure 3.4 shows the distribution of the conductors for the curves of `small` and `large`, obtained by sorting $\log N$ into buckets of width 0.1 and scaling so that the largest bucket has size 1. Its multi-modality is very obvious, indicating that it is probably the sum of several separate distributions. If we distinguish among curves by the sign of the functional equation, we get figure 3.5, and it appears safe to conjecture that each peak corresponds to a separate rank. Note that the position of the peaks seems to remain fixed between the two samples, except that the rightmost peak (rank 8) has moved towards larger conductors in the larger sample.

3.2.5 The distribution of conductor with rank

We have not computed ranks for all of the curves from `small` or `large`: for the latter, `mwrank` takes about one minute per curve and so we work with only the 9832 curves with more than 78 points. But, working only with the curves whose

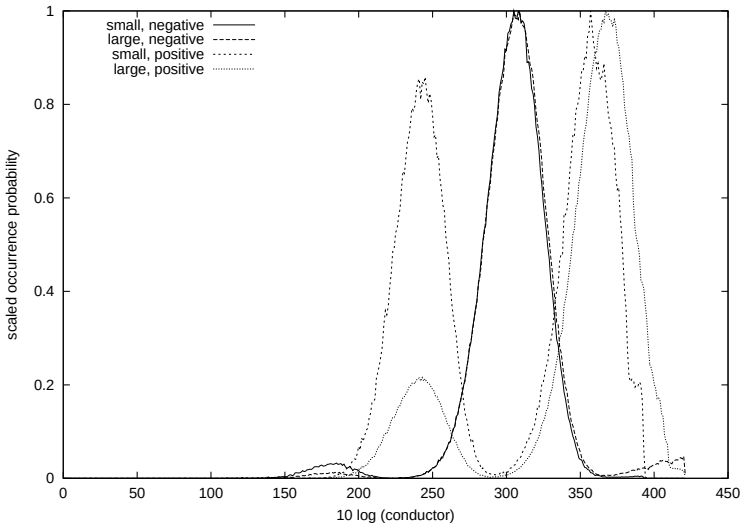


Figure 3.5: Probability density functions for $\log N$, for the small and large samples, and the two different signs of the functional equation

rank we know, we can construct figures 3.6 and 3.7.

3.2.6 Distribution of point count with rank

Figure 3.8 displays the distribution of point-counts for the curves of ranks 5, 6 and 7 from `small` – there were not enough curves of larger rank for the display to make sense. You will notice that each of the distributions is approximately exponential, with a slope decreasing as the rank does – observe the crossing-over of the rank-6 and rank-7 curves: this explains why looking at high-point-count curves serves to give us examples of large rank, since, as we increase the point-count threshold, a larger proportion of the curves at that point count come from the higher-ranked distributions.

3.2.7 Smallest-observed conductors for given ranks

Table 3.1 is obtained by combining information from many sources, and lists the five curves of smallest conductor that I found for each rank between 3 and 9.

The rank-three curves are those of smallest conductor from Cremona's [21] tables, which exhaustively enumerate elliptic curves with conductor up to 20000: there are four more such curves with N between 13766 and 20000. Some of the rank-four and rank-five curves arise from unpublished work of Cremona, who

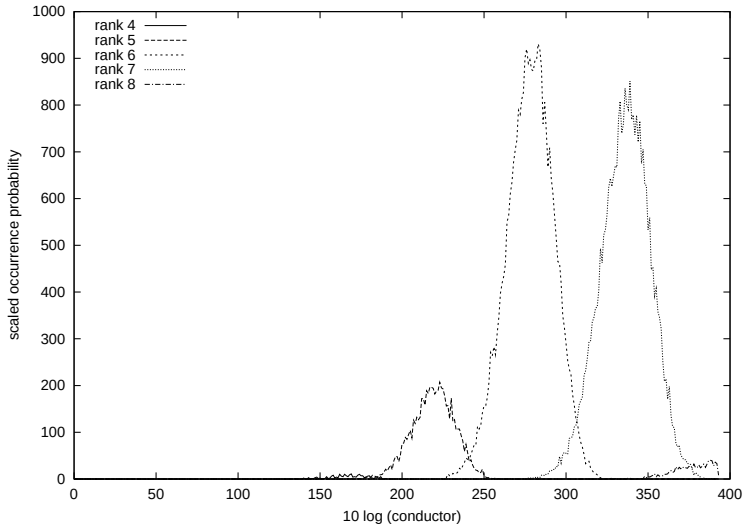


Figure 3.6: Probability density functions for $\log N$, for the small sample, split out by rank

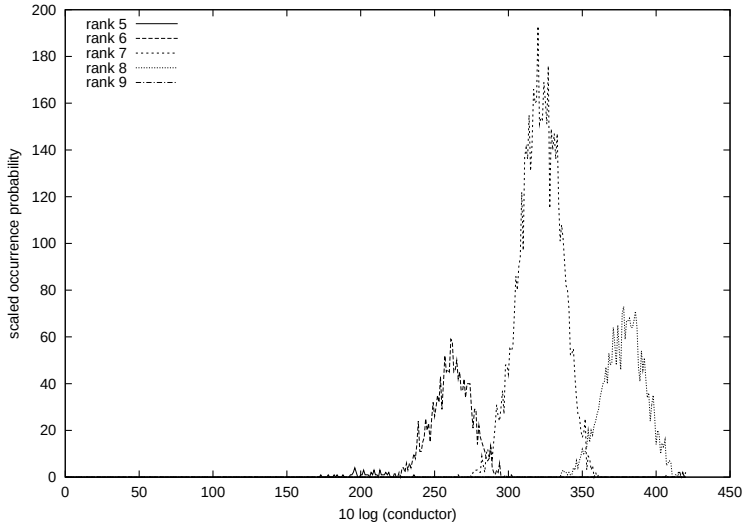


Figure 3.7: Probability density functions for $\log N$, for the large sample, somewhat partial, and split out by rank

	Curve	Conductor
Rank 3 :	$y^2 + y = x^3 - 7x + 6$	5077
	$y^2 + xy + y = x^3 - x^2 - 6x$	11197
	$y^2 + xy = x^3 - x^2 - 16x + 28$	11642
	$y^2 + y = x^3 - x^2 - 10x + 12$	12279
	$y^2 + xy + y = x^3 - 23x + 42$	13766
Rank 4 :	$y^2 + xy = x^3 - x^2 - 79x + 289$	234446
	$y^2 + y = x^3 + x^2 - 72x + 210$	501029
	$y^2 + y = x^3 - 7x + 36$	545723
	$y^2 + xy + y = x^3 - 35x + 90$	556838
	$y^2 + xy = x^3 - x^2 - 34x + 64$	614066
Rank 5 :	$y^2 + y = x^3 - 79x + 342$	19047851
	$y^2 + xy = x^3 - 22x + 219$	20384311
	$y^2 + y = x^3 - 247x + 1476$	22966597
	$y^2 + xy = x^3 - x^2 - 415x + 3481$	34672310
	$y^2 = x^3 - 532x + 4420$	37396136
Rank 6 :	$y^2 + xy = x^3 + x^2 - 2582x + 48720$	5187563742
	$y^2 + y = x^3 - 7077x + 235516$	5258110041
	$y^2 + xy = x^3 - x^2 - 2326x + 43456$	5739520802
	$y^2 + y = x^3 - 547x - 2934$	6756532597
	$y^2 + xy = x^3 - x^2 - 1486x + 21688$	6895251302
Rank 7 :	$y^2 = x^3 - 10012x + 346900$	382623908456
	$y^2 + y = x^3 - 36673x + 2704878$	814434447535
	$y^2 + xy + y = x^3 - 5983x + 164022$	1005276094726
	$y^2 = x^3 - 101647x + 12379090$	1022298908216
	$y^2 = x^3 - 12979x + 405826$	1074680679376
Rank 8 :	$y^2 + y = x^3 - x^2 - 68520x + 6724532$	395623692381639
	$y^2 + y = x^3 + x^2 - 23846x + 1022562$	409086620841461
	$y^2 = x^3 - x^2 - 98310x + 12416121$	452976789140724
	$y^2 + y = x^3 - 23737x + 960366$	457532830151317
	$y^2 + y = x^3 - 63973x + 6649278$	468796409109295
Rank 9 :	$y^2 + y = x^3 - 121849x + 38046702$	509558981564991851
	$y^2 + y = x^3 + x^2 + 44518x + 49327152$	1056142691473638331
	$y^2 + y = x^3 + x^2 - 11720x + 49465500$	1057096963922857483
	$y^2 + y = x^3 + x^2 - 27980x + 52049010$	1169348464809508603
	$y^2 + y = x^3 - 121387x + 55813176$	1231256525818173083

Table 3.1: The five curves of smallest conductor at each rank between 3 and 9, observed in the tables [20], in Cremona's search of $|a_4|, |a_6| \leq 500$, and in the searches `small` and `large` of section 3.2.1

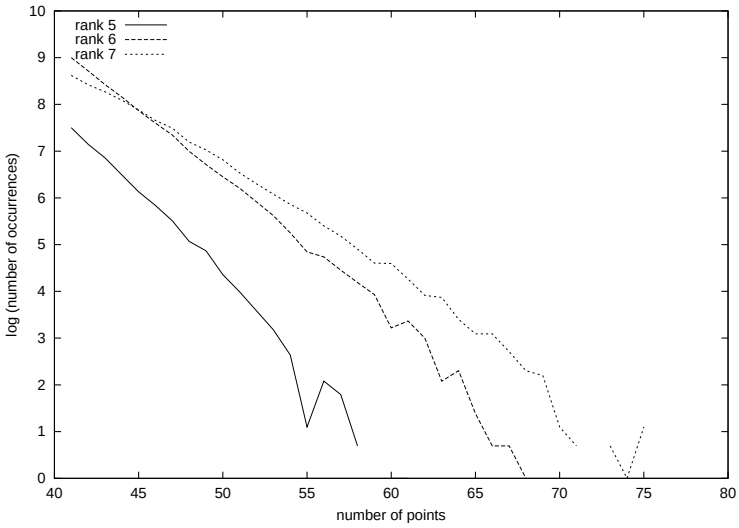


Figure 3.8: Distribution of point-counts for the three ranks for which we have an adequate sample of curves from `small`

applied `mwrank` to all the curves with

$$a_1 \in \{0,1\}, a_2 \in \{-1,0,1\}, a_3 \in \{0,1\}, |a_4| \leq 500, |a_6| \leq 500.$$

No curves in that region have 2-Selmer rank greater than five.

3.2.8 Other work in this field

Stein and Watkins have been endeavouring to construct very large tables of elliptic curves with interesting properties: a preliminary report appears in [68]. In mid-2002, independently of my work, Watkins [74] used tests which he called the H^8 and H^9 methods to look for curves of high rank. These found a number of extremely good examples, including ones better than any my search offered at ranks 8 and 9, and ones missing from my top-five table at ranks 6 and 7; I list them in table 3.2. It is reassuring that his table of rank-five curves of low conductor agrees exactly with mine.

The curves that appear in table 3.2 are ones for which the minimal model has $a_1 \neq 0$; the substitution required to fix this increases the coefficients so much that the a_4 and a_6 become too large to be found by my search.

3.2.9 What about regulators?

From `small`, I have full sets of generators for 70,767 curves. I computed regulators, smoothed the data by computing the sum of a narrow Gaussian centred

	Curve	Conductor
Rank 6	$y^2 + xy = x^3 + x^2 - 2582x + 48720$	5187563742
	$y^2 + y = x^3 - 7077x + 235516$	5258110041
	$y^2 + xy = x^3 - x^2 - 2326x + 43456$	5739520802
	$y^2 + xy = x^3 - x^2 - 16249x + 799549$	6601024978
	$y^2 + xy + y = x^3 - x^2 - 63147x + 6081915$	6663562874
Rank 7	$y^2 = x^3 - 10012x + 346900$	382623908456
	$y^2 + xy + y = x^3 - 14733x + 694232$	536670340706
	$y^2 + y = x^3 - 36673x + 2704878$	814434447535
	$y^2 + xy = x^3 - x^2 - 92656x + 10865908$	858426129202
	$y^2 + xy = x^3 - x^2 - 18664x + 958204$	896913586322
Rank 8	$y^2 + xy = x^3 - x^2 - 106384x + 13075804$	249649566346838
	$y^2 + xy = x^3 - x^2 - 71899x + 5522449$	314658846776578
	$y^2 + xy = x^3 - x^2 - 124294x + 14418784$	315734078239402
	$y^2 + y = x^3 - 135109x + 18252072$	323954505704623
	$y^2 + xy = x^3 + x^2 - 69607x + 6711985$	325724094713742
Rank 9	$y^2 + xy = x^3 - x^2 - 139246x + 36766576$	205034814784919398

Table 3.2: The five curves of smallest conductor, for ranks between 6 and 8, obtained by the "H^8" and "H^9" methods of Watkins [74]

at each of the $\log r$, and, in figure 3.9, plotted $\mathrm{P}(\log r = N)$ for the rank 5, 6, 7, 8 curves, scaled so that the highest peak has height 1. Similar data from the 9,832 curves from `large` appears in figure 3.10.

The multi-modal distributions, looking similar for each rank (if we translate to line up the highest peaks, the other peaks align very well), are very striking: they are unlikely simply to indicate that I sometimes computed a generating set of index > 1 in $E(\mathbb{Q})$, since the primary and secondary peaks are too far apart to correspond to indices 1 and 2. The separation between the primary and secondary peaks is consistently around 2.15, corresponding to a factor between 8 and 9 in the regulator; there is no striking difference in the distribution of the a_1, a_2, a_3 coefficients between the curves with low and high regulators of a given rank (where 'low' corresponds to the first peak and 'high' to the region after the first peak).

