# Problems in Parameterization, Zero-Estimates, Point-Counting, and O-Minimality


John Armitage
St Peter's College
University of Oxford

A thesis submitted for the degree of
*Doctor of Philosophy*

Hilary 2021

# Acknowledgements

Firstly, I would like to thank my supervisor Jonathan Pila for his invaluable guidance and support, and generosity with his time, throughout my studies.

I would like to thank my fellow students in the Logic and Number Theory groups for making the department such a welcoming and enjoyable place to work, and for many enjoyable lunches. I would especially like to thank Vahagn Aslanyan and Sebastian Eterović for their advice and encouragement, and Guy Fowler and Oliviero Cassani, with whom I have had many productive discussions. I am also grateful to the staff of the Mathematical Institute, especially to Sandhya Patel, who have helped the administrative side of my studies run smoothly.

I would like to thank my family for their continual love and support.

Finally, I would like to thank Abi for many helpful discussions, and her companionship, especially throughout the coronavirus lockdowns.

# Abstract

In this thesis we consider problems of effectivity for zero-bounds, point-counting, computation, and class numbers.

Using the tameness (in the sense of o-minimality and Khovanskii [48]) of the modular function ($j$-function) and Weierstrass elliptic functions on certain complex arcs we prove new zero-bounds for some polynomials involving these functions. Using the method of mild parameterizations [67], we prove a new result on the number of algebraic points on a curve involving the $j$-function, and give a new proof of the six exponentials theorem.

We develop new algorithms to invert the $j$-function without using the elliptic curve representation of a $j$-invariant, to determine the torsion subgroup of rational elliptic curves in the manner of Doud [28], and to test an elliptic curve for complex multiplication. All these algorithms either improve on previous methods, or are of equivalent optimal complexity. We develop a new algorithm to find numbers satisfying a set of modular conditions, which is significantly faster than the fastest previous method [81], with which we find some new Cunningham chains of primes.

Finally we consider Gauss' conjecture [36] that there are no negative discriminants with one class of binary quadratic forms in each genus which have at least 32 genera. We apply our algorithm to find numbers satisfying modular conditions to prove a new general lower bound for such discriminants, adapt the methods of Watkins [86] in his determination of discriminants with class number $\leq 100$ to prove a large lower bound for discriminants indivisible by 2, 3 or 5, and adapt Baker's [6] method for class numbers one and two to prove that such a discriminant cannot have a particularly large prime factor.

# Contents

# List of Figures

# Chapter 1

# Introduction

This thesis consists of loosely connected results around modular and elliptic functions: we consider problems of effectivity for zero-bounds, point-counting, computation, and class numbers.

The $j$-function parameterizes the moduli space of elliptic curves, and has been studied since the nineteenth century in connection with the theory of complex multiplication — the theory of elliptic curves with endomorphism ring larger than $\mathbb{Z}$.

Every elliptic curve admits a uniformization $\mathbb{C}/\Lambda$, with $\Lambda$ a lattice in $\mathbb{C}$. Endomorphisms of $E$ correspond to complex $x \in \mathbb{C}$ such that $x\Lambda \subset \Lambda$. All elliptic curves have such endomorphisms by $\mathbb{Z}$, and if $x \notin \mathbb{Z}$, then $x\omega_1 = n\omega_1 + m\omega_2$ with $m \neq 0$, and $x\omega_2 = n'\omega_1 + m'\omega_2$ with $n' \neq 0$; thus $\frac{\omega_2}{\omega_1}$ is a (non-real) root of $mz^2 + (n + m')z + n'$, so is an imaginary quadratic irrational, as is $x$. The endomorphism ring is then an order contained in the ring of integers of an imaginary quadratic field. It is readily shown that $j$ takes algebraic values at quadratic irrationals, and it is further known by the theory of complex multiplication that $j$ takes algebraic integer values at these points. These algebraic integer values are called *singular moduli*. Schneider's theorem provides the converse that $j$ is transcendental at other algebraic arguments.

The theory of complex multiplication provides an analogue to the Kronecker–Weber theorem, which states that every finite abelian extension of $\mathbb{Q}$ is contained in some cyclotomic field. In other words, the maximal abelian extension of $\mathbb{Q}$ is obtained by adjoining the *special values* of the exponential function, the roots of unity. The analogous result for abelian extensions of imaginary quadratic fields is as follows: the maximal unramified abelian extension of an imaginary quadratic field $K$, with ring of integers $\mathbb{Z}[\tau]$, is obtained by adjoining $j(\tau)$ to $K$, and Hasse proved [39] that the maximal abelian extension of $K$ is generated over $K(j(\tau))$ by the (suitably normalized) values of Weierstrass elliptic function $\wp(z, \tau)$ evaluated at the division points of the lattice $\langle \tau, 1 \rangle$.

Associated to these special values are finiteness problems — there ought to be "few" special points on varieties except for those coming from "special subvarieties". A prototypical example is the question of when a plane curve may contain an infinite number of points $(x, y)$ both roots of unity, for which we have an *effective* answer.

**Theorem** ([11]). *Let $f = 0$ be an algebraic plane curve, and define the cyclotomic points of $\mathbb{C}^2$ to be points $(x, y)$ where $x$ and $y$ are roots of unity. Then $f = 0$ contains either at most $22V(f)$ cyclotomic points, where $V(f)$ is the volume of the Newton polytope of $f$, or infinitely many, in which case $f$ has a factor $x^i y^j - \omega$, for $(i, j) \neq (0, 0)$.*

Here we are considering the exponential function, the special points of which are roots of unity, and the special subvarieties of which are the factors $x^i y^j - \omega$.

The Manin–Mumford conjecture, proved by Raynaud [73] [74], is one such "special point" problem for abelian varieties.

**Theorem.** *Let $C$ be a curve of genus $\geq 2$ defined over a number field $K$, and $J$ be its Jacobian. Fixing an embedding $C \to J$ defined over $K$, the set $C(\overline{K}) \cap J(\overline{K})_{\text{tors}}$ is finite.*

In relation to the $j$-function and its analogues for Shimura varieties, we have the André–Oort conjecture, which concerns the behaviour of analogous special points.

**Conjecture.** *Let $S$ be a Shimura variety, and $V$ be an irreducible closed subvariety of $S$. Then $V$ contains only finitely many maximal special subvarieties.*

For the special case of products of modular curves,

$$V \subset Y = Y_1 \times \cdots \times Y_n,$$

where $Y_i = \Gamma_i \backslash \mathbb{H}$ for congruence subgroups $\Gamma_i$, one may consider $\Gamma_i = \mathrm{SL}_2(\mathbb{Z})$, and $Y_i = \mathbb{C}$. The special subvarieties are here induced by the modular polynomials $\Phi_N(X, Y)$. The case for $n = 2$ was obtained by André [4], and in this case one also has effective results of Bilu–Masser–Zannier [13] and Kühne [49]. The general case of $n$ was proved by Pila [68], who also considered a more general setting of products of modular curves, elliptic curves, and the multiplicative group of complex numbers; these in turn correspond to considering the special points of $j$, Weierstrass elliptic functions, and the exponential function.