For the datasets of [20], you observe in figure 3.11 that the distributions for each rank look much more unimodal; for the curves from section 3.5, the data was so noisy that I smoothed it with a Gaussian of twice the width of that used for the other regulator-distribution graphs, and again the distribution at each rank looks unimodal, though the peak at the left-hand side of the rank-1 graph in figure 3.12 appears to be a real feature of the data.

For [20], indeed, we find that the peak of the rank-one distribution is to the right of that for rank two: but the ratio of rank-one to rank-two curves in that data set seems very atypical. For the section 3.5 curves, of course, there is the

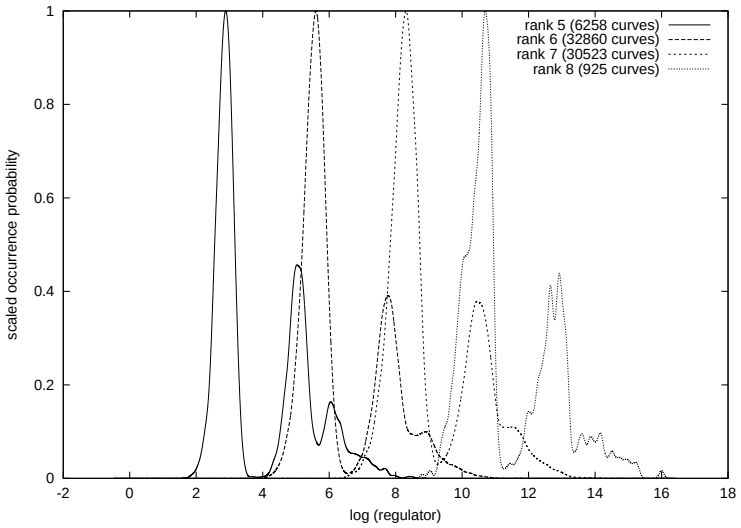


Figure 3.9: Distribution of regulators for the curves from `small` of various ranks.

problem that we do not have regulators for the curves on which we failed to find points, which we expect to be precisely those with the largest points and hence probably the greatest regulators.

3.3 Curves with non-trivial torsion groups

There are finitely many possible torsion subgroups for an elliptic curve defined over $\mathbb{Q}$, and the elliptic curves with a given non-trivial torsion subgroup lie in a one-parameter family. Explicit descriptions of this family – that is, explicit functions $E_G(t)$ where $E_G(t)$ has torsion subgroup G for all but finitely many t – are given in a paper of Kubert [46].

The coefficients of $E_G(t)$ can be quite large; for $G = \mathbb{Z}/12\mathbb{Z}$, the discriminant for large t is proportional to t^{10}, and its denominator is proportional to t^{24}, whilst the discriminant of the minimal model is proportional to t^{48}. However, the factorisation of the discriminant of a minimal model often involves very large powers, and so, for t of small naïve height, the conductor is often surprisingly accessible. I computed analytic ranks for $E_G(t)$ for all permitted G of size ≥ 4 (apart from $(\mathbb{Z}/2\mathbb{Z})^2$) and for $t = \pm n/d$ with $n + d \leq 200$, looking for the curve of smallest conductor with a given rank and torsion group.

I found curves of rank two with all the torsion groups, and for $\mathbb{Z}/4\mathbb{Z}$, $\mathbb{Z}/5\mathbb{Z}$, $\mathbb{Z}/6\mathbb{Z}$, $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/6\mathbb{Z}$ I found curves of rank three; I also found a number of quite large values of $Ш_{\mathrm{anal}}$. However, my results have been significantly improved on by Kulesz [47], who has found families defined over

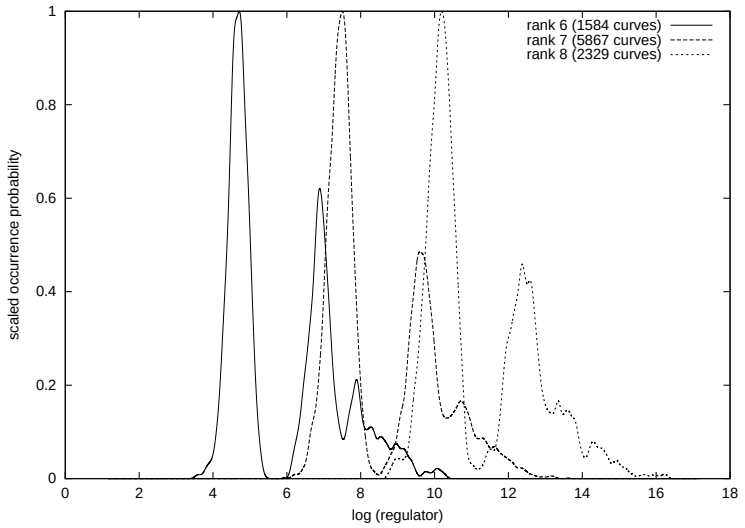


Figure 3.10: Distribution of regulators for the curves from `large` of various ranks.

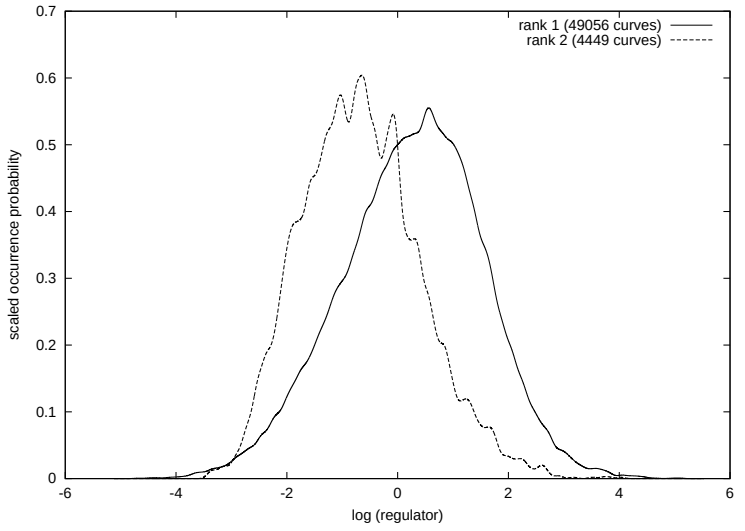


Figure 3.11: Distribution of regulators for the curves from [20] of various ranks.

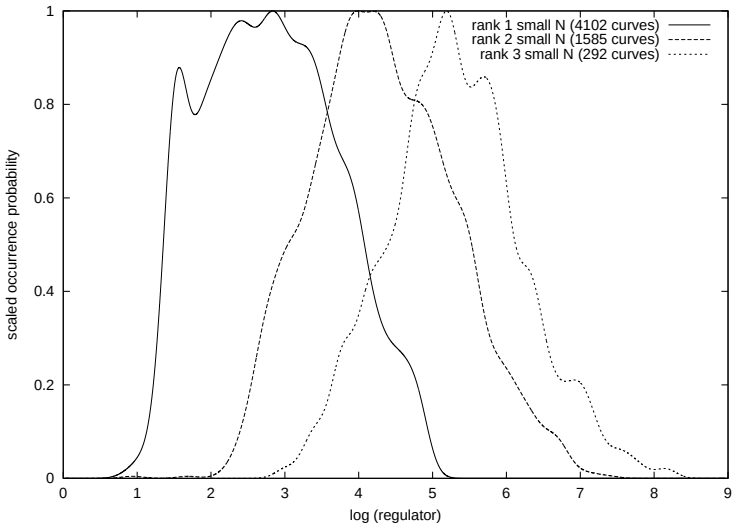


Figure 3.12: Distribution of regulators for the curves from section 3.5 of various ranks, listing only curves for which all generators are known

$\mathbb{Q}(t)$ of elliptic curves of rank ≥ 1 and given torsion group, and Dujella; Dujella maintains a Web site [31] giving the best-known ranks for each torsion group.

3.4 Mordell curves

3.4.1 Finding large-rank examples with `ecsieve`

The `ecsieve` program works very well for the Mordell curves $y^2 = x^3 + K$, so I examined these for all $|K| < 2^{40}$, counting integer points with x co-ordinate between -10400 and 2^{20}. These Mordell curves for $|K| < 10^4$ were studied exhaustively in a paper of Gebel, Petho and Zimmer [39], who later extended their search to $|K| < 10^5$ and found the smallest examples at rank 5. Some generators were missing, and were found in later work by Wildanger [76].

Notice that the point (a^2x, a^3y) on $y^2 = x^3 + a^6K$ gives the point (x, y) on $y^2 = x^3 + K$, and that, since $y^2 = x^3 + K$ has complex multiplication by $\sqrt{-3}$, there exists a 3-isogeny between the curves $y^2 = x^3 + K$ and $y^2 = x^3 - 27K$: explicitly, if we take

$$x' = x + \frac{4K}{x^2}, y' = \frac{y^3 - 9ky}{x^3}$$

we have ${x'}^3 - 27K = {y'}^2$.

Using all $|K| < 2^{40}$, rather than filtering to cube-free K, had the same effect as including points with small denominators; for example, $y^2 = x^3 + 47550317$ has only seven points with integer x co-ordinate in the range searched, whilst

$y^2 = x^3 - 3^9 \times 47550317$ has 28, so I found the latter curve and was then led to consider the former.

The discriminants of the curves with many integer points were in general too large for a two-descent to finish in a reasonable length of time, whilst their conductors were much too large for it to be feasible to compute the analytic rank using the $O(N^{1/2})$ algorithm of section 3.1.2. Instead, `findinf` was used to find independent rational points of small height, and the sign of the functional equation was computed; assuming the parity conjecture, this lets us know whether `findinf` has missed out one point.

The sign of the functional equation was also useful in selecting where to search; recall from figures 3.6 and 3.7 that, at least under the situation prevalent there, the distributions of conductor with rank do not overlap much between ranks r and $r+2$. Similarly we find that, when one lists Mordell curves with > 20 rational points and sorts by conductor, the curves of a given rank tend to clump together, and so the sign of the functional equation tends to be constant for long intervals. Hence, if we find a curve with $\varepsilon = +1$ in the middle of a long run of $\varepsilon = -1$, this indicates a rank either one more or one less than the 'expected' rank for that conductor, so worth investigating.

The results of the search are presented as part of table 3.4.1. The largest rank of a curve that I found was nine; Mordell curves of rank > 9 are known, from constructions by Quer [53] and Elkies [32], but Quer's constructions gave substantially larger values of K at a given rank than appeared in my search.

Finding points on a Mordell curve can also be considered as a problem over a quadratic number field – $y^2 = x^3 + K$ means $(y-\theta)(y+\theta) = x^3$ in $\mathbb{Q}(\theta) : \theta^2 = K$; it is this kind of argument that [49] and [53] rely on, and they find the elliptic curves incidentally, their goal being imaginary quadratic number fields whose class group had high 3-rank.

3.4.2 Deeper investigations with two-descents

The N values for the minimal positive and negative Mordell curves of rank six found by the sieving process were small enough that it was feasible to perform a two-descent on every N in the range $[-1000000, 1400000]$, to confirm that no other rank-six curves exist in this range. Using `mwrank`, with its default bounds for point-search on the two-coverings, it took two months on three 850MHz Athlon computers to perform the descents and obtain the results in table 3.4. The lower bound in that table is the number of generators found for the curve: all these generators are recorded, and the data is available on CDR by request.

To continue the exhaustive search up to the K of the rank-7 curve would be computationally difficult, though not completely impractical – I estimate it

Rank	Smallest positive	Largest negative	Reference
0	1	-1	trivial
1	2	-2	long-known
2	15	-11	long-known
3	113	-174	[39]
4	2089	-2351	[39]
5	66265	-28279	[39]
6	1358556	-975379	this work
7	47550317	-56877643	this work
8	1632201497 ([32] 1999)	-2520963512 (this work)	
9	185418133372	-463066403167	this work
10	68513487607153 ([32] 2001)	-56736325657288 ([49])	
11	35470887868736225 ([32] 1999)	-46111487743732324 ([53])	
12	176415071705787247056	-6533891544658786928	[53]

Table 3.3: Smallest-known K with $y^2 = x^3 + K$ of a given rank. For rank ≤ 6, these are smallest-possible. If a reference is given in the "Reference" column it applies to both entries; if by a single entry, it applies to that one only

would take three hundred 2GHz computers a year, which is not a hopelessly large amount of computation compared with the `GIMPS`, `Seti@Home` or `RC5/64` projects. Verifying rank 8 seems utterly impractical with current algorithms and foreseeable future computers.

3.4.3 Exploring the inexact realm using four-descents

As in the caption of table 3.4, I call a curve inexact if the number of generators found is not equal to the 2-Selmer rank. There are obviously too many inexact curves to perform four-descents on all of them, so I picked one thousand examples uniformly at random from the 248,078 curves with rank bound $[0, 2]$. Each example gave three quartics; if none of them gave a four-descendent we have $Ш[2^\infty] = (\mathbb{Z}/2\mathbb{Z})^2$, and otherwise all three will give two four-descendents, telling us that $Ш[4] = (\mathbb{Z}/4\mathbb{Z})^2$. In that case, I computed $Ш_{\text{anal}}$ using the program of section 3.1.2, allowing me to distinguish between $Ш[2^\infty] = (\mathbb{Z}/4\mathbb{Z})^2$ and $Ш[2^\infty] = (\mathbb{Z}/8\mathbb{Z})^2$ – all the analytic ranks were zero.

Sha structure	With $N < 0$	With $N > 0$	Total
$\mathbb{Z}/2\mathbb{Z}^2$	391	432	823
$\mathbb{Z}/4\mathbb{Z}^2$	94	73	167
$\mathbb{Z}/8\mathbb{Z}^2$	9	1	10

These were very much not the results I expected; I had expected $(\mathbb{Z}/4\mathbb{Z})^2$ to be rare and $(\mathbb{Z}/8\mathbb{Z})^2$ vanishingly so, and I was surprised that a χ^2 test on the table gives $p = 0.0042$, indicating that the distribution of Ш structures for Mordell curves is significantly different between positive and negative N values. There are, of course, more positive than negative N values since the range is not

Rank bound	**Number with $K > 0$**	**Number with $K < 0$**	**Total**	**Proportion among $K > 0$**	**Proportion among $K < 0$**
[0]	297831	247366	545197	0.213	0.247
[1]	514000	272450	786450	0.367	0.272
[2]	208450	147632	356082	0.149	0.148
[3]	40036	26371	66407	0.0286	0.0264
[4]	3679	2547	6226	0.00263	0.00255
[5]	173	98	271	0.00012	0.00010
[6]	1	1	2	7×10^{-7}	10^{-6}
inexact	**335830**	**303535**	**696365**	**0.240**	**0.304**
[0, 1]	141103	136452	277555	0.101	0.136
[1, 2]	9388	14018	23406	0.0067	0.014
[2, 3]	95	77	172	6.7×10^{-5}	7.7×10^{-5}
[0, 2]	127622	120456	248078	0.091	0.120
[1, 3]	52706	26639	79345	0.0376	0.0266
[2, 4]	2703	2062	4765	0.0019	0.0021
[3, 5]	46	16	62	3.2×10^{-5}	1.6×10^{-5}
[0, 3]	1203	2595	3798	0.00086	0.0026
[1, 4]	2	20	22	1.4×10^{-6}	2×10^{-5}
[0, 4]	901	1189	2090	0.00064	0.0012
[1, 5]	61	11	72	4.4×10^{-5}	1.1×10^{-5}

Table 3.4: Results of 2.4 million 2-descents on curves $y^2 = x^3 + K$, for $K \in [-10^6, 1.4 \times 10^6]$. If the number of generators found is equal to the 2-Selmer rank r_2, I write $[r_2]$; if not, I write $[r_1, r_2]$ where r_1 generators were found, and call the curve inexact.

symmetrical around zero; but restricting to $|N| \leq 10^6$, we find an even stronger effect ($p_{\chi^2} = 0.0011$):

Sha structure	**With** $N \in [-10^6, 0)$	**With** $N \in (0, 10^6]$	**Total**
$\mathbb{Z}/2\mathbb{Z}^2$	391	309	700
$\mathbb{Z}/4\mathbb{Z}^2$	94	40	134
$\mathbb{Z}/8\mathbb{Z}^2$	9	1	10

Christophe Delaunay, working by analogy with the results of Cohen-Lenstra for the class groups of quadratic fields given in [17], gave in his thesis [29] an analytic argument that, for a curve of Mordell-Weil rank zero, $Ш_2 \neq \{1\}$ with probability 0.580577: this is entirely incompatible with my experimental evidence. He also claimed that $Ш[p^\infty]$ is isomorphic to $(\mathbb{Z}/p^2\mathbb{Z})^2$ precisely $1/p$ as often as it is isomorphic to $(\mathbb{Z}/p\mathbb{Z})^2$; this is also unsupported.

Enlivened by the above findings, I turned my attention to the curves with 2-Selmer rank 4 on which I had found either 0 or 2 generators. There are 2090 of the former type and 4765 of the latter; I picked 200 of each type uniformly at random, calculated the quartics, and performed four-descents on all of them, computing $\mathfrak{D}_4^{\text{alg}}$ but not calculating explicit matrices.

For the $[2, 4]$ curves, most (502/600) of the two-coverings turned out to be resolvable, so little local-solvability work was needed; 72 of the 98 irresolvable two-coverings were immediately shown not to be solvable at 2. The whole calculation took about 80 minutes, and there was not a single four-descendent among all 600 two-coverings examined; the most time spent on one two-covering was about five minutes, almost entirely on computing the unit group of the number field L. Overall, 45% of time was spent on computing class groups, 40% on unit groups, 12% on calculating 2-Selmer groups in the number field, and the rest mostly on local solvability.

For the sample of $[0, 4]$ curves, there are of course 3000 non-trivial two-coverings to consider, so the calculation takes a good deal longer, and we run across situations where the class group calculation alone takes several hours. Again, most (2372/3000) of the two-coverings were resolvable; this time, 65% of the time was spent on class groups, 13% on unit groups, and 22% on 2-Selmer groups. The whole run took a little under 26 hours; on seven of the curves (three with $N < 0$, four with $N > 0$), three of the two-coverings gave eight four-descendents each, indicating $(\mathbb{Z}/2\mathbb{Z})^2 \times (\mathbb{Z}/4\mathbb{Z})^2 \leq Ш[2]$. Analytic rank calculations gave $Ш_{\text{anal}} = 64$ for all those curves.

3.5 A Mordell-like family, avoiding the origin

One can argue that Mordell curves are not good examples of random curves, if only because they all have complex multiplication by $\mathbb{Q}[\sqrt{-3}]$. Also, all my

Selmer rank bounds	**Number in $K_0 = 314159$**	**Number in $K_0 = 3141592$**
0	2255	2279
1	3138	2521
2	1517	1273
3	292	255
4	24	28
5	1	2
$[0,1]$	1201	1729
$[1,2]$	53	146
$[2,3]$	0	1
$[0,2]$	1086	1258
$[1,3]$	361	381
$[2,4]$	29	30
$[3,5]$	0	3
$[0,3]$	31	79
$[0,4]$	12	16
$[1,5]$	1	0
Total inexact	**2774**	**3643**

Table 3.5: Results from two-descent on curves of the form $y^2 = x^3 + 17x + K_0 + N, N \in [-5000..5000]$

experiments have suggested that the distribution of ranks and of Ш is not uniform, particularly not near the origin; I wanted to construct two samples so that I could test if the distributions observed were statistically similar.

So I considered two samples of curves of the form $y^2 = x^3 + 17x + K_0 + K$ for $K \in [-5000, 5000]$, with $K_0 = 314159$ and $K_0 = 3141592$, and performed 2-descents with `mwrank`'s default bounds; for the smaller K_0 a P4/2400 computer manages approximately 90 2-descents per CPU-minute, for the larger, approximately 12.5. I recorded all the everywhere-locally-solvable two-coverings on which no global point was found, for later four-descent. The results from this initial two-descent are presented in table 3.3.

3.5.1 Results from four-descent

For all the curves for which the 2-Selmer rank was not equal to the number of generators found, I performed four-descents, though not all were conclusive; thanks to time limitations, I often allowed only two minutes of CPU time per curve, and thanks to a bug in the factorisation of polynomials over $\mathbb{Q}_p$ for very large p present in the version of `magma` I was using, sometimes the construction of $\mathfrak{D}_4^{\mathrm{alg}}$ failed.

Failures in $\mathfrak{D}_4^{\mathrm{alg}}$ are not hopeless, provided that we accept the conjecture that Ш is finite and therefore $\|Ш[2^\infty]\|$ is a square – for example, if $\|Ш[2]\| = 16$,

the possibilities for Ш[4] indicate that either 0, 3 or 15 of the two-coverings will have descendents, so we can survive failures on any two.

The single curve with rank bound $[2,3]$ revealed its third generator upon attack with `mwrank -b14` (see appendix C; this means that the point on the two-covering had X co-ordinate less than $\exp 14 \approx 1.2 \times 10^6$); however, searches with that great a naïve height bound take about five minutes per two-covering, so are impractical for more than a few hundred curves – moreover, five minutes is longer than the average four-descent takes.

For the smaller K_0, none of the bound-$[1,5]$ and bound-$[2,4]$ curves had four-descendents; similarly for the $[3,5]$ curves for the larger K_0, and all but one of its $[2,4]$ curves. $y^2 = x^3 + 17x + 3140012$, sadly, has too large a conductor for an analytic-rank computation to be reasonable.

Of the 31 $[0,3]$ curves for the smaller K_0, points were found on eleven by a search with `-b12`, and on seven more by four-descent. Five curves found a 4-descendent but no point; the remaining eight crashed `magma`; a deeper search on two of the five found points. Of the 79 $[0,3]$ curves for the larger K_0, 22 gave points after the higher-bound search, 5 found a point after four-descent, 12 found a descendent but no point, and the remaining 40 broke the machinery enough that they found no descendent.

Of the 12 $[0,4]$ curves for the smaller K_0, although there were crashes for some of the two-coverings, enough survived to demonstrate that all the curves had $Ш[2^\infty] = Ш[2]$. Of the sixteen for the larger K_0, seven exhausted time limits on more than two of the quartics, $y^2 = x^3 + 17x + 3139145$ clearly had $(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/4\mathbb{Z})^2 \leq Ш[2^\infty]$, and the other eight had $Ш[2^\infty] = Ш[2] = \mathbb{Z}/2\mathbb{Z}^4$.

Seven of the 372 $[1,3]$ curves at the smaller K_0 had four-descendents (and hence $Ш_{\text{anal}} = 16$), and two crashed magma. Of the 402 $[1,3]$ curves for the larger K_0, 14 had four-descendents and 16 crashed magma.

Of the 53 curves for the smaller K_0 with bound $[1,2]$, points were found on 16 by `mwrank -b12` ; from a four-descent carried out on the remaining 37, 13 caused internal errors, and 24 found a four-descendent and a point on it. Of the 146 such curves for the larger K_0, `mwrank -b12` found points on 42, and a four-descent found points on 37 of the remainder – 8 gave four-descendents and no point, and the remaining 59 crashed magma. Deeper search was applied to the four-descendents lacking a point but with coefficients all < 100, and found points of canonical height 141.4 and 160.5.

Of the 1201 curves for the smaller K_0 with bound $[0,1]$, 159 admitted to a point after attack with `mwrank -b11`. Of the remaining 1042, after 29 hours of calculation, 316 computed a four-descendent and found a point on it, 339 computed a four-descendent and failed to find a point, 58 ran out of time and 329 caused an internal error.

This calculation was performed before the implementation of the Elkies search of section 2.9, and before a bug in the reduction code was fixed; it was repeated afterwards for the 339 curves, with Elkies search up to $|x_i| \leq 2^{18}$ [1] on each of them, finding points on 97 (42 of these because the corrected reduction gave a more reasonable four-covering on which to hunt the point), with canonical heights between 53 and 136. The four-coverings are available should anyone want to perform burn-in tests on a large cluster of computers by performing a much more extensive search for points on them.

Of the 1086 curves for the smaller K_0 with bound $[0, 2]$, 883 demonstrated $Ш[2] = Ш[2^\infty]$, 140 found non-trivial elements of S_4 – of these 140, $Ш_{\text{anal}} \approx 16$ for 130, and $Ш_{\text{anal}} \approx 64$ for the rest. The remaining 63 did not complete.

Of the 1258 curves for the larger K_0 with bound $[0, 2]$, 945 had $Ш[2] = Ш[2^\infty]$, 144 found non-trivial elements of S_4, and 169 did not complete.

Processing 1561 curves for the larger K_0 with bound $[0, 1]$ took 62 hours of CPU time: 576 ran out of time, 492 caused an internal error, 304 successfully computed a four-descendent but found no point, and 189 computed a four-descendent and found a point on it.

The results of this computation – around three hundred P4/2400-hours in total – are summarised in table 3.6.

3.6 How useful has four-descent been in practice?

After the substantial experience with four-descent recorded in this section, we conclude that it is a very useful tool, though by no means perfect. The p-adic factorisation bug in `magma` will be resolved in later versions, which removes the most critical practical problem and will make four-descent a very useful tool for curves with trivial $Ш[2]$ and generators of large height.

However, it is not as useful a tool as I had hoped for curves with non-trivial $Ш[2]$ if ones goal is to find generators, since, for a generator of large height, searches have to be conducted to substantial depth on at least $\|Ш[2]\|$ four-coverings in the knowledge that at most one can yield a usable result.

I was surprised to discover how common it is for $Ш$ to contain elements of order four or more – I had expected them to be rare enough that four-descent would be sufficient for all practical purposes, whilst I have several examples for which even eight-descent would not suffice – and interested to note that large-order elements of $Ш$ appeared significantly rarer as the rank of the curve

[1] this arbitrary bound was picked to require about $4\frac{1}{2}$ minutes per curve, so the 339 curves would take 24 hours

Mordell-Weil rank	**Tate-Shaferevich group**	**Number in $K_0 = 314159$**	**Number in $K_0 = 3141592$**
0	trivial	2255	2279
1	trivial	3713	2878
2	trivial	1557	1355
3	trivial	292	256
4	trivial	24	28
5	trivial	1	2
$\in [0,1]$	unknown	$242 + 385$ [1]	$304 + 1068$
$\in [1,2]$	unknown	$0 + 13$	$5 + 59$
$\in [0,2]$	unknown	$0 + 63$	$0 + 169$
$\in [0,3]$	unknown	$3 + 8$	$12 + 40$
$\in [0,4]$	unknown	0	$0 + 7$
$\in [1,3]$	unknown	$0 + 2$	$0 + 16$
0	$(\mathbb{Z}/2\mathbb{Z})^2$	883	945
0	$(\mathbb{Z}/4\mathbb{Z})^2$	130	N/K
0	$(\mathbb{Z}/8\mathbb{Z})^2$	10	N/K
0	$(\mathbb{Z}/2\mathbb{Z})^4$	12	8
0	$\geq (\mathbb{Z}/2\mathbb{Z})^2 \times (\mathbb{Z}/4\mathbb{Z})^2$	0	1
0	$\geq (\mathbb{Z}/4\mathbb{Z})^2$	0	144
1	$(\mathbb{Z}/2\mathbb{Z})^2$	371	373
1	$(\mathbb{Z}/4\mathbb{Z})^2$	7	N/K
1	$(\mathbb{Z}/2\mathbb{Z})^4$	1	0
1	$\geq (\mathbb{Z}/4\mathbb{Z})^2$	0	14
2	$(\mathbb{Z}/2\mathbb{Z})^2$	29	29
2	$\geq (\mathbb{Z}/4\mathbb{Z})^2$	0	1
3	$(\mathbb{Z}/2\mathbb{Z})^2$	0	3

Table 3.6: Results from 4-descents on curves of the form $y^2 = x^3 + 17x + K_0 + N$, $N \in [-5000, 5000]$. "$x+y$" in a cell means x curves where a four-descendent was found but no point with co-ordinates all < 25000 (except for the cell marked with [1] where the search bound was 2^{18}), and y where the four-descent software failed to complete in two minutes. The distinction between $(\mathbb{Z}/4\mathbb{Z})^2$ and $(\mathbb{Z}/8\mathbb{Z})^2$ is made by examining L(E,1); it is not made for the larger K_0 since the analytic rank computations take too long, hence the $\geq (\mathbb{Z}/4\mathbb{Z})^2$ rows. Mordell-Weil ranks are precisely the number of generators found.

increased.