In order to state the main point-counting result used by Pila, we first give a definition of o-minimality. As we do not make use of o-minimality in this thesis, we refer the reader to [29] for an introduction to o-minimality and its theory.

**Definition** ([90]). *A* pre-structure *$S$ is a sequence $\mathcal{S}_n$, $n \geq 1$, where each $\mathcal{S}_n$ is a collection of subsets of $\mathbb{R}^n$. It is called a* structure *over $\mathbb{R}$ if, for all $n, m \geq 1$, the following four conditions are satisfied:*

1. *$\mathcal{S}_n$ is a boolean algebra under the usual set theoretic operations,*
2. *$\mathcal{S}_n$ contains every semi-algebraic subset of $\mathbb{R}^n$,*
3. *if $A \in \mathcal{S}_n$ and $B \in \mathcal{S}_m$, then $A \times B \in \mathcal{S}_{n+m}$,*
4. *if $m \geq n$, $A \in \mathcal{S}_n$, then $\pi[A] \in \mathcal{S}_n$, where $\pi : \mathbb{R}^m \to \mathbb{R}^n$ is projection onto the first $n$ coordinates.*

*If, further, the boundary of every set in $\mathcal{S}_1$ is finite, then $\mathcal{S}$ is called an* o-minimal *structure over $\mathbb{R}$.*

Some examples, also described in [90], are the structure $\mathbb{R}_{\text{an}}$, where each $\mathcal{S}_n$ is the collection of globally subanalytic subsets of $\mathbb{R}^n$, and $\mathbb{R}_{\exp}$, where $\mathcal{S}_n$ is the collection of sets $\pi(f^{-1}(0))$, where $f : \mathbb{R}^m \to \mathbb{R}$ is a real exponential polynomial, and $\pi$ is projection onto the first $n \leq m$ coordinates. The main difficulty in proving that these structures are o-minimal is their closure under complementation. For $\mathbb{R}_{\text{an}}$, this is a consequence of Gabrielov's theorem [33], and for $\mathbb{R}_{\exp}$ this was established by Wilkie [89]. An o-minimal structure in which the $j$-function is definable is $\mathbb{R}_{\text{an, exp}}$, the structure generated by $\mathbb{R}_{\text{an}}$ and $\mathbb{R}_{\exp}$ under the operations implicit in the definition of o-minimal structures. This structure was proved to be o-minimal by Van den Dries and Miller [30]. In general, the structure generated by two o-minimal structures together may not be o-minimal.

For $X \subset \mathbb{R}^n$, define $X^{\text{alg}}$ to be the union of connected positive dimensional semi-algebraic subsets of $X$, and $N(Y, H, d)$ the number of points $(y_1, \ldots, y_n)$ in $Y$ such that the degree of each $y_i$ is bounded by $d$ and the absolute multiplicative height (c.f. [16]) of each $y_i$ is bounded by $H$. With these definitions we have the following theorem of Pila and Wilkie.

**Theorem** ([70] for rational points, [66] in general). *Let $X \subset \mathbb{R}^n$ be definable in an o-minimal structure over $\mathbb{R}$, let $d \geq 1$ and $\epsilon > 0$. Then there exists a constant $c(X, d, \epsilon)$ such that*

$$N(X - X^{\text{alg}}, H, d) \leq c(X, d, \epsilon) H^{\epsilon}.$$

This result is the best possible, as in [64] it is demonstrated that for any function $\epsilon(t) \to 0^+$ as $t \to \infty$, one can construct a transcendental analytic function $f$ on $[0,1]$ and a sequence of integers $T_i$ such that on its graph $X$,

$$N(X, H, 1) \geq T_i^{\epsilon(T_i)}.$$

However, for some o-minimal structures, one expects polylogarithmic bounds, for example in the conjecture of Wilkie.

**Conjecture** ([70]). *Let $X$ be definable in $\mathbb{R}_{\exp}$. Then there exist constants $c_1(X)$, $c_2(X)$ such that*

$$N(X - X^{\text{alg}}, H, 1) \leq c_1(X)(\log H)^{c_2(X)}.$$

One expects also that the constants $c_1(X)$, $c_2(X)$ would depend in a readily computable and reasonable manner on the complexity of defining $X$.

As the $j$-function is definable in the structure $\mathbb{R}_{\text{an,exp}}$, returning to our product of modular curves, the Pila–Wilkie theorem may be applied to the definable set

$$Z = \{(x_1, \ldots, x_n) | (j(x_1), \ldots, j(x_n)) \in V\},$$

with $\mathbb{C}^n$ considered as a subset of $\mathbb{R}^{2n}$. As the pre-special points are quadratic irrationals, the number of such pre-special points on $Z - Z^{\text{alg}}$ is bounded by $c(\epsilon)H^{\epsilon}$. The Galois orbits of the special points of $X$ will then provide the contrasting lower bounds.

The lower bounds in the situation of modular curves are the class numbers of the discriminants corresponding to the points of $X$, which may be bounded below by Siegel's theorem.

**Theorem** ([80]). *Let $\chi$ be a non-principal Dirichlet character of modulus $k$, and $\epsilon > 0$. Then there exists a constant (ineffective) $c(\epsilon) > 0$ such that*

$$L(1, \chi) \geq \frac{c(\epsilon)}{k^{\epsilon}}.$$

In particular, by Dirichlet's class number formula for negative discriminants,

$$h(-k) \geq \frac{c(\epsilon)}{\pi} k^{1/2-\epsilon}.$$

This gives an ineffective lower bound for the size of the Galois orbit of $j(\tau)$ when $\tau$ is a pre-special (quadratic irrational) point. The discriminant of a quadratic irrational point grows polynomially with its height — it is shown in [68] that $D_\tau \geq H(\tau)$.

Fixing say $\epsilon = 1/4$ in Siegel's theorem, and $\epsilon = 1/5$ in the Pila–Wilkie theorem, for sufficiently large height, these upper and lower bounds would contradict if such a point were not to lie in $Z^{\mathrm{alg}}$, and the proof of Pila's theorem proceeds by playing these bounds off against each other.

This thesis then considers problems of effectivity in this circle of problems. For the upper bounds on the number of algebraic points, we consider two prototypical situations involving the $j$-function and Weierstrass elliptic functions — polynomials in $z$ and $j(z)$ and in $z$ and $\wp(z)$. For a polynomial in $z$ and $j(z)$, we obtain a bound on its number of zeros in the whole fundamental domain of $j$. Effective zero-bounds of this sort are necessary to obtain effective bounds on the number of algebraic points of bounded height and degree via transcendence type arguments. It is natural to treat the $j$-function on the whole of its fundamental domain, and we believe our method of proof will admit some use in more general problems involving the $j$-function.