Class- and unit-group calculations are nearly always very efficient, but there still exist bad examples, and of course these appear as significant obstructions when doing long unattended runs; `magma` takes more than six hours to find the class group of $\mathbb{Q}(\theta)$ where $-11\theta^4-12\theta^3-228\theta^2+1944\theta+924=0$. It may be that these bad examples could be detected more quickly given access to the internals of the `ClassGroup` routine, but at present they show up at random and make it impossible to predict how long a four-descent on a given two-covering might take; the timings can vary enormously even among the different two-coverings of a single elliptic curve.

I am slightly surprised at how great the height of a generator on a rank-one elliptic curve can become: I have several examples where a search up to $N=10^7$ on the single four-descendent of a curve with Selmer rank one failed to find the point, which would suggest that the canonical height of the generator exceeds 150. The Elkies search parallelises perfectly, so a search up to $N=10^9, h\approx 200$ is reasonable over a fortnight on a single fast computer or a weekend on a small cluster; but I doubt one will often want to devote that much computation to a single curve. Lang [48] conjectures, by bounding one by one the terms in the Tate–Shaferevich formula for the regulator, that in the rank-one case

$$|Ш|R = o(\max(a_4^{1/4}, a_6^{1/6})\log N N^{c_1(\log N \log\log N)^{-1/2}}),$$

and if this is true there may be many curves for which the generator P has $h(P)$ large enough that any method taking time $O(\alpha^{h(P)})$ is hopeless.

At rank one, the method of Heegner points comes to the rescue, taking as it does time polynomial in N and $h(P)$ to find a generator; indeed, they have already proven essential for the computation of generators for many of the curves in Cremona's tables [20]. But there is no known analogue of Heegner points at ranks above one.

3.7 An unexpected lack of independence

I noticed that, in computing several hundred analytic Ш values for the above work, I had never observed one not a power of two. This led me to wonder whether $2|\,\|Ш\|$ and $3|\,\|Ш\|$ were in fact independent events, and to be fairly certain that the $P(3|\,\|Ш\|)\approx 0.361$ from [29] is wrong. Looking at the $98,616$ curves of $N<15000$ in [20], we see $4,310$ with $Ш_{\mathrm{anal}}>1$; considering small primes, we find

p	Number of times $p\|Ш_{\text{anal}}$	Proportion
2	3316	0.03363
3	860	0.00872
5	116	0.00118
7	28	0.00028
11	4	0.00004
13	1	0.00001

Fitting a curve suggests that $P(p|Ш_{\text{anal}}) \approx p^{-4}$, and a χ^2 test on the table

	$2 \not| Ш_{\text{anal}}$	$2|Ш_{\text{anal}}$
$3 \not| Ш_{\text{anal}}$	94453	3303
$3|Ш_{\text{anal}}$	847	13

indicates that the events $2|Ш_{\text{anal}}$ and $3|Ш_{\text{anal}}$ are not independent with $p = 0.002$.

Of course, the frequency of $2|Ш_{\text{anal}}$ in the above table is much lower than the 14% or so observed in table 3.5: perhaps this is because the curves of small conductor include a much higher proportion with non-trivial torsion.

Chapter 4

Invariant-theory-based 2-descent over $\mathbb{Z}[i]$

Over $\mathbb{Z}$, the classical 2-descent algorithm, due to Birch and Swinnerton-Dyer and substantially refined over many years by Cremona, relies on bounds for the coefficients of a reduced quartic polynomial with given I and J invariants. In that context, a quartic is reduced iff an associated covariant positive definite quadratic form is reduced, which happens iff a point in the upper half-plane $\mathfrak{H}^2$ associated with the form lies in a specific fundamental region: we pull back the equations defining the region to get the coefficient bounds.

There are approaches, dating back to Hermite and Julia and carried through by Cremona and Stoll [70], which produce a similar form of reduction algorithm for curves with coefficients in $\mathbb{C}$; this work describes the approach from a very computational point of view, and constructs the bounds on the coefficients of a reduced quartic with coefficients in $\mathbb{Z}[i]$ and given I and J invariants, or of a reduced cubic with coefficients in $\mathbb{Z}[i]$ and given discriminant.

4.1 The model – reduction of positive definite quadratic forms

This is very standard material, but I go through it absolutely explicitly since almost every step is followed in the Hermitian case.

We find an $\mathrm{SL}_2(\mathbb{R})$-contravariant map ϕ from the set of positive definite binary quadratic forms to $\mathfrak{H}^2$, and define a positive definite quadratic form as reduced iff its associated point in $\mathfrak{H}^2$ is reduced. To define reduction on $\mathfrak{H}^2$, we find a fundamental region – a connected subset of $\mathfrak{H}^2$ containing precisely one point from each orbit of $\mathrm{SL}_2(\mathbb{Z})$ – and let the reduced form z' of a point

z be the representative point for the orbit $\mathrm{SL}_2(\mathbb{Z}) \cdot z$. Given a quadratic form t, let $z = \phi(t)$, and reduce z by finding M to satisfy $Mz = z'$. Since ϕ is a contravariant map, $M \cdot \phi(t) = \phi(M^{-1}t)$, so we can use $M^{-1}t$ as the reduced form of t.

Specifically, if we write (a, b, c) for the positive definite quadratic form

$$F(x, y) = ax^2 + bxy + cy^2,\ a, b, c \in \mathbb{R}$$

with discriminant $\Delta(F) = b^2 - 4ac$ (since F is positive definite, $\Delta(F) < 0$) we can define $\phi(F) = \frac{\sqrt{\Delta} - b}{2a}$ (the root of $F(x, 1) = 0$ in the upper half-plane): $\Delta < 0$ so $\sqrt{\Delta}$ can be taken with strictly-positive imaginary part.

$\mathrm{SL}_2(\mathbb{R})$ acts on quadratic forms by change of variable:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} A \\ B \\ C \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} a^2A + acB + c^2C \\ 2abA + (ad + bc)B + 2cdC \\ b^2A + bdB + d^2C \end{pmatrix}^{\mathrm{T}}$$

and also on points in $\mathfrak{H}^2$ by Möbius transforms :

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d},$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot x + iy = \frac{ac(x^2 + y^2) + (ad + bc)x + bd \quad + iy}{(cx)^2 + (cy + d)^2}$$

and we have, as desired, the result

Theorem 4.1.1. $\phi(M \cdot F) = M^{-1} \cdot \phi(F) \quad \forall M \in \mathrm{SL}_2(\mathbb{R})$.

whose proof is standard.

Next, we obtain a fundamental region. $\mathrm{SL}_2(\mathbb{Z})$ is generated by $S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The actions of the generators on $\mathfrak{H}^2$ are given by $S(z) = z + 1$ and $T(z) = -z^{-1}$. Let

$$R_0 = \left\{ z \in \mathfrak{H}^2 : \Re z \in \left(-\frac{1}{2}, \frac{1}{2} \right] \right\};$$

clearly a suitable power of S will translate any point of $\mathfrak{H}^2$ into R_0. Considering the action of T then leads us to consider

$$R_1 = \left\{ z \in \mathfrak{H}^2 : \Re z \in (-1/2, 1/2], |z| > 1 \right\}$$

and

$$R_2 = \left\{ z \in \mathfrak{H}^2 : |z| = 1, 0 \leq \Re z \leq 1/2 \right\}$$

$R = R_1 \cup R_2$ will be a fundamental region. We add also the two omitted edge points i and $\frac{1}{2}(1 + \sqrt{-3})$, fixed points of T and of TS respectively.

The shape of the fundamental region leads us immediately to

Lemma 4.1.1. Given any positive definite binary quadratic form F, there is an equivalent form F' such that the real part of $\phi(F')$ lies in $(-1/2, 1/2]$ and the imaginary part is greater than $\sqrt{3/4}$.

Examining the definition of ϕ gives us

Lemma 4.1.2. Given any positive definite binary quadratic form F of determinant $\Delta < 0$, there is an equivalent form $F' = (A', B', C')$ with $|B'| \leq A'$ and $3A'^2 \leq -\Delta$

and as an immediate consequence

Lemma 4.1.3. Given a determinant $\Delta < 0$, there are only finitely many inequivalent positive definite binary quadratic forms with integer coefficients and that determinant

Proof. Any form with that determinant must be equivalent to one with $1 \leq A \leq \sqrt{|\Delta|/3}$ and $B \in [-A, A]$, and there are only finitely many forms with those coefficient bounds. Every (A, B) in that set satisfying $(B^2 + \Delta) = 4AC$ for integral C gives such a form, and they are pairwise inequivalent since they correspond to distinct points in R. □

This result can be very much extended; Birch and Merriman in [4] show that, for any number field K, there are only finitely many equivalence classes under the action of $\mathrm{SL}_2(\mathcal{O}_K)$ of binary forms with coefficients from $\mathcal{O}_K$ and a given degree and discriminant.

4.2 Hermitian forms: definitions, group actions and contravariant maps

Definition 4.2.1. A Hermitian form is a function $\mathbb{C}^2 \to \mathbb{R}$, given by four real parameters $a, \Re b, \Im b, c$: we write

$$F(z_1, z_2) = az_1\overline{z_1} + bz_1\overline{z_2} + \overline{b}\overline{z_1}z_2 + cz_2\overline{z_2}$$

for $a, b, c \in \mathbb{C}$, but you will note that, to guarantee $F(z_1, z_2) \in \mathbb{R}$, a and c must be real. We will abbreviate the above F as $[a, b, c]$; for convenience of notation, we will call the set of Hermitian forms $\mathcal{H}$.

Definition 4.2.2. The discriminant of the Hermitian form $[a, b, c]$ is $\Delta = b\overline{b} - ac$, a real number. A *positive definite Hermitian form* is one for which $F(z_1, z_2) \geq 0$ with equality only at $z_1 = z_2 = 0$; this requires $\Delta < 0$ and $a, c > 0$. Let $\mathcal{H}^+$ be the set of positive definite Hermitian forms.

Definition 4.2.3. Upper half-space, $\mathfrak{H}^3$, is defined by $\mathfrak{H}^3 = \mathbb{C} \times \mathbb{R}^+$; we write a point $p \in \mathfrak{H}^3$ as $p = (z, t)$ where $z \in \mathbb{C}$ and $t \in \mathbb{R}^{>0}$.

It has been known since the work of Hermite in the nineteenth century that $\mathrm{SL}_2(\mathbb{C})$ has an action on $\mathfrak{H}^3$; it can be derived in several ways, but it's easiest to note that this action is precisely what is required to make a natural map constructed later contravariant.

Definition 4.2.4. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{C})$. If we define

$$g \cdot (z, t) = \left(\frac{(az + b)(\overline{cz + d}) + a\overline{c}t^2}{|cz + d|^2 + |ct|^2}, \frac{t}{|cz + d|^2 + |ct|^2} \right)$$

then we have a left action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathfrak{H}^3$: we can check that $M_1 \cdot (M_2 \cdot H) = (M_1 M_2) \cdot H$ for $M_1, M_2 \in \mathrm{SL}_2(\mathbb{C})$, and, since the denominator $|cz + d|^2 + |ct|^2$ is positive, the sign of t' in $(z', t') = g \cdot (z, t)$ will be the same as the sign of t.

We also have a natural change-of-variables action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathcal{H}$ by

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \cdot F(z_1, z_2) = F(\alpha z_1 + \beta z_2, \gamma z_1 + \delta z_2).$$

In our triplet notation for elements of $\mathcal{H}$, this gives us

Lemma 4.2.1. The element

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C})$$

sends the Hermitian form $H = [a, b, c]$ to $H' = [a', b', c']$ with

$$\begin{aligned} a' &= \alpha(a\overline{\alpha} + b\overline{\gamma}) + \gamma(\overline{b\alpha} + c\overline{\gamma}) \\ b' &= \alpha(a\overline{\beta} + b\overline{\delta}) + \gamma(\overline{b\beta} + c\overline{\delta}) \\ c' &= \beta(a\overline{\beta} + b\overline{\delta}) + \delta(\overline{b\beta} + c\overline{\delta}) \end{aligned}$$

with $\operatorname{disc} H' = |\det M|^2 \operatorname{disc} H$. This is a right action of $\mathrm{GL}_2(\mathbb{C})$ on $\mathcal{H}$: $(A \cdot M_1) \cdot M_2 = A \cdot M_1 M_2$. And it sends positive definite Hermitian forms to positive definite Hermitian forms, since $a' = H(\alpha, \gamma)$, $c' = H(\beta, \delta)$ (so they are positive if H was positive definite) and the sign of the discriminant remains unchanged.

Definition 4.2.5. Let ψ be the map $\mathcal{H}^+ \to \mathfrak{H}^3$ given as

$$\psi([a,b,c]) = \left(-\frac{b}{a}, \frac{\sqrt{-\Delta}}{a}\right).$$

Observe that, if $(z,t) = \psi([a,b,c])$, we have $|z|^2 + |t|^2 = \frac{c}{a}$.

We know that this maps into the upper half-plane since $\Delta < 0$ and $a > 0$ for positive definite Hermitian forms. Whilst it discards too much information to be a bijection, it is certainly surjective on the region $|z| > t$ since $\psi([1, -z, z\overline{z} - t^2]) = (z,t)$.

Moreover, we have, and this is the justification for defining the action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathfrak{H}^3$ above, that

$$\psi(H \cdot M) = \overline{M^{-1}} \cdot \psi(M);$$

the verification is straightforward by working symbolically in

$$\mathbb{C}(\alpha, \beta, \gamma, \delta, \sqrt{-\Delta}, a, b, c, \overline{\alpha}, \overline{\beta}, \overline{\gamma}, \overline{\delta}, \overline{b})$$

and using the identities $\det M = \det \overline{M} = 1$ and $\sqrt{-\Delta}^2 = ac - b\overline{b}$. Since the latter identity is required, the result does not hold for H not positive definite; but since it is obtained symbolically it holds for all $M \in \mathrm{SL}_2(\mathbb{C})$.

4.3 Bounds on reduced Hermitian positive definite quadratic forms

We next need fundamental regions for $\mathrm{SL}_2(\mathcal{O}_K)$, which I give for $\mathbb{Z}[i]$ and $\mathbb{Z}[\sqrt{-2}]$. A paper of Swan [72] gives generators and fundamental regions for a number of imaginary quadratic number fields, including some with class number not 1. Let

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \; S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

be the generators for $\mathrm{SL}_2(\mathbb{Z})$; S has order 4 and TS has order 6.

Lemma 4.3.1. $\mathrm{SL}_2(\mathbb{Z}[i])$ is generated by S, T, and also

$$T_i = \begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}, \quad K = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

K, KT, KS, KT_i, T_iK and TK all have order 4.

Lemma 4.3.2. $\mathrm{SL}_2(\mathbb{Z}[\sqrt{-2}])$ is generated by S, T and $T_2 = \begin{pmatrix} 1 & \sqrt{-2} \\ 0 & 1 \end{pmatrix}$.

The following table gives the actions of the five generators mentioned above on $\mathfrak{H}^3$ and on $\mathcal{H}^+$: they are all self-explanatory except for the operation of S on $\mathfrak{H}^3$, which corresponds to inversion in a hemisphere of radius 1 centred at the origin.

Generator	Action on $[a,b,c] \in \mathcal{H}^+$	Action on $(z,t) \in \mathfrak{H}^3$
T	$[a, a+b, a+b+\overline{b}+c]$	$(z+1, t)$
T_i	$\left[a, b-ia, a+i(b-\overline{b})+c\right]$	$(z+i, t)$
T_2	$\left[a, b-\sqrt{-2}a, a+\sqrt{-2}(b-\overline{b})+c\right]$	$(z+\sqrt{-2}, t)$
K	$[a, -b, c]$	$(-z, t)$
S	$[c, -b, a]$	$(-\frac{\overline{z}}{\lvert z\rvert^2+\lvert t\rvert^2}, \frac{t}{\lvert z\rvert^2+\lvert t\rvert^2})$

For $\mathrm{SL}_2(\mathbb{Z}[i])$, it is clear that actions by suitable powers of T and T_i can send any point in $\mathfrak{H}^3$ to lie in the prism

$$P_1 = \left\{(z,t) \in \mathfrak{H}^3 : -\frac{1}{2} \leq \Im z < \frac{1}{2}, -\frac{1}{2} \leq \Re z < \frac{1}{2}\right\}.$$

The action of K rotates this prism 180° around the axis passing through $z = 0$, so we can replace P_2 with the half-prism

$$P_2 = \left\{(z,t) \in \mathfrak{H}^3 : -\frac{1}{2} \leq \Im z < 0, -\frac{1}{2} \leq \Re z < \frac{1}{2}\right\}.$$

Following the model of section 4.1, we now consider the action of S. This interchanges the regions $R_1 : |z|^2 + |t|^2 < 1$ and $R_2 : |z|^2 + |t|^2 > 1$; the interior of our fundamental domain will be $P_2 \cap R_2$.

We also need to get the boundaries of the fundamental region correct; at the moment we have that the plane $\Im z = -\frac{1}{2}$ is fixed by $T_i^{-1}K$ (which acts on this plane by negating the real part), and the plane $\Re z = -\frac{1}{2}$ is fixed by $T^{-1}K$, whilst the line $z = 0$ is fixed by K. Decompose the plane $\Im z = -\frac{1}{2}$ into $R_7 \cup R_8$, where

$$R_7 = \left\{(z,t) \in \mathfrak{H}^3 : \Im z = -\frac{1}{2}, \Re z < 0\right\}$$

and R_8 is its complement. However, we want the line $L : z = -\frac{1}{2} - \frac{1}{2}i$ to lie in the fundamental domain even though it lies in R_7 which we will be excluding.

On the region $R_3 : |z|^2 + |t|^2 = 1$, S acts by $(z,t) \to (-z,t)$; we decompose

R_3 as $R_4 \cup R_{5\{a,b,c\}} \cup R_6$, where

$$\begin{aligned}
R_4 &= \left\{(z,t) \in \mathfrak{H}^3 : \Re z < 0, |z|^2 + |t|^2 = 1\right\} \\
R_{5a} &= \left\{(z,t) \in \mathfrak{H}^3 : \Re z = 0, \Im z > 0, |z|^2 + |t|^2 = 1\right\} \\
R_{5b} &= \left\{(z,t) \in \mathfrak{H}^3 : \Re z = 0, \Im z < 0, |z|^2 + |t|^2 = 1\right\} \\
R_{5c} &= \{(0,1)\} \\
R_6 &= \left\{(z,t) \in \mathfrak{H}^3 : \Re z > 0, |z|^2 + |t|^2 = 1\right\}
\end{aligned}$$

S interchanges R_4 and R_6 and R_{5a} and R_{5b}, whilst leaving R_{5c} (the single point $(0,1)$ corresponding to multiples of the form $|x|^2 + |y|^2$) fixed; SK acts as the identity on the whole of R_3. So we define

$$\mathfrak{F}_i = (R_{5c} \cup R_{5b} \cup (P_2 \cap (R_2 \cup R_6)) \setminus R_7) \cup L.$$

For $\mathbb{Z}[\sqrt{-2}]$, the calculation is similar though less convoluted; define the prism

$$P_{\sqrt{-2}} : \left\{(z,t) \in \mathfrak{H}^3 : -\frac{1}{2} \leq \Re z < \frac{1}{2}, -\frac{\sqrt{2}}{2} \leq \Im z < \frac{\sqrt{2}}{2}\right\}$$

into which any point of $\mathfrak{H}^3$ may be moved by a unique element of $T' = \langle T, T_2 \rangle$, and once again the interior of the fundamental domain is $P_{\sqrt{-2}} \cap R_2$. We do not have the confounding element K, so our contribution from the region R_3 can be $R_6 \cup R_{5b} \cup R_{5c}$ as before but without having to intersect with P_2:

$$\mathfrak{F}_{\sqrt{-2}} = (P_{\sqrt{-2}} \cap R_2) \cup R_6 \cup R_{5b} \cup R_{5c}.$$

The perils of finite precision

Our fundamental regions are hyperbolic polyhedra with some faces open and some closed; we have some elements which act trivially on codimension-1 or greater subspaces of $\mathfrak{H}^3$. There will always be trouble in checking whether an element of $\mathfrak{H}^3$ given with finite precision lands in one of these infinitely-thin spaces, and, whilst we do not make actually-transcendental extensions at any point in this chapter, it is as yet impractical to work in $\overline{\mathbb{Q}}$: we have to use finite precision.

The problem crops up most obviously when we work with Hermitian forms with rational coefficients, and in that case we try to work in $\mathcal{H}^+$ as much as we can, since the only precision loss comes from calculating the square root in the ψ map of definition 4.2.5. This means we have to translate the equations describing our fundamental domain in $\mathfrak{H}^3$ into restrictions in $\mathcal{H}^+$.

Let $F = [A, B_r + iB_i, C]$ be a Hermitian form with coefficients in $\mathbb{Q}[i]$.

$\Re z = -\frac{1}{2}$ means $A = -2B_r$, $\Im z = -\frac{1}{2}$ means $A = -2B_i$. The region P_1 has $(B_r, B_i) \in ([-A/2, A/2))^2$, R_3 has $A = C$, R_1 has $A > C$ and R_2 has $A < C$. R_4 and R_6 are $B_r < 0$ and $B_r > 0$ respectively, whilst the fine separation of R_5 into R_{5a}, R_{5b}, R_{5c} comes from $B_r = 0$ and $B_i > 0, B_i < 0, B_i = 0$.

4.3.1 Bounds on A, B, C for $\psi([A, B, C]) \in \mathfrak{F}$

Throughout P_2, we have $|z|^2 \leq \frac{1}{2}$, and throughout $R_2 \cup R_3$ we have $|z|^2 + |t|^2 \geq 1$; so $|t|^2 \geq \frac{1}{2}$ in $\mathfrak{F}_i$; for $\mathfrak{F}_{\sqrt{2}}$, we have $|z|^2 \leq \frac{3}{4}$ in $P_{\sqrt{2}}$, and so $|t|^2 \geq \frac{1}{4}$. Call this lower bound $t_{\max}$ in what follows.

By definition, $|t|^2 = \frac{-\Delta}{A^2}$; so $|t|^2 \geq t_{\max} \implies A^2 \leq -t_{\max}^{-1}\Delta$. A is positive and, since $\Delta < 0$ and $B\overline{B} > 0$, we cannot have $A = 0$, so $1 \leq A \leq \sqrt{-t_{\max}^{-1}\Delta}$.

But, if we know A and we know that $\psi([A, B, C]) \in P_2$, we have $-\frac{A}{2} \leq \Re B < \frac{A}{2}$ and $-\frac{A}{2} \leq \Im B < 0$, or, if we know $\psi([A, B, C]) \in P_{\sqrt{-2}}$, $-\frac{A}{2} \leq \sqrt{\frac{1}{2}}\Im B < \frac{A}{2}$– so we have a finite number of choices for A and B, and, just as in section 4.1, we have shown that there are only finitely many $\mathrm{SL}_2(\mathbb{Z}[i])$- or $\mathrm{SL}_2(\mathbb{Z}[\sqrt{-2}])$-inequivalent positive definite Hermitian quadratic forms with a given discriminant and coefficients from any discrete subset of $\mathbb{C}$ desired.

4.3.2 Computing the reduced form of an H.p.d.q.f.

Write $a \approx b$ if $|a - b| < 10^{-20}$; with the standard precisions in `magma`, this is a reasonable bound to give for floating-point precision. We use three-valued indicator variables, for "inside", "outside" and "on boundary".

Consider a point $P = (z, t) \in \mathfrak{H}^3$, and define $S(P) = |z|^2 + |t|^2$.

- If $S(P) \approx 1$, set $P_s = 2$; otherwise, if $S(P) < 1$ set $P_s = 0$ and otherwise set $P_s = 1$.
- If $\Re z \approx -0.5$, set $P_r = 2$; otherwise, if $\Re z < -0.5$ or $\Re z \geq 0.5$, set $P_r = 0$, and otherwise $P_r = 1$.
- If $\Im z \approx 0$ or $\Im z \approx -0.5$, set $P_i = 2$; otherwise, set $P_i = 1$ unless $\Im z > 0$ or $\Im z < -0.5$.
- If any of P_i, P_r or P_s is zero, P is outside $\mathfrak{F}_i$.
- If $P_s = 2, P_r \neq 2$ and $\Re z < 0$, we are in region R_4, so P is outside $\mathfrak{F}_i$.
- If $\Re z \approx 0.5$, we are outside P_1, so P is outside $\mathfrak{F}_i$.
- If $P_i = 2, P_r \neq 2$ and $\Re z < 0$, we are in the region R_7, so P is outside $\mathfrak{F}_i$.
- Otherwise, P lies in $\mathfrak{F}_i$.

To move a point into $\mathfrak{F}_i$, follow the following procedure (this is pseudo-code, so you update z at each stage and use the value from the last stage rather than the value at the beginning of the loop) until the algorithm above tells you the point is in $\mathfrak{F}_i$:

- Let $P = \lfloor \Re z + \frac{1}{2} \rfloor + i \lfloor \Im z + \frac{1}{2} \rfloor$, and, if P is non-zero, subtract P from z; this is an action by $T^{-\Re P} T_i^{-\Im P}$.
- If $\Im z > 0$ then negate z by acting by K.
- If $\Im z \approx 0$ and $\Re z < 0$ then we are in the region R_{5a}; replace z by $-\Re z$ (this sets $\Im z$ to exactly zero, which helps maintain precision), and record an action by K.
- If $\Im z \approx -0.5$ and $\Re z < 0$ then we are in the region R_7; replace z by $-\overline{z}$ to negate the real part, and record an action by $T_i^{-1} K$.
- If $\Re z \approx 0.5$ then set $z = -0.5 + i\Im z$ and record an action by T^{-1}.
- If $\Re z \approx -0.5$ then set $z = -0.5 + i\Im z$; this simply deals with precision loss
- If $S((z,t)) < 1$, or if $S((z,t)) \approx 1, \Re z < 0$ and $z \neq -0.5$ – that is, if we are in $R_2 \cup R_4$ – act on Z by S.