In relation to the lower bounds of Galois orbits, we consider Gauss' so-called one class per genus problem — that there are no negative discriminants with one class of binary quadratic forms in each genus beyond those which we already know. Just as the class number one problem corresponds to $j$ being an integer, at imaginary quadratic points of a discriminant with one class of binary quadratic forms in each genus, $j$ is totally real.

Associated to these problems are computational issues — we consider the problems of identifying elliptic curves with complex multiplication, or equivalently numbers which are singular moduli, and if so by what order in what field, and determining the torsion subgroup of elliptic curves. Most previous algorithms for these problems are based on the algebraic properties of the associated elliptic curves. This thesis presents improved algorithms for these two problems which consider the problem from the "lattice" point of view: detecting CM may be done by determining whether the lattice $\Lambda = \langle \omega_1, \omega_2 \rangle$ associated to an elliptic curve has the property that $\omega_1/\omega_2$ is a quadratic irrational; determining the torsion subgroup in a field $K$ of an elliptic curve may be done by determining if the image of a division point of $\Lambda$ is in $E(K)$. This approach is due to Doud [28], which we consider for $K = \mathbb{Q}$.

All of these problems have higher dimensional analogues — for abelian varieties, the analogue of having an endomorphism ring larger than $\mathbb{Z}$ is that $\mathrm{End}_{\mathbb{Q}}(A)$, the tensor product of $\mathrm{End}(A)$ with $\mathbb{Q}$, has a commutative subring of dimension twice the dimension of $A$. Such abelian varieties are said to be *of* CM *type*, and the analogues of the quadratic imaginary fields are complex quadratic extensions of totally real fields, which are termed CM-fields. This theory is described in [52].

The theory of complex multiplication has its generalization in class field theory, though the problem of explicit generation of the maximal abelian extension is a much more difficult problem, and in general unsolved. Just as with imaginary quadratic fields, where one needs more than $j$ and the exponential function, for CM-fields one needs more than the analogues of $j$ and the Weierstrass elliptic functions.

One extension of André–Oort to higher dimensions is that of Pila–Tsimerman [69] and Tsimerman [85], who obtained the André–Oort conjecture unconditionally for the moduli space of principally polarized abelian varieties $\mathcal{A}_g$.

We now describe in more detail the sections of this thesis.

## 1.1 Chapters in detail

### 1.1.1 Pfaffian control of polynomials involving $j$ and Weierstrass elliptic functions

In Chapter 2, we obtain bounds on the number of zeros of polynomials in $z$ and $j(z)$, and polynomials in $z$ and Weierstrass elliptic functions with rectangular fundamental domains. On certain complex lines, where $j(z)$ and these $\wp(z)$ are real, the derivatives of their inverses satisfy a particularly simple set of differential equations — they are Pfaffian functions, introduced by Khovanskii [48], who proved bounds on the number of zeros of such functions. This allows us to bound the winding number of $j$ and $\wp$ in regions bounded by these complex lines. We obtain an entirely new bound on the zeros of polynomials $P(z, j(z))$, polynomial in the degree of $P$, in the whole fundamental domain of $j$. Previous results of Binyamini [14] were restricted to compact subsets of the fundamental domain, and of a multiply-exponential order. For Weierstrass elliptic functions, non-uniform bounds $c(\wp)d^2$ of Masser [59] are known to hold for the number of zeros of polynomials $P(z, \wp(z))$ in a fundamental domain of $\wp(z)$. For Weierstrass elliptic functions whose associated lattice is rectangular we obtain a *uniform* bound of the form $cd^2$. This also improves, in this restricted case, uniform bounds of Jones and Schmidt [44] derived from their Pfaffian definition of $\wp(z)$, which are of the order $cd^{11}$.

### 1.1.2 Point-counting via mild parameterizations

In Chapter 3 we apply Pila's [67] method of mild parameterization to bound the number of algebraic points on graphs relating to the $j$-function and to give a proof of the six exponentials theorem. This is a real variable interpolation determinant

method for bounding the number of algebraic points of bounded height in a number field on transcendental surfaces in $(0,1)^n$.

### 1.1.3 Inverting the $j$-function and testing complex multiplication

In Chapter 4 we develop an algorithm to invert the $j$-function in quasilinear time and apply this to the problem of determining if an elliptic curve has complex multiplication. There are methods of inverting $j$ via the construction of an elliptic curve with $j$-invariant $j$, and computing its periods, such as that of Cremona and Thongjunthug [27], or the methods of Dupont [31], but it is considered of interest to study $j$ *without* passing through the elliptic curve representation. A previous algorithm of Alwaise [3] does this by different methods, but does not analyze the running time of their algorithm. The time complexity of our algorithm is essentially optimal, and of the same complexity as that of Cremona and Thongjunthug's and Dupont's algorithms.

As previously described, our algorithm to determine complex multiplication is based on the fact that an elliptic curve has complex multiplication if and only if its the $j$-invariant is the image under $j(z)$ of a quadratic irrational. In order to use this fact to show that an elliptic curve in fact *does* have complex multiplication we make use of the algebraic properties of both the $j$-invariant, coming from the data of the elliptic curve, and the algebraic properties of singular moduli.

Firstly, the discriminant of any potential quadratic irrational $\tau$ for which the $j$-invariant of $E$ is equal to $j(\tau)$ may be bounded in terms of the data of the elliptic curve. If the inverse of the $j$-invariant is sufficiently close to a quadratic irrational $\tau$ of discriminant at most this bound, then the $j$-invariant of $E$ and $j(\tau)$ — which are *both* algebraic numbers of height and degree bounded in terms of $E$ — will also be close. Unless they are equal, their difference is bounded below by Liouville's inequality. This algorithm has a lower time complexity than previous algorithms (either conditional, probabilistic, or unconditional).

### 1.1.4 Calculating rational torsion of elliptic curves

In Chapter 5 we use a recent result on the computation of Jacobi's theta functions to determine the torsion subgroup of a rational elliptic curve quasilinear time. Torsion points of elliptic curves correspond to division points of its associated lattice, and elliptic curves of the form $E : y^2 = x^3 + Ax + B$, with $A, B$ integers, have the property that all rational torsion points have integer coordinates. This allows us

to compute the potential torsion points of $E$ by evaluating the Weierstrass elliptic function associated to the lattice of $E$ at division points, which corresponds to the $x$-coordinate of our potential torsion point. Taking the nearest integer to our computed value of $x$, and the corresponding $y$, we may test whether these points lie on $E$. If so we may test whether they are torsion, by the group law of $E$. We show that one may use the aforementioned recent result of Labrande [50] on the computation of Jacobi's theta functions to reduce the computational complexity of computing the Weierstrass elliptic function associated to $E$.