It is clear that, if this algorithm terminates, it will do so with $[A, B, C]$ corresponding to a point in the fundamental region. Since we perform a meaningful action by T only when $|z|^2 + |t|^2 < 1$ – on R_4 the action of T is simply to negate the real part – and, in that case, the action of T increases t, we have a quantity growing larger at every step through the loop.

In fact, if we worked in $\mathcal{H}^+$ rather than $\mathfrak{H}^3$, the procedure above performs the Euclidean algorithm on A and C: the action of S^{-1}, S_i^{-1} replace C by $C - A, C - iA$ respectively, and the action of T swaps A and C. Since the Euclidean algorithm terminates after $O(\log(\max|A|, |C|))$ division steps, the reduction procedure above terminates after $O(\log(\max|A|, |C|))$ applications of T. Since only finitely many other transformations occur between each application of T, the procedure is finite.

4.4 $\mathrm{SL}_2(\mathcal{O}_K)$-reducing higher-degree forms with complex coefficients

Let

$$f(x) = \sum_{i=0}^{n} a_i x^i = a_n \prod_{i=1}^{n} (x - \alpha_i)$$

be a polynomial with coefficients a_i and roots α_i lying in $\mathbb{C}$. Cremona's paper [23] handles the case where $a_i \in \mathbb{R}$ and $n = 3, 4$; his joint paper [70] with Stoll proves, amongst other things,

Theorem 4.4.1. Let $\mathbb{C}[X, Z]_n$ denote the set of binary degree-n forms with coefficients from $\mathbb{C}$. With any form $Q(X, Z)$ we can associate a Hermitian form

$$f_Q(X, Z) = \sum_{j=1}^{n} \frac{|X - \alpha_j Z|^2}{|Q'(\alpha_j)|^e}$$

where the exponent is given as $e = 2/(n-2)$, and hence a point $\varphi(Q) = \psi(f_Q) \in \mathfrak{H}^3$ where ψ is the map from Hermitian forms to $\mathfrak{H}^3$ given in definition 4.2.5. For $n = 3, 4$, this map $Q \to \varphi(Q)$ is the **unique** covariant map $\mathbb{C}[X, Z]_n \ \to \ \mathfrak{H}^3$.

Given this covariant map, we define a polynomial as reduced over an imaginary quadratic number field K iff its associated point in $\mathfrak{H}^3$ lies in $\mathfrak{F}(\mathrm{SL}_2(\mathcal{O}_K))$, and obtain a reduction $f_1 = M^{-1} \cdot f$ where $M \cdot \psi(f) \in \mathfrak{F}_{\mathcal{O}_K}$.

4.4.1 Bounding coefficients for reduced polynomials

The aim of this section is to show

Theorem 4.4.2. For a polynomial Q of degree n, with leading coefficient a, constant term z and discriminant Δ, the $X\overline{X}$ and $Y\overline{Y}$ coefficients of the Stoll covariant f_Q are bounded in terms of $|a|$, $|z|$ and $|\Delta|$.

Proof. Let the roots of Q be $\alpha_1 \dots \alpha_n$, and consider the Hermitian form f_Q defined in theorem 4.4.1; we can write

$$f_Q = r_0 X\overline{X} + r_1 X\overline{Y} + \overline{r_1 X}Y + r_2 Y\overline{Y}$$

with

$$r_0 = \sum_{i=1}^{n} |Q'(\alpha_i)|^{-2/n-2}, r_2 = \sum_{i=1}^{n} |\alpha_i|^2 |Q'(\alpha_i)|^{-2/n-2}.$$

Let e be the exponent $\frac{-2}{n-2}$, and $f = 2n-2$ be the power of the leading coefficient which appears when you expand the discriminant of Q as $\Delta = a^f \Psi$ with $\Psi = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j)^2$.

Expanding by the product rule, $Q'(\alpha_i) = a \prod_{1 \le j \le n, j \ne i} \alpha_i - \alpha_j$, so r_0 is a sum of n positive terms with product

$$P = \prod_{i=1}^{n} \left(|a| \prod_{1 \le j \le n, j \ne i} |\alpha_i - \alpha_j| \right)^e .$$

A given term $|\alpha_k - \alpha_l|$ will appear precisely $n-2$ times in this product, being omitted only in $Q'(\alpha_k)$ and $Q'(\alpha_l)$; the aggregate power of $|a|$ is just ne, and so we get $P = |a|^{ne}|\Psi|^{-1}$. Obviously, we compare with $|\Delta|^{-1} = |a|^{-f}\Psi^{-1}$ to get $P = |a|^{ne+f}|\Delta|^{-1}$.

Now, applying the arithmetic-geometric mean inequality

$$\sum_{i=1}^{n} x_i \geq n \left(\prod_{i=1}^{n} x_i \right)^{1/n}$$

we get $r_0 \geq n|\Delta|^{-1/n}|a|^{f/n+e}$. For the r_2 term, the product is increased by a factor $|\prod \alpha_i|^2$, which by the standard symmetric-polynomials argument is just $|\frac{z}{a}|^2$, and so we have $r_2 \geq n|\Delta|^{-1/n}|a|^{e+f/n}|\frac{z}{a}|^{2/n}$.

□

In the cubic case, $n = 3, e = -2, f = 4$, and

$$r_0 \geq 3|\Delta|^{-1/3}|a|^{-2/3};$$

for quartics, $n = 4, e = -1, f = 6$, and

$$r_0 \geq 4|\Delta|^{-1/4}|a|^{1/2}.$$

4.5 Explicit bounds on the leading coefficient, in the cubic and quartic cases

4.5.1 Cubics

From the arguments of section 4.4.1, we have that, for a cubic $Q(x,y) = ax^3 + bx^2y + cxy^2 + dy^3$ with roots $\alpha_1, \alpha_2, \alpha_3$, the modified Stoll covariant $|\Delta| f_Q = [r_0, r_1, r_2]$ has $r_0 \geq 3|a|^{2/3}|\Delta|^{1/3}$ and $r_2 \geq 3|d|^{2/3}|\Delta|^{1/3}$. [1]

This result gives us a bound on the leading coefficient of a reduced cubic. For, if $|\Delta| f_Q(x,y)$ is reduced, its covariant point in $\mathfrak{H}^3$ will have t co-ordinate not less than some (positive) constant $t_{\min}$ dependent only on the field we are reducing over. From section 4.2.5, this t co-ordinate is given as $\sqrt{-\Delta_f}/r_0$, where $\Delta_f = r_0 r_2 - |r_1|^2$ is the discriminant of $|\Delta| f_Q$. Multiplying out, we discover that $-\Delta_f = 3|\Delta|$, so we have $\sqrt{3|\Delta|} \geq t_{\min} r_0$.

Substituting in the bound for r_0 above, we find

$$3^{1/2}|\Delta|^{1/2} \geq 3t_{\min}|a|^{2/3}|\Delta|^{1/3}.$$

[1] A similar lower bound for r_1 is not available; if $\alpha_2 = \omega\alpha_1$ and $\alpha_3 = \omega^2\alpha_1$ for $\omega = \exp(2\pi i/3)$ (ie the roots are arranged in an equilateral triangle), then $r_1 = 0$. But we know r_1 once we know r_0, r_2, Δ.

Dividing both sides by $3t_{\min}|\Delta|^{1/3}$, we find

$$|a|^{2/3} \leq 3^{-1/2} t_{\min}^{-1} |\Delta|^{1/6};$$

both sides are positive, so we can raise to the power $3/2$ to obtain

$$|a| \leq 3^{-3/4} t_{\min}^{-3/2} |\Delta|^{1/4}.$$

For the $\mathbb{Z}[i]$ case, where $t_{\min} = 2^{-1/2}$, this bound is $(2/3)^{3/4}$; if we were reducing **real** cubics over $\mathrm{SL}_2(\mathbb{Z})$, we would have $t_{\min} = 4/3^{-1/2}$ since the fundamental region in $\mathfrak{H}^2$ has a higher lowest point, and would get $|a| \leq \frac{2\sqrt{2}}{3\sqrt{3}}$, the coefficient obtained on p.72 of [23].

Now, we use the uniqueness of the covariant for cubics. By the classical covariant theory from [23], any cubic Q has a cubic covariant $g_3(Q)$, with leading coefficient $U = 2b^3 + 27a^2d - 9abc$ and discriminant $729\Delta^3$. Now, $g_3(Q)$ will have a Stoll covariant point $\psi(g_3(Q))$. But this will be a covariant point of Q (since covariants of covariants are covariant), and therefore will just be $\psi(Q)$. So, if Q is reduced, $\psi(Q)$ must be reduced, and by the argument above we have $|U| \leq 3^{3/4} t_{\min}^{-3/2} |\Delta|^{3/4}$ (where the $3^{3/4}$ arises as $3^{-3/4} \times 729^{1/4}$).

If we know a, U, Δ, we essentially know all about the cubic, because we can get the other seminvariant $P = b^2 - 3ac$ from the syzygy $4P^3 = U^2 + 27\Delta a^2$; we have Δ fixed, we can search over the regions found above for U and a, and check whether the syzygy's result is four times a cube of an element of the search space – there is clearly potential for a two-dimensional sieve approach at this stage.

Once we have a value of P from the syzygy, we can pick $b = 0$, solve (trivially) $P = b^2 - 3ac$ and $U = 2b^3 + 27a^2d - 9abc$ to get c and d (rejecting values which do not fall in the right regions of search space), and reduce the resultant cubic, knowing that picking b corresponded only to a translation of the cubic.

Hence we have an algorithm for finding all equivalence classes of cubics with coefficients from any discrete subset of $\mathbb{C}$ and a given discriminant, the equivalence classes being defined under the action of $\mathrm{SL}_2(\mathbb{Z}[i])$ or $\mathrm{SL}_2(\mathbb{Z}[\sqrt{-2}])$.

4.5.2 Quartics

The arguments here are very similar, though complicated by the fact that the discriminant of f_Q is not uniquely determined by knowing $|\Delta|$ (which we might have expected, since quartics have two invariants).

However, we get from [70] that

$$\Delta_f = \sum_{1 \le j < k \le 4} \|\alpha_j - \alpha_k\|^2 t_j t_k$$

with $t_j = \|Q'(\alpha_j)\|^{-1}$ (the version in [70] has an additional constant factor 4 because they define r_1 differently). This is a sum of six positive terms, and, proceeding in the same way as in section 4.4.1, we find that their product is $|a|^{-12}|\Psi|^{-2}$ (each term contributes $|a|^{-2}$ from the two t_k components, the $|\alpha_j - \alpha_k|^2$ gives one $|\Psi|$, and in the $\prod_{1 \le j < k \le 4} t_j t_k$ we have each $|\alpha_i - \alpha_j|^{-1}$ appearing six times for a total $|\Psi|^{-3}$).

Fortuitously, $|a|^{-12}|\Psi|^{-2} = \Delta^{-2}$; we apply the AGM to get $|\Delta|_f \ge 6|\Delta|^{-1/3}$, and the arguments of section 4.4.1 then give $r_0 \ge 4|\Delta|^{-1/4}|a|^{1/2}$.

Applying $\sqrt{\Delta}_f \ge t_{\min} r_0$, we then have $|a| \le 3/8 t_{\min}^{-2} |\Delta|^{1/6}$. In the $\mathbb{Z}[i]$ case, the constant here is $\frac{3}{4}$.

And this is in fact enough; quartics have a quartic covariant g_4 with leading coefficient $-H = 3b^2 - 8ac$ and discriminant $4096 J^2 \Delta$ (where J is the J-invariant of the original quartic). By the same uniqueness argument as in the cubic case, g_4 is reduced iff the original quartic is, so we get a bound on H as $|H| \le 3/2 |J|^{1/3} t_{\min}^{-2} |\Delta|^{1/6}$.

And so we have finite search regions for a and H, whilst we know I and J by hypothesis. Again, we have a seminvariant syzygy $H^3 - 48Ia^2H + 64Ja^3 = -27R^2$, so we look over a and H values to find things which are -27 times a square; at this point we know R. We can take $b = 0$, solve for c from a and H, note that $R = b^3 + 8a^2 d - 4abc$ and solve for d, and find e from $I = 12ae - 3bd + c^2$.

4.6 Applications to descent

As in the rational case, this reduction theory is essential in theory for the two-descent based on invariant theory, and useful in practice for the two-descent based on algebraic number theory: Simon [66] can now use the latter technique to perform two-descents over number fields fairly efficiently, and obtain a complete set of two-coverings representing all the elements of the 2-Selmer group. However, there is no reason for these two-coverings to have pleasant coefficients, and an application of reduction can make the coefficients much more manageable – after all, we know the invariants of the associated quartics, and have absolute bounds for the sizes of the coefficients of a quartic with given invariants. However, I have not found an example where searching for rational points on a two-covering is the limiting factor for Simon's two-descent.

Using the very precise enumeration in Serf's thesis [58] of the possibilities for

the I and J invariants of a minimal quartic over a quadratic number field, we can find up to four (I, J) pairs which together cover all the reduced minimal quartics that could appear in a two-covering for a given elliptic curve. The technique of section 4.5.2 will find all the $\mathbb{Z}[i]$- or $\mathbb{Z}[\sqrt{-2}]$-reduced minimal quartics with each of the invariants; chapter 2.2 of [58] contains an algorithm for checking local solvability of the associated two-coverings, and a clever argument of Cremona in [22] allows you to check equivalence between pairs of two-coverings by looking at the factorisation of various polynomials over the number field.