### 1.1.5 Combining linear congruences

In Chapter 6 we describe an algorithm to combine linear congruences. We wish to find solutions to sets of linear congruences $x \mod p \in \{a_1, \ldots, a_k\}$, which our algorithm does by taking some products of primes $P = p_1 \cdot \cdots \cdot p_n$, $Q = q_1 \cdot \cdots \cdot q_m$, determining those $a$ which are admissible modulo $P$, and $b$ which are admissible modulo $Q$. By the Chinese remainder theorem all numbers $< PQ$ which are admissible modulo the $p_i$ and $q_i$ are of the form $am + bn - PQ[(am + bn)/PQ]$, which may then be tested, an idea due to Bernstein [8]. We describe an algorithm which follows this scheme, our novelty being a method of simultaneously testing many numbers $N = am + bn \mod PQ$. Instead of directly testing $N$, we construct a bit-vector of length $B$ which represents $a_i m + bn \mod PQ$ for $1 \leq i \leq B$, and special vectors for each prime $p$, so that in a single hardware operation (a $B$-bit OR) we test all $a_i m + bn \mod PQ$ simultaneously. In testing our algorithm proves to be roughly 50 times faster per-core than the previous fastest algorithm of Sorenson [81], though we are unable to perform the comparison on the same hardware. With this algorithm we also find some new Cunningham chains of primes.

### 1.1.6 Discriminants with one class of binary quadratic forms in each genus

Chapter 7 concerns negative discriminants with one class of binary quadratic forms in each genus.

In Section 7.1 we apply our algorithm to combine linear congruences to give a new lower bound of $10^{21}$ for any further such discriminants. In Section 7.2 we consider large lower bounds for these discriminants. In a similar manner to Watkins' [86] determination of discriminants with class number at most 100, we make use of $L$-functions with small zeros to eliminate certain ranges of discriminants. It is known

that this method may be used to eliminate the possibility of such discriminants in the interval $10^{68} \leq |d| \leq 10^{7000}$, so the problem is to eliminate discriminants $10^{21} \leq |d| \leq 10^{68}$. To this end we prove, in a computationally assisted manner, a new bound on the remainder terms of Epstein zeta functions associated to binary quadratic forms of the form $ax^2 + axy + cy^2$ and the character $\left(\frac{-17923}{\cdot}\right)$. By considering separately each possible number of prime factors of $d$, we are able, with some further computation, to obtain large lower bounds for discriminants not divisible by 2, 3, or 5. Further (incomplete) work indicates that certainly the case $2, 3 \nmid d$ is obtainable, and we believe that obtaining the lower bound for odd discriminants is feasible, and perhaps for all $d$ by reasonable improvements of our methods, and greater computational resources.

In Section 7.3 we consider discriminants with a particularly large prime factor $P \gg |d|^{1/2}(\log|d|)^5 \log\log|d|$. In a similar manner to Baker's [6] solution of the class number one problem, again making use of products of $L$-functions, we show that none may exist.

## 1.2 Background

We will now go over some background for the problems considered in this thesis.

### 1.2.1 Pfaffian functions

Pfaffian functions were introduced by Khovanskii [48] in his theory of so-called "fewnomials". The main idea is that real polynomials with few terms should be geometrically simple, having few zeros, and provides a way of formally describing the "tameness" of a function. Given a real domain $G \subset \mathbb{R}^n$, a sequence of functions $f_i : G \to \mathbb{R}$, $1 \leq i \leq r$ is a *Pfaffian chain* if there exist polynomials $P_{i,j}$ satisfying the triangular differential relations

$$\frac{\partial f_i(\boldsymbol{x})}{\partial x_j} = P_{i,j}(\boldsymbol{x}, f_1(\boldsymbol{x}), \ldots, f_i(\boldsymbol{x})),\ 1 \leq i \leq r,\ 1 \leq j \leq n.$$

The *degree* of such a Pfaffian chain is the maximum total degree of the $P_{i,j}$, and its *order* is $r$. A Pfaffian function is a function $f : G \to \mathbb{R}^n$ such that there exists a Pfaffian chain $f_1, \ldots, f_r$, such that $f = P(x, f_1, \ldots, f_r)$ for some polynomial $P$. The *order* of a Pfaffian function is $r$, and its *degree* is $(\alpha, \beta)$, where $\beta$ is the maximum total degree of $P$, and $\alpha$ is the degree of the Pfaffian chain $f_1, \ldots, f_r$.

A version of the main theorem on which this is based is as follows.

**Theorem** (§2.3, Theorem 2, [48], c.f. Lemma 2.1, [57]). *Let $G : \mathbb{R}^{n+1} \to \mathbb{R}^1$ be a smooth function with nondegenerate level set $M^n$. Let $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be a smooth proper map, and $\tilde{F} : M^n \to \mathbb{R}^n$ its restriction to $M^n$. Let, further, $\hat{J}$ be any smooth function on $\mathbb{R}^n$ that coincides on $M^n$ with the Jacobian $J$ of the map $(F, G) : \mathbb{R}^{n+1} \to \mathbb{R}^n \times \mathbb{R}^1$. Under these conditions the following holds: the maximum number of nondegenerate preimages of any point in the range of of the map $\tilde{F} : M^n \to \mathbb{R}^n$ is bounded by that of the map $(F, \hat{J}) : \mathbb{R}^{n+1} \to \mathbb{R}^n \times \mathbb{R}^1$.*

The definition of Pfaffian functions allow the repeated application of the theorem to reduce the problem of bounding the number of zeros of a function to that of a collection of algebraic curves. As an example we consider bounding the number of zeros of a polynomial in $x$ and an exponential function of the form $P(x, e^{x^2})$. Introducing the variable $u = e^{x^2}$, the zeros of $P(x, e^{x^2})$ are the solutions to the system of equations

$$\begin{aligned} P(x, u) &= 0 \\ u - e^{x^2} &= 0, \end{aligned}$$

the Jacobian of which is

$$J(x, u) = P_x(x, u) - \left(u - 2xe^{x^2}\right) P_y(x, u),$$

and restricted to $u = e^{x^2}$, $J(x, u) = P_x(x, u) - u(1 - 2x)P_y(x, u)$. The number of non-degenerate roots of $P(x, e^{x^2})$ is then bounded by the number of preimages of any point in the image of the function

$$(P(x, u), P_x(x, u) - u(1 - x)P_y(x, u)),$$

which by Bézout's theorem is bounded by $d(d+1)$, where $d$ is the degree of $P$. Khovanskii obtained the following general bound on the number of solutions of systems of Pfaffian functions.

**Theorem** ([48]). *Let $f_1, \ldots, f_k$ be Pfaffian functions on a domain $G \subset \mathbb{R}^k$ with a common chain of order $r$, and degrees $(\alpha, \beta_i)$. Then the number of non-degenerate zeros of the system $f_1 = \cdots = f_k = 0$ is bounded by*

$$2^{r(r-1)/2}\beta_1 \cdots \beta_k(\min\{n, r\}\alpha + \beta_1 + \cdots + \beta_k - n + 1)^r.$$

There are similar bounds for the number of connected components and the multiplicity of zeros of systems of Pfaffian functions. Khovanskii' theory, and this theorem, suggest that structures constructed from the usual arithmetic operations and Pfaffian functions are o-minimal, but does not prove it, as one needs closure under projections and complementation, so o-minimality is a substantial further step.

### 1.2.2 Mildness

Pila's method of mild parameterization [67] is a real-variable interpolation determinant method for counting algebraic points on the transcendental parts of surfaces $X \subset (0,1)^n$. Fixing a parameterization $\theta : (0,1)^k \to (0,1)^n$, the unit box is divided into cubes of side-length $r < 1$, and for algebraic points $x^{(i)}$, $1 \leq i \leq N$, in a given box, the determinant of the matrix

$$\left(\theta_1(z^{(i)})^{j_1} \cdots \theta_n(z^{(i)})^{j_n}\right)_{\substack{1 \leq i \leq N \\ j_1 + \cdots + j_n \leq M}}$$

is considered, where $z^{(i)}$ is the preimage under $\theta$ of $x^{(i)}$. On expanding $\theta(z^{(i)})$ as a Taylor series to a certain number of terms about the centre $z^{(0)}$ of the box, the lower order terms cancel out, leaving only terms with large powers of $z_j^{(0)} - z_j^{(i)}$, which is bounded in absolute value by $r/2$ for each ordinate. With bounds on the growth of the derivatives of $\theta_i$ this allows us to bound the determinant above. Lower bounds, supposing the determinant is non-zero, are furnished by the algebraicity of the points $\theta(z^{(i)}) = x^{(i)}$. In this way it is shown that the determinant is zero, and so the $x^{(i)}$ lie on a hypersurface of degree $\leq M$. To conclude we then need bounds on the number of points of intersection of $X$ with a hypersurface.

### 1.2.3 Weierstrass elliptic functions and the $j$-function

The following may be found in [53]. Weierstrass elliptic functions are meromorphic functions of the complex plane which are periodic with two periods — given two periods $\omega_1$, $\omega_2$, the Weierstrass elliptic function $\wp(z, \omega_1, \omega_2)$ may be defined as

$$\wp(z, \omega_1, \omega_2) = \frac{1}{z^2} + \sum_{m\omega_1 + n\omega_2 \neq 0} \left( \frac{1}{(z - m\omega_1 - n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right).$$

Then $\wp(z, \omega_1, \omega_2)$ satisfies the differential relation

$$(\wp(z, \omega_1, \omega_2)')^2 = 4\wp(z, \omega_1, \omega_2) - g_2\wp(z, \omega_1, \omega_2) - g_3\wp(z, \omega_1, \omega_2),$$

where

$$g_2 = g_2(\omega_1, \omega_2) = 60 \sum_{m\omega_1 + n\omega_2 \neq 0} \frac{1}{(m\omega_1 + n\omega_2)^4},$$

$$g_3 = g_3(\omega_1, \omega_2) = 140 \sum_{m\omega_1 + n\omega_2 \neq 0} \frac{1}{(m\omega_1 + n\omega_2)^6}.$$

This differential equation corresponds to an elliptic curve $y^2 = 4x^3 - g_2 x - g_3$, and $(\wp(z), \wp'(z))$ parameterizes this curve. The *discriminant* of this curve is $g_2^3 - 27g_3^2$.

If the discriminant is non-zero, then $E$ is a non-singular curve of genus one. Its $j$-invariant is defined to be $1728\frac{g_2^3}{g_2^3-27g_3^2}$, which is invariant under isomorphisms of $E$. As $\wp(z,\omega_1,\omega_2)$ is homogeneous of degree $-2$, it is natural to consider such functions with $\omega_2 = 1$. In this case, for $\tau$ in the upper half plane, $g_2(\tau)$ and $g_3(\tau)$ are scalings of the weight 4 and weight 6 Eisenstein series respectively. By the homogeneity of $g_2$ and $g_3$, the $j$-invariant may be viewed as a function of $\tau = \omega_1/\omega_2$.

The $j$-function is a holomorphic function of the upper half plane, invariant under $\mathrm{SL}_2(\mathbb{Z})$, that is invariant under transformations of the form

$$z \mapsto \frac{az+b}{cz+d}, \qquad ad-bc=1, \quad a,b,c,d \in \mathbb{Z},$$

and $j(z)$ generates the field of modular functions. The $j$-function has a standard fundamental domain

$$\mathcal{F} = \left\{ z \Big| -\frac{1}{2} < \mathrm{Re}(z) \leq \frac{1}{2}, |z| > 1 \right\} \cup \left\{ z \Big| 0 \leq \mathrm{Re}(z) \leq \frac{1}{2}, |z| = 1 \right\},$$

and all points in the upper half plane are equivalent under $\mathrm{SL}_2(\mathbb{Z})$ to some point in $\mathcal{F}$. Furthermore $j$ has a $q$-expansion

$$j(z) = q^{-1} + 744 + 196844q + 21493706q^2 + \cdots$$

where $q = e^{2\pi i z}$, which we will make use of at various points — primarily that $j$ is dominated by the first terms of this expansion when $\mathrm{Im}(z)$ is large. Schneider [78] proved that at algebraic points of degree $\geq 3$, $j(z_0)$ is transcendental, and by the theory of complex multiplication, $j$ is algebraic at points of degree 2. Associated to the $j$-function are the modular polynomials, $\Phi_N(Y)$, $N > 0$, the minimal polynomial of $j(Nz)$ (which, similarly to $j$ for $\Gamma_0$, generates the field of modular functions for the congruence subgroup $\Gamma_0(N)$) over the field $\mathbb{C}(j)$. Replacing every instance of $j$ in the coefficients of $\Phi_N(Y)$ by $X$, we obtain an integer bivariate polynomial $\Phi_N(X,Y)$ satisfying $\Phi_N(j(z), j(Nz)) \equiv 0$. This polynomial is symmetric for $N > 1$, and may be factorized as

$$\Phi_N(j(\tau), z) = (z - j(N\tau)) \prod_{\substack{ad=N \\ 0 \leq b < d}} \left( z - j\left( \frac{a\tau+b}{d} \right) \right).$$

We make use of the modular polynomial $\Phi_2(X,Y)$ in our algorithm to invert $j$ in order to obtain $j(2\tau), j\left(\frac{\tau}{2}\right)$, and $j\left(\frac{\tau+1}{2}\right)$ from $\Phi_2(j(\tau), z)$.