Combining all of these, we end up with a single representative for each non-trivial element of $S_2(E(K))$. We finish in the same way as the direct 2-descent over $\mathbb{Q}$, by searching for points on each of these representatives.

Searching for points on a two-covering over a number field is not a well-studied task at the moment; the best techniques I am aware of use quadratic sieves at primes with degree-1 residue fields to accelerate the evaluation and checking for squareness of

$$f(z^{-1}\sum x_i\theta^i)$$

for a range of integral x_i and z, and θ a basis of $\mathcal{O}_K$. It may turn out worthwhile instead to write the $[K:\mathbb{Q}]$ simultaneous equations in $2\,[K:\mathbb{Q}]+2$ unknowns with integral coefficients, corresponding to

$$z_1^{-1}\left(\sum y_i\theta^i\right)^2 = z_2^{-1} f\left(\sum x_i\theta^i\right),$$

and use a variation on [33].

Chapter 5

Directions for further work

The implementations of four-descent and L-function derivatives described in this thesis will be incorporated into a release of `magma` to appear sometime after September 2003, along with a routine for finding generators for curves with analytic rank 1 and small conductor using Heegner points; that routine was almost entirely developed by Cremona, and accordingly is not discussed here.

The construction of $\mathfrak{D}_4^{\mathrm{alg}}$ extends to number fields immediately, and an implementation, using the same sorts of relative-extension work as used in Simon [66], would not be too difficult. The minimisation will be fairly similar, at least at degree-one unramified primes: Serf's extension in chapter five of [58] of the minimisation of quartic forms over quadratic fields is the model to follow, and the method based on $\mathbb{F}_p[x, y]$ appears to extend immediately. It is not, however, clear how to extend reduction, or how to work with primes where the residue field is not $\mathbb{F}_p$.

Naturally, any improvement in the computation of class and unit groups for general number fields will make the computation of $\mathfrak{D}_4^{\mathrm{alg}}$ easier; however, unless you are checking that $\|Ш[4]\| = \|Ш[2]\|$ for long lists of curves, and so almost never generating explicit elements, the lion's share of the calculation time is taken by Stoll reductions; this is because those are calculated by default at a very high precision since the current implementation is not robust against precision loss. This effect is worst when the generators of the unit group have very large coefficients with respect to an integral basis, since these will lead to similarly-complicated elements of $L'(S, 2)$ which require significant reduction effect to come down to single-digit coefficients.

The present implementation feels very inefficient in the fairly common case where $Ш[2]$ is non-empty, and we have found g generators for the Mordell-Weil group and believe there is one missing. The set of non-trivial elements of the 2-Selmer group is of size $2^{g+s+1} - 1$; taking a quotient by the part generated

by the known generators gives $2^{s+1}-1$ individual two-coverings for the four-descent. In the best case, where $Ш[4] = Ш[2]$, we only find four-descendents on one of them – but we find 2^{g+s+1} four-descendents, and, while only one of these can have an interesting rational point, we have to check all of them to find it. It may be that only a further descent could resolve this issue.

It would be interesting to find a non-experimental way of determining the resolvability of a two-covering; this would lead to a novel definition of "reduced" for two-coverings, namely replacement with an equivalent two-covering with smallest possible $\left\|\mathfrak{D}_4^{\mathrm{alg}}\right\|$, which is more relevant for the four-descent than the classical reduction of binary quartic forms.

It is possible to search for $\mathbb{Q}_p$-points on the two-coverings that generate the local image more efficiently than I do at the moment.

Whilst it is not clear how to extend the `ecsieve` algorithm to handle curves $y^2 + xy + y = f(x)$, it is trivially possible to use it for curves in obviously non-minimal models – arbitrary values of a_2, for instance. I did not do this in the work to date because it was not obvious to me what range of a_2 values it would be sensible to use, other than the $\{-1, 0, 1\}$ which provide possibly-minimal models.

The techniques of chapter four were carried out over real quadratic number fields by Serf in [58]. For more general number fields K, the natural space in which to find covariant points is $(\mathfrak{H}^2)^r \otimes (\mathfrak{H}^3)^s$ where r and s are the number of real and complex embeddings of K, and it is not clear what the correct analogue of the fundamental region would be: for real quadratic fields, the bound obtained was on the product of the imaginary parts of the embeddings.

What is clear is that even four-descent is not a complete solution to the hunt for generators for elliptic curves: whilst it opens up the previously-inaccessible range of heights between about 70 and about 150, it would appear, from the existence of curves with small a_4 and a_6, Selmer rank one, and a single four-descendent, with small coefficients, on which there is no point with coordinates $< 10^7$, that a reasonable fraction of curves have generators of even greater heights. I suspect that deeper descent will in practice provide a better attack on these curves than the method of Heegner points offers, even though the asymptotics would seem to be in the latter method's favour; I do not know how third and subsequent two-descents will be performed, but I am sure that some day I will read a paper performing them.

Appendix A

Invariants and covariants of a binary quartic form

A binary quartic form $f(x,y) = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$ has (see for example [42]) the two invariants

$$I = 12ae - 3bd + c^2$$

and

$$J = 72ace + 9bcd - 27ad^2 - 27eb^2 - 2c^3.$$

An argument involving generating functions shows that any other invariant will be an isobaric polynomial in I and J; in particular, the discriminant of the quartic is $\Delta = \frac{1}{27}\left(4I^3 - J^2\right)$.

As well as these two degree-zero invariants, we have degree-four and degree-six algebraic covariants

$$g_4 = (3b^2-8ac)x^4+4(bc-6ad)x^3+2(2c^2-24ae-3bd)x^2+4(cd-6be)x+(3d^2-8ce)$$

$$\begin{aligned} g_6 = {} & (b^3 + 8a^2d - 4abc)x^6 + 2(16a^2e + 2abd - 4ac^2 + b^2c)x^5 \\ & + 5(8abe + b^2d - 4acd)x^4 + 20(b^2e - ad^2)x^3 - 5(8ade + bd^24bce)x^2 \\ & - 2(16ae^2 + 2bde - 4c^2e + cd^2)x - (d^3 + 8be^2 - 4cde) \end{aligned}$$

Writing $Q(x) = f(x,1)$, we find that the covariants are related by the syzygy

$$27g_6^2 = g_4^3 - 48IQ^2g_4 - 64JQ^3;$$

rearranging, we find that, if (x, y) satisfies $y^2 = Q(x)$, the point

$$(t, u) = \left(\frac{3g_4(x)}{4y^2}, \frac{27g_6(x)}{8y^3} \right)$$

lies on the elliptic curve $u^2 = t^3 - 27It - 27J$. This is the two-covering map $\xi : \mathcal{C} \to E$ used in section 2.2.3.

The leading coefficients of the g_4 and g_6 covariants are called the H and R seminvariants respectively; there is also a Q seminvariant equal to $\frac{1}{3}(H^2 - 16a^2I)$. These seminvariants are relevant for finding the number of real roots of a quartic polynomial, and for constructing the covariant binary quadratic form used for reducing quartic polynomials (and hence two-coverings) over $\mathrm{SL}_2(\mathbb{Z})$.

A.1 The map $\theta_P : \mathcal{C} \to E$

Let $\mathcal{C}$ be a curve $y^2 = f(x)$ with f quartic, and let $P = (X, Y)$ lie on $\mathcal{C}$. Translate to make the X co-ordinate of P zero, giving a curve

$$\mathcal{C}' : y^2 = ax^4 + bx^3 + cx^2 + dx + q^2.$$

Then, following [18], the map

$$\begin{aligned} u &= x^{-2}(2q(y + q) + dx) \\ v &= x^{-3}(4q^2(y + q) + 2q(dx + cx^2) - d^2x^2/2q) \end{aligned}$$

takes $(x, y) \in \mathcal{C}$ to (u, v) on the elliptic curve

$$v^2 + \frac{d}{q}uv + 2bqv = u^3 + \left(c - \frac{d^2}{4q^2} \right) u^2 - 4q^2u + a(d^2 - 4q^2c).$$

Appendix B

The invariant map up from a four-covering

This appendix presents material from [51].

Let $\mathcal{H} = (M_1, M_2)$ be a four-covering expressed as a pair of 4×4 matrices; let $\mathbf{x} = (x_1 : x_2 : x_3 : x_4)$ be the vector of homogeneous co-ordinates, and $\mathbf{p}$ be a point with $\mathbf{p}M_i\mathbf{p}^{\mathrm{T}} = 0$. Let f_1 and f_2 be the quaternary quadratic functions $f_i(\mathbf{x}) = \mathbf{x}M_i\mathbf{x}^{\mathrm{T}}$.

Consider the quartic polynomial $Q = \det(M_1 - \lambda M_2)$ (which defines the sigma invariants by $\det(M_1 - \lambda M_2) = \sum_{i=1}^{5} \sigma_i \lambda^{i-1}$), and the two adjoint matrices $r_i = \operatorname{adj} M_i$.

Consider also the double-adjoint matrix $R = \operatorname{adj}(xr_1 + r_2)$, which is 4×4 with its entries cubic quadratic polynomials; decompose it as

$$R = R_3x^3 + R_2x_2 + R_1x + R_0$$

with $R_i \in \mathrm{GL}_4(\mathbb{Z})$, and compute the quadratic functions $d_1(\mathbf{x}) = \sigma_1^{-1}\mathbf{x}R_1\mathbf{x}^{\mathrm{T}}$ and $d_2(\mathbf{x}) = \sigma_5^{-1}\mathbf{x}R_2\mathbf{x}^{\mathrm{T}}$.

The fifth invariant of the pair of quadrics is G, a quartic function of the x_i equal to one sixteenth of the determinant of the 4×4 Jacobian matrix

$$\begin{pmatrix} \frac{\partial d_1}{\partial x_1} & \frac{\partial d_1}{\partial x_2} & \frac{\partial d_1}{\partial x_3} & \frac{\partial d_1}{\partial x_4} \\ \frac{\partial d_2}{\partial x_1} & \frac{\partial d_2}{\partial x_2} & \frac{\partial d_2}{\partial x_3} & \frac{\partial d_2}{\partial x_4} \\ \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \frac{\partial f_1}{\partial x_3} & \frac{\partial f_1}{\partial x_4} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \frac{\partial f_2}{\partial x_3} & \frac{\partial f_2}{\partial x_4} \end{pmatrix}.$$

With these rather cumbersome definitions, we have, for

$$x = -d_1(\mathbf{p})/d_2(\mathbf{p}), y = G(\mathbf{p})/d_2(\mathbf{p})^2,$$

that $Q(x) = y^2$.

B.1 Following the maps down

For debugging purposes and the generation of test data, it is useful, given an elliptic curve E, a point P on the curve, and a four-covering $\mathcal{H}$, to construct a point $R \in \mathcal{H}$ whose image on E is P. As always in this kind of work, we go via a point Q on an intermediate two-covering T.

Recall from appendix B that, if we have $Y^2 = f(X)$ for some X and Y, and f some quartic polynomial, and we consider the invariants I, J, g_4 and g_6 of the quartic, we find that $3g_4(X)/4Y^2$ is the X coordinate of a point on the elliptic curve $y^2 = x^3 - 27Ix - 27J$.

This is fairly easily reversible; suppose we have f and a point (x, y) on the attached elliptic curve. Then, for (X, Y) to lift to (x, y), we must have $Y^2 = f(X)$ and $4xY^2 = 3g_4(X)$; substituting, X must be a rational root of $4xf(t) - 3g_4(t) = 0$.

The four inverse images on the two-covering of a point on the elliptic curve will, by construction, differ by a 2-torsion point; if we assume that the elliptic curve was without 2-torsion, which is credible because we'd not be using the general two-descent on a curve with 2-torsion, the polynomial above has at most one rational root.

Suppose next that we have a four-covering, of the form $q_1(\mathbf{x}) = q_2(\mathbf{x}) = 0$, and a rational point $\mathbf{p}$ lying on it. From the previous part, we have an associated quartic $Q : \sum_{i=1}^{5} \sigma_i \lambda^{i-1}$, a pair of functions $d_1(\mathbf{x}) = \sigma_1^{-1}\mathbf{x}R_1\mathbf{x}^{\mathrm{T}}$ and $d_2(\mathbf{x}) = \sigma_5^{-1}\mathbf{x}R_2\mathbf{x}^{\mathrm{T}}$, with $X = -d_1(\mathbf{p})/d_2(\mathbf{p})$ being the X co-ordinate of a point on Q.

Reversing this construction gives us three quadratic equations in four homogenous variables:

$$\begin{aligned} q_1(\mathbf{p}) &= 0 && (e_1) \\ q_2(\mathbf{p}) &= 0 && (e_2) \\ d_1(\mathbf{p}) + X d_2(\mathbf{p}) &= 0 && (e_3) \end{aligned}$$

To solve such a set of equations, we compute resultants in all possible permutations:

$$\begin{aligned} R_{123} &:= \mathrm{Res}(e_1, e_2; z) \\ R_{124} &:= \mathrm{Res}(e_1, e_2; t) \\ R_{233} &:= \mathrm{Res}(e_2, e_3; z) \\ R_{234} &:= \mathrm{Res}(e_2, e_3; t). \end{aligned}$$

Now, we can take

$$\begin{aligned} e_y &:= \mathrm{Res}(R_{123}, R_{233}; t) \\ e_t &:= \mathrm{Res}(R_{123}, R_{233}; y) \text{ and} \\ e_z &:= \mathrm{Res}(R_{124}, R_{234}; y) \end{aligned}$$

to get homogenous degree-sixteen equations in x and y, z and t respectively; we homogenise by setting $x = 1$, and solve the equation to find the possible y, z or t-coordinates for points on $\mathcal{H}$.