### 1.2.4 Discriminants and binary quadratic forms

The following is due to Gauss, and may be found in [36]. The discriminant of a binary quadratic form $f(x,y) = ax^2 + bxy + cy^2$ is defined to be $d = b^2 - 4ac$, and two binary quadratic forms are said to be *equivalent* if there exists a substitution

$$f(\alpha x + \beta y, \gamma x + \delta y) = g(x,y),$$

where $\alpha\delta - \beta\gamma = 1$, $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$. A class of equivalent binary quadratic forms has a canonical representative — the unique binary quadratic form in the class with $0 < a \leq \sqrt{|d|/3}$, and $-a < b \leq a \leq c$, and this may be found by an algorithmic process. Gauss defined the *composition* of two binary quadratic forms, and showed that the set of classes of binary quadratic forms of a given discriminant form a group under this operation. The *principal* binary quadratic form of a given discriminant is the unique binary quadratic form with $a = 1$, and is of the form

$$\begin{aligned} &x^2 - \frac{d}{4}y^2 && \text{if } d \text{ is even,} \\ &x^2 + xy + \frac{1-d}{4}y^2 && \text{if } d \text{ is odd.} \end{aligned}$$

A binary quadratic form $f$ is defined to be *ambiguous* if $f$ composed with itself is equivalent to the principal form of its discriminant, and reduced ambiguous binary quadratic forms must be of the form

$$\begin{aligned} &ax^2 + cy^2, \\ &ax^2 + axy + cy^2, \text{ or} \\ &ax^2 + bxy + ay^2. \end{aligned}$$

The *genus* of a binary quadratic form $f$ relates to the nature of those numbers which may be represented by it. Gauss showed that for any odd prime $p$ dividing $d$, all numbers represented by $f$ which are not divisible by $p$ agree in that they are all quadratic residues, or are non-residues, modulo $p$. This relationship to $p$ is termed a *particular character* of $f$. There may be one or two further particular characters (relating to the congruence classes of numbers represented by $f$ modulo 8) depending on the value of $d$ modulo 8. The set of genera is then a subdivision of the classes of binary quadratic forms of discriminant $d$ into sets which share *all* particular characters.

Genera may also be defined in terms of local equivalence: two quadratic forms are in the same genus if they are equivalent over all $\mathbb{Z}_p$ and $\mathbb{R}$. For binary quadratic

forms, this condition then reduces to Gauss' definition. This provides the natural extension of the notion of genus to integral quadratic forms of higher dimension.

In his investigations of binary quadratic forms Gauss raised the following questions (among many others) about the class groups of negative discriminants:

1. Are there finitely many negative discriminants with class number one? If so, are there any beyond those known?

2. Are there finitely many negative discriminants with one class of binary quadratic forms in each genus? If so, are there any beyond those known?

Heilbronn [41] proved that there are finitely many negative discriminants with class number 1, and it was proved that there were no further such discriminants by Heegner [40], Baker [5], and Stark [83]. It was proved by Chowla [23] that there are finitely many negative discriminants with one class of binary quadratic forms in each genus, and it was proved by Weinberger [88] that there is at most one such discriminant beyond those known.

# Chapter 2

# Pfaffian control of polynomials involving $j$ and Weierstrass elliptic functions

In this chapter we obtain some new bounds on the number of zeros of polynomials in $z$ and $j(z)$, and polynomials in $z$ and $\wp(z)$, where $\wp$ is a Weierstrass $\wp$-function associated to a lattice of the form $\langle 1, i\tau \rangle$, where $\tau$ is real. By the argument principle, the zeros of a holomorphic function in a closed, compact, simply connected region of the complex plane are controlled by behaviour of the function on its boundary. If the function is tame on the boundary, then we will obtain control over the number of zeros within the region. For $j$ and $\wp$ we use Pfaffian definitions of their inverses on suitable contours, and results of Khovanskii [48] to bound the zeros.

For polynomials in $z$ and $j(z)$, we obtain the following theorem.

**Theorem 2.1.** *Let $P(X, Y)$ be a complex polynomial of degree at most $d$ in either variable. Then $P(z, j(z))$ has at most $2^{80}d^{10}$ zeros in the standard fundamental domain.*

This we believe is a new result, giving a bound on the whole (non-compact) fundamental domain. It may be compared to Binyamini's result of a similar nature for Noetherian functions on relatively compact domains [14]. An application of this result here would yield a bound which depends on the size of a domain in which the zeros lie, and the size of $|j|$ on it. We also note that $j$ is o-minimal when restricted to its standard fundamental domain, which gives an ineffective finiteness result. The zero-bound depends polynomially on degree, and in that sense is not too far from the truth, as an obvious lower bound is $\gg d^2$. It is known that the inverse of $j(ix)$ is real and Pfaffian on the imaginary axis, a fact which may be deduced from its

expression in terms of the Gaussian hypergeometric functions, which are themselves Pfaffian on an interval. This is used in Jones and Schmidt's [44] Pfaffian definition of $\wp$ on the whole fundamental domain, and Bianconi's [12] work on the model theory of hypergeometric functions. To control the behaviour of the polynomial on our contour, we make use of a result from Ramanujan's theory of elliptic functions to alternative bases to obtain a direct and simple expression for the inverse of $j$ in terms of Gaussian hypergeometric functions, and near the cusps we use the fact that $j$ is dominated by the $q^{-1}$ term in its $q$-expansion. Though $j$ is a Noetherian function, it is unknown whether its real and imaginary parts are Pfaffian in the whole fundamental domain which would directly furnish a zero-bound, so considering a contour containing the fundamental domain bypasses this.

For the Weierstrass $\wp$ functions, we restrict our attention to those with rectangular lattice. In this case $\wp$ is real on the boundary of any fundamental domain, which improves the estimates under consideration, and obtain the following theorem.

**Theorem 2.2.** *Suppose that $\wp(z)$ is the Weierstrass $\wp$ function associated to the lattice $\langle 1, i\tau\rangle$, where $\tau > 0$ is real, and let $P(X, Y)$ be a complex polynomial of degree at most $d$ in either variable. Then $P(z, \wp(z))$ has at most $8d^2 + 14d + 5$ zeros in each fundamental domain.*

This may be compared with the non-uniform bound $c(\wp)d^2$ (see for example [59]), which holds for all $\wp$. A similar result is that of Jones and Schmidt [44] — they give a Pfaffian definition of $\wp$ on the whole fundamental domain, which yields a uniform bound of $cd^{11}$, where $c$ is an absolute constant. For non-real invariants, a similar result by our method of inferior quality could be obtained, as the need to take real and imaginary parts complicates the Pfaffian definitions involved.

## 2.1 Preliminaries

We first have two lemmas bounding the winding number related to the argument principle. We will apply these to segments of a closed contour, the first to estimate where there is a dominant term, and the second in order to apply bounds on zeros of real functions. Firstly we define the winding number of a path.

**Definition 2.1** (Definition 7.2, [84])**.** *A* continuous choice of argument *along a path $\gamma : [0, 1] \to \mathbb{C}\backslash\{0\}$ is a continuous map $\theta : [0, 1] \to \mathbb{R}$ such that*

$$e^{i\theta(t)} = \frac{\gamma(t)}{|\gamma(t)|}.$$

**Definition 2.2** (Definition 7.6, [84]). *The winding number of a path* $\gamma : [0,1] \to \mathbb{C}\backslash\{0\}$ *round the origin is*

$$\frac{\theta(1) - \theta(0)}{2\pi}$$

*for a continuous choice of argument* $\theta$ *along* $\gamma$*, and this is independent of the particular continuous choice of argument taken.*

For a holomorphic function $f$ and a (not necessarily closed) non-self-intersecting contour $\Gamma$ on which $f$ is non-zero, we will denote by $(f, \Gamma)$ the winding number of the path $f(\gamma(t))$, where $\gamma : [0,1] \to \mathbb{C}$ parameterizes the contour $\Gamma$. It is clear from the definition of a continuous choice of argument along a path that $(fg, \Gamma) = (f, \Gamma) + (g, \Gamma)$, and that given contours $\Gamma_1, \Gamma_2$, parameterized by $\gamma_1, \gamma_2$ such that $\gamma_1(1) = \gamma_2(0)$, then $(f, \Gamma_1 \cup \Gamma_2) = (f, \Gamma_1) + (f, \Gamma_2)$. Finally if $\Gamma$ is a closed non-self-intersecting contour, then the number of zeros of $f$ within $\Gamma$ is equal to $|(f, \Gamma)|$ — this is the argument principle.

Our first lemma is an analogue of Rouché's theorem for open paths.

**Lemma 2.1.** *Suppose that* $|f(z)| > 2|g(z)|$ *on the not necessarily closed contour* $\Gamma$*. Then*

$$|(f+g, \Gamma) - (f, \Gamma)| \leq \frac{1}{6}.$$

*Proof.* Firstly, as $|f(z)| > 2|g(z)| \geq 0$, $f(z) \neq 0$ and $(f+g)(z) \neq 0$ on $\Gamma$, so we may take

$$\begin{aligned}(f+g, \Gamma) &= \left(f \cdot \left(1 + \frac{g}{f}\right), \Gamma\right) \\ &= (f, \Gamma) + \left(1 + \frac{g}{f}, \Gamma\right),\end{aligned}$$

so that

$$|(f+g, \Gamma) - (f, \Gamma)| = \left|\left(1 + \frac{g}{f}, \Gamma\right)\right|.$$

Now for any $z \in \Gamma$, as $\frac{|g(z)|}{|f(z)|} \leq \frac{1}{2}$, the values taken by the function $h(z) := 1 + \frac{g(z)}{f(z)}$ lie in the ball of radius $\frac{1}{2}$ about 1, and so its argument is constrained to be $\leq \arctan\left(\frac{1}{\sqrt{3}}\right) = \frac{\pi}{6}$ in absolute value. We may then take $\theta(t) = \arg(h(\gamma(t)))$ as a continuous choice of argument for $h$, where arg is the principal argument $\arg(1) = 0$. We then have $|\theta(t)| \leq \frac{\pi}{6}$, so that

$$|(h, \Gamma)| = \left|\frac{\theta(1) - \theta(0)}{2\pi}\right| \leq \frac{1}{6}.$$

□

Our second lemma allows us to use zero estimates on the real or imaginary parts of holomorphic functions to bound their winding numbers on a path.

**Lemma 2.2.** *Let $f$ be a holomorphic function, non-zero on the not necessarily closed contour $\Gamma$. Then*

$$|(f,\Gamma)| \leq \#\{z \in \Gamma | \mathrm{Re}(f(z)) = 0\}/2 + 1$$

*and*

$$|(f,\Gamma)| \leq \#\{z \in \Gamma | \mathrm{Im}(f(z)) = 0\}/2 + 1.$$

*Proof.* It is clear that progressing through one revolution about the origin, the path $f(\gamma(t))$ must intersect the real axis at least twice, so

$$\#\{z \in \Gamma | \mathrm{Im}(f(z)) = 0\} \geq 2[|(f,\Gamma)|],$$

where $[|(f,\Gamma)|]$ is the integer part of $|(f,\Gamma)|$, and similarly for the real part. □

The following definition is less general than that of Pfaffian functions in [48], but is easier to work with (and restricted to one dimension).

**Definition 2.3.** *Let $f_1, \ldots, f_r$ be a sequence of analytic functions on the interval $(a,b)$. Then $f_1, \ldots, f_r$ is a* Pfaffian chain *if, for $1 \leq i \leq r$ and $x \in (a,b)$, there exist polynomials $P_i$ such that*

$$\frac{df_i(x)}{dx} = P_i(x, f_1(x), \ldots, f_i(x)).$$

*The* degree *of a Pfaffian chain is the maximum total degree of each $P_i$, and the* order *of a Pfaffian chain is its length. A* Pfaffian function *of order $r$ and degree $(\alpha,\beta)$ is a function $P(x, f_1(x), \ldots, f_s(x))$ where $P$ is a polynomial of total degree at most $\beta$, and $f_1(x), \ldots, f_s(x)$ are members of a Pfaffian chain of order $r$ and degree $\alpha$.*

We make use of the following bound, due to Khovanskii, on the number of zeros of a Pfaffian function.

**Theorem 2.3** ([48])**.** *Let $f$ be a Pfaffian function of order $r$ and degree $(\alpha,\beta)$ on the open interval $(a,b)$. Then the number of zeros of $f$ in $(a,b)$ is at most*

$$2^{r(r-1)/2}\beta(\alpha+\beta)^r.$$

In the proofs of our theorems we will bound the winding number of our function on a contour by decomposing it into paths. The absolute values of the winding numbers on these paths we bound by Lemma 2.2 together with Theorem 2.3 when we can make use of a Pfaffian definition, or by Lemma 2.1 when our function is dominated by a simple term.

## 2.2 Polynomials in $z$ and $j(z)$

Here we make use of an expression for the inverse of $j$ in terms of the Gaussian hypergeometric function, which is given, for $|z| < 1$, by

$$ {}_2F_1(a,b,c;z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, $$

where $(a)_n = \prod_{k=0}^{n-1}(a+k)$ is the rising factorial, and $c$ not a non-positive integer. Gauss determined 15 relations between those ${}_2F_1$ whose parameters differ by integers. We make use of the following two relations, where we have suppressed the parameters of the function, and use the notation $F(c\pm) = {}_2F_1(a,b,c\pm 1;z)$.

**Theorem 2.4** (Gauss, [35], Art. 11, Eq. 16; DLMF [1], Eq. 15.5.21)**.**

$$ \begin{aligned} z\frac{dF}{dz} &= (c-1)(F(c-) - F) \\ &= z\frac{(c-a)(c-b)F(c+) + c(a+b-c)F}{c(1-z)}. \end{aligned} $$

Now these allow us to construct a Pfaffian chain which $j^{-1}$ will be defined over on the imaginary axis.

**Lemma 2.3.** *The sequence of functions*

$$ \frac{1}{x}, \qquad \frac{1}{1-x}, \qquad \frac{F}{F(c+)}, \qquad F(c+), \qquad F, \qquad \frac{1}{F} $$

*is a Pfaffian chain of order* 6 *and degree* 6 *on* $(0,1)$.