Note that the only input we have here is the X co-ordinate of the point on the two-covering; to check that we have lifted $Q = (X, Y)$ rather than $-Q = (X, -Y)$, we take the candidate point $\mathbf{p} \in \mathcal{H}$ and check that the sign of $G(\mathbf{p})$ (where G is the Jacobian determinant defined in the previous part) is equal to the sign of Y.

Appendix C

Computer programs used in and developed during this work

The four-descent software developed during this work runs under the computer algebra package `magma` [7], developed by Cannon et al at the University of Sydney. For ease-of-interfacing reasons, and because its algorithms for counting points on elliptic curves are very substantially quicker for small p (because `magma`'s compute extra data for the curve over $\mathbb{F}_p$), some of the large computations and all work on analytic ranks were performed using `pari` [2] instead.

For performing two-descents, I used Cremona's `mwrank` [19]. To search for points on curves $y^2 = f(x)$, I use Cremona's `findinf`, which is a small wrapper around the `ratpoint` routine, which was based on code of Elkies and later refined and improved by Stoll and Stahlke – `mwrank` has this functionality incorporated. The only configuration option I used for `mwrank` was the `-b` setting, which controls the naïve height up to which points are searched for on the two-coverings; `-bN` corresponds to an x and y co-ordinate of absolute value $\leq \exp N$, for which there are $\exp \frac{3}{2}N$ possibilities – experiment suggests that the search takes time roughly proportional to $\exp 1.8N$, with a very small constant factor thanks to the efficiency of the sieves used: for $N = 13$, the computer I use requires about 20 seconds per curve.

The current version of the four-descent software can be obtained from

`http://tom.womack.net/d4.tar.gz`

as a collection of magma package files; decompress this file into an appropriate directory with the `tar zxf` command, and after issuing the command

```
AttachSpec(``d4spec.spec'');
```

`d4([a,b,c,d,e])` will perform a four-descent on the two-covering $y^2 = ax^4 + bx^3 + cx^2 + dx + e$, producing a set of minimised, reduced four-descendents. `ElkiesSearch4(dd, limit)` will search for a point on the four-covering `dd` with $\sum x_i^2 < 4\texttt{limit}^2$, and `FourDescendentAscent(dd, point)` will give the point on the associated elliptic curve corresponding to the point `point` on the four-covering `dd`. A cleaner version of this functionality should be integrated into a version of magma released sometime after September 2003, probably invisible to the user but accelerating the `MordellWeilGroup` command substantially.

C.1 Sample data sets

Various people provided data sets which were useful in the development of this software. When I needed two-coverings known to represent elements of Ш[2], I applied `mwrank` to the curves marked as rank zero with $\#Ш[2] = 4$ in Cremona's tables of curves with $N \leq 12000$.

Bibliography

[1] Greg W. Anderson, pre-print entitled "Wick's Theorem and the calculation of Jacobians of genus-one curves", available at time of writing from his Web page at `http://www.math.umn.edu/~gwanders`

[2] C Batut, K Belabas, D Bernardi, H Cohen, M Olivier, *User's Guide to PARI-GP*, version 2.1.1 (see also `http://www.parigp-home.de`)

[3] B J Birch, D J Lewis, T G Murphy, *Simultaneous Quadratic Forms*, Amer. J. Math **84** (1962), 110–115

[4] B J Birch, J R Merriman, *Finiteness theorems for binary forms with given discriminant*, Proc. London Math. Soc. (3) **24** (1972), 385–394

[5] B J Birch, H P F Swinnerton-Dyer, *Notes on elliptic curves I*, J. Reine Angew. Math, **212** (1963), 7–25

[6] R Bölling, *Die Ordnung der Schafarewitsch-Tate-Gruppe kann beliebig gross werden*, Math. Nachr. **67** (1975), 157–179

[7] W Bosma, J Cannon and C Playoust, *The* `magma` *algebra system I: the user language*, J Symbolic Comput **24** (1997), 235–265; see also `http://www.maths.usyd.edu.au:8000/u/magma`

[8] A Brumer, K Kramer, *The Rank of Elliptic Curves*, Duke Math J **44** (1977), 715–743

[9] A Brumer, O McGuinness, *The behavior of the Mordell-Weil group of elliptic curves*, Bull. AMS **23** (1990), 375 – 382

[10] D A Buchbaumm, D Eisenbud, *Algebra structures for finite free resolutions, and some structure theorems for ideals of codimension 3*, Amer. J. Math. **99** (1977), 447 – 485

[11] J Buchmann, *On the computation of units and class numbers by a generalisation of Lagrange's algorithm*, J. Number Theory **26** (1987), 8 – 30

[12] J P Buhler, B H Gross, D B Zagier, *On the conjecture of Birch and Swinnerton-Dyer for an elliptic curve of rank 3*, Math. Comp. **44** (1985), 473-481

[13] J W S Cassels, *Lectures on Elliptic Curves*, LMS Student Texts **24**, pub. Cambridge 1991; pp 66-72

[14] J W S Cassels, *The Mordell-Weil Group of Curves of Genus Two*, in Arithmetic and Geometry Papers Dedicated to I R Shafarevich on the Occasion of his Sixtieth Birthday, ed. M Artin and J Tate, pub. Birkhäuser 1983; pp 29–60

[15] H Cohen, *A Course in Computational Algebraic Number Theory*, Graduate Texts in Math. **138**, pub. Springer-Verlag 1996

[16] H Cohen, *Advanced Topics in Computational Algebraic Number Theory*, Graduate Texts in Math. **193**, pub. Springer-Verlag 2000

[17] H Cohen and H W Lenstra, *Heuristics on class groups of number fields*, Lecture Notes in Math. **1068**, pub. Springer-Verlag 1984; pp 33–62

[18] I Connell, *Elliptic Curve Handbook*, may be downloaded from `http://www.math.mcgill.ca/connell/`

[19] John Cremona's `mwrank` program, still under active development and available from his Web page at `http://www.maths.nottingham.ac.uk/personal/jec/ftp/progs`

[20] John Cremona's `allbsd` tables, available at time of writing from `http://www.maths.nottingham.ac.uk/personal/jec/ftp/data`

[21] J E Cremona, *Algorithms for modular elliptic curves*, second edition pub. Cambridge 1997

[22] J E Cremona, *Classical invariants and 2-descent on elliptic curves*, J Symbolic Computation **31** (2001), 71 – 87

[23] J E Cremona, *Reduction of binary cubic and quartic forms*, LMS J Comput Math **2** (1999), 62–92

[24] J E Cremona, D Rusin, *Efficient solution of rational conics*, Math Comp **72** (2003), 1417–1441

[25] J E Cremona, P Serf, *Computing the rank of elliptic curves over imaginary quadratic fields of class number one*, Math Comp **68** (1999), 1187–1200

[26] J E Cremona, M Stoll, T Fisher, C O'Neil, D Simon, *Explicit 3-descent on an elliptic curve* (in preparation)

[27] J E Cremona, M Stoll, *Minimisation and reduction of ternary cubics* (in preparation)

[28] H Darmon, *Heegner points, Heegner cycles and congruences*, CRM Proceedings and Lecture Notes volume **4** (1994) ed. H Kisilevsky and M Marty, 45–59

[29] C Delaunay, *Heuristics on Tate-Shafarevitch Groups of Elliptic Curves Defined Over* $\mathbb{Q}$, Experiment. Math **10** (2001), 191–196

[30] Z Djabri, N P Smart, *A comparison of direct and indirect methods for computing Selmer groups of an elliptic curve*, Proceedings of ANTS-III, Springer LNCS **1423** (1998), 502–513

[31] A Dujella, record ranks for elliptic curves with given torsion group over $\mathbb{Q}$ and $\mathbb{Q}(t)$, available from `http://www.math.hr/~duje/tors/tors.html`

[32] N D Elkies, personal communication

[33] N D Elkies, *Rational Points Near Curves and Small Nonzero* $|x^3 - y^2|$ *via Lattice Reduction*, Proceedings of ANTS-IV, Springer LNCS **1838** (2000), 33–64

[34] T Fisher, *Some examples of 5 and 7 descent for elliptic curves over* $\mathbb{Q}$, J. Eur. Math. Soc. **3** (2001), 169–201

[35] T Fisher, *On 5- and 7-descents for elliptic curves*, PhD thesis, Cambridge University 2000 [1]

[36] T Fisher, *Invariants for the elliptic normal quintic*; preprint

[37] T Fisher, *Curves of genus 1 defined by Pfaffians*; preprint

[38] C F Gauss, *Disquisitiones Arithmeticæ*, pub. Göttingen 1801; English translation by A A Clarke, pub. Yale 1965

[39] J Gebel, A. Petho, H.G. Zimmer, *On Mordell's Equation*, Compositio Mathematica **110** (1998), 335–367

[40] M J Greenberg, *Lectures on Forms in Many Variables*, pub. W A Benjamin Inc

[1] available from his Web site at `http://www.dpmms.cam.ac.uk/~taf1000`

[41] B H Gross, *Kolyvagin's work on modular elliptic curves*, in *Proceedings of the Durham Symposium on L-functions and arithmetic (1989)*, pub. Cambridge University Press 1991; pp 235 – 256

[42] D Hilbert, *Theory of Algebraic Invariants*, English translation pub. Cambridge 1993

[43] G Julia, *Etude sur les formes binaries non quadratiques*, Mem. Acad. Sci. d'Inst. France **55** (1917), 1-293

[44] R Kloosterman, *The p-part of Tate-Shaferevich groups of elliptic curves can be arbitrarily large*, preprint `math.NT/0303143`, 12 March 2003

[45] A W Knapp, *Elliptic Curves*, pub. Princeton 1992

[46] D S Kubert, *Universal bounds on the torsion of elliptic curves*, Proc. LMS **33** (1976), 193-237

[47] L Kulesz, *Elliptic curves of high rank with non-trivial torsion group over $\mathbb{Q}$, Experiment. Math* **10** *(2001), 475–480*

[48] S Lang, *Conjectured Diophantine Estimates on Elliptic Curves*, in Arithmetic and Geometry Papers Dedicated to I R Shafarevich on the Occasion of his Sixtieth Birthday, ed. M Artin and J Tate, pub. Birkhäuser 1983, pp 156–171

[49] P Llorente and J Quer, *On the 3-Sylow subgroup of the class group of quadratic fields*, Math. Comp. **50** (1988), 321–333.

[50] R Martin, W McMillen, *An elliptic curve of rank at least 24*, `NMBRTHRY` posting May 2000

[51] J R Merriman, S Siksek, N P Smart, *Explicit 4-descents on an elliptic curve*, Acta Arithmetica **77** (1996), 385–403

[52] M Prickett, forthcoming PhD thesis from University of Nottingham

[53] J Quer, *Corps quadratiques de 3-rang 6 et courbes elliptiques de rang 12*, C.R. Acad Sc. Paris I **305** (1987), 215–218

[54] G Salmon, *Higher Plane Curves*, 3rd edition printed Cambridge 1879

[55] Sang Yook A, Seog Young K, D C Marshall, S H Marshall, W G McCallum and A R Perlis, *Jacobians of genus one curves*, J Number Theory **40** (2001), 304–315

[56] E F Schaefer and M Stoll, *How to do a p-descent on an elliptic curve*, preprint 2001

[57] Suzanne Schmitt's thesis at Universität des Saarlandes

[58] P Serf, *The rank of elliptic curves over real quadratic number fields of class number 1*, PhD dissertation, Universität des Saarlandes, Saarbrücken 1995

[59] G Shimura, *Introduction to the arithmetic theory of automorphic functions*, pp 20–21

[60] Victor Shoup's `NTL` library, available under the Gnu Public Licence from `http://www.shoup.net`

[61] S Siksek, *Sieving for rational points on hyperelliptic curves*, Math Comp **70** (2001), 1661–1674

[62] S Siksek, *Descents on Curves of Genus 1*, PhD thesis, University of Exeter, 1995

[63] J H Silverman, *The arithmetic of elliptic curves*, GTM **106**, Springer-Verlag, New York, 1986

[64] J Silverman, *Computing Heights on Elliptic Curves*, Math. Comp. **51** (1988), 339-358

[65] D Simon, *Solving norm equations in relative number fields using S-units*, Math. Comp. **71** (2002), 1287–1305

[66] D Simon, *Computing the Rank of Elliptic Curves over Number Fields*, LMS JCM **5** (2002), 7–17

[67] D Simon, *Solving quadratic equations using reduced unimodular quadratic forms*, pre-print

[68] W A Stein and Mark Watkins, *A Database of Elliptic Curves – First Report*, Proceedings of ANTS-V, Springer LNCS **2369** (2002), 267–275

[69] M Stoll, personal communication

[70] M Stoll and J Cremona, *On the Reduction Theory of Binary Forms*, J Reine Angew. Math. (to appear)

[71] M Stoll and J Cremona, *Minimal models for 2-coverings of elliptic curves*, LMS JCM **5** (2002), 220–243

[72] R G Swan, *Generators and Relations for certain Special Linear Groups*, Advances in Mathematics **6** (1971), 1–77

[73] J A Todd, *Projective and Analytical Geometry*, pub. Pitman 1947

[74] Mark Watkins (Penn State University), personal communication

[75] A Weil, *Number Theory, An Approach Through History*, pub. Birkhäuser 1984

[76] K Wildanger, *Über das Lösen von Einheiten- und Indexformgleichungen in algebraischen Zahlhörpern mit einer Anwendung auf die Bestimmung aller ganzen Punkte einer Mordellschen Kurve*, PhD Thesis, TU Berlin, 1997