*Proof.* We first note that $\frac{1}{x}, \frac{1}{1-x}$ is a Pfaffian chain of order 2 and degree 2. By Gauss' contiguous relations, we have

$$ \begin{aligned} x\frac{dF(c+)}{dx} &= c(F - F(c+)) \\ \frac{dF}{dx} &= \frac{(c-a)(c-b)F(c+) + c(a+b-c)F}{c(1-x)}. \end{aligned} $$

Consider

$$ \begin{aligned} \frac{d}{dx}\frac{F}{F(c+)} &= \frac{(c-a)(c-b)F(c+) + c(a+b-c)F}{c(1-x)F(c+)} - \frac{c(F - F(c+))F}{xF(c+)^2} \\ &= \frac{(c-a)(c-b)}{c(1-x)} + \left(\frac{a+b-c}{c(1-x)} - \frac{c-1}{x}\right)\frac{F}{F(c+)} - \frac{c}{x}\left(\frac{F}{F(c+)}\right)^2. \end{aligned} $$

Hence $\frac{1}{x}, \frac{1}{1-x}, \frac{F}{F(c+)}$ is a Pfaffian chain of order 3 and degree 3. Now consider

$$\frac{dF(c+)}{dx} = \frac{c(F - F(c+))}{x} = \frac{c}{x}\left(F(c+)\frac{F}{F(c+)} - F(c+)\right),$$

so $\frac{1}{x}, \frac{1}{1-x}, \frac{F}{F(c+)}, F(c+)$ is a Pfaffian chain of order 4 and degree 3. Taking the derivative of $F = F(c+)\frac{F}{F(c+)}$, we see that $\frac{1}{x}, \frac{1}{1-x}, \frac{F}{F(c+)}, F(c+), F$ is a Pfaffian chain of order 5 and degree 4. Finally taking the derivative of $1/F$ we see that

$$\frac{1}{x}, \qquad \frac{1}{1-x}, \qquad \frac{F}{F(c+)}, \qquad F(c+), \qquad F, \qquad \frac{1}{F}$$

is a Pfaffian chain of order 6 and degree 6. □

We make use of the following consequence, which is clear considering the above argument for ${}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} + \frac{y}{2}\right)$ and ${}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right)$.

**Lemma 2.4.** $\frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} + \frac{y}{2}\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right)}$ *is Pfaffian function of order* 9 *and degree* $(6, 2)$ *on* $(0, 1)$.

*Proof.* The argument runs the same, and yields the chain

$$\frac{1}{1+y}, \qquad \frac{1}{1-y}, \qquad \frac{F\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} + \frac{y}{2}\right)}{F\left(\frac{1}{6}, \frac{5}{6}, 2; \frac{1}{2} + \frac{y}{2}\right)}, \quad F\left(\frac{1}{6}, \frac{5}{6}, 2; \frac{1}{2} + \frac{y}{2}\right),$$
$$F\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} + \frac{y}{2}\right), \quad \frac{F\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right)}{F\left(\frac{1}{6}, \frac{5}{6}, 2; \frac{1}{2} - \frac{y}{2}\right)}, \quad F\left(\frac{1}{6}, \frac{5}{6}, 2; \frac{1}{2} - \frac{y}{2}\right), \quad F\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right),$$
$$\frac{1}{F\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right)},$$

and $\frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} + \frac{y}{2}\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1; \frac{1}{2} - \frac{y}{2}\right)}$ is then the product of the fifth and ninth functions in this chain. □

Now we express the inverse of $j$ on the imaginary axis in terms of the Gaussian hypergeometric functions by way of the following theorem, an inversion formula from the theory of elliptic functions to alternative bases — in this case the sextic theory.

**Theorem 2.5** (Theorem 4.10, $r = 1$, [25])**.** *Let $q$ be a real number in the interval $0 < q < 1$. Then*

$$q = \exp\left(-2\pi \frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, 1 - x\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, x\right)}\right),$$

*where, letting $P, Q, R$ be Ramanujan's Eisenstein series,*

$$x(1 - x) = \frac{Q(q)^3 - R(q)^2}{4Q(q)^3}.$$

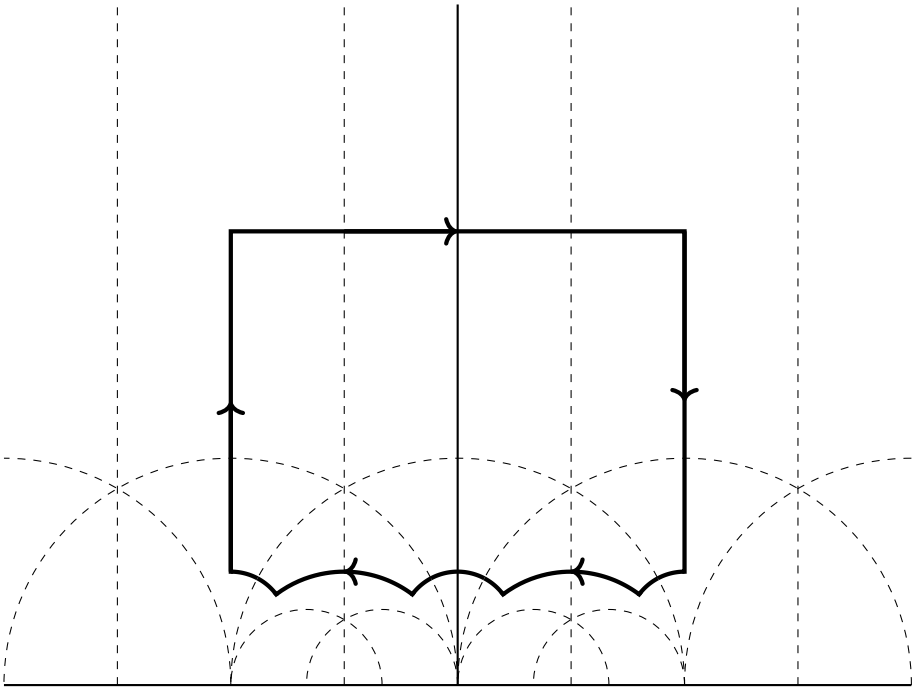

Figure 2.1: Our contour for $j$

Letting $q = e^{2\pi i\tau}$, and $\tau$ be purely imaginary, we take

$$\alpha = \frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{1728}{j(\tau)}},$$

which satisfies

$$\alpha(1-\alpha) = \frac{1728}{4j(\tau)} = \frac{Q(q(\tau))^3 - R(q(\tau))^2}{4Q(q(\tau))^3},$$

so that by Theorem 2.5,

$$\tau = i\frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{1728}{j(\tau)}}\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{1728}{j(\tau)}}\right)}.$$

So for $x \geq 1728$, letting $J(x) = j(ix)$,

$$J^{-1}(x) = \frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{1728}{x}}\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} - \frac{1}{2}\sqrt{1 - \frac{1728}{x}}\right)}.$$

*Proof of Theorem 1.* We apply the argument principle to a truncated fundamental domain together with its copies or half-copies under $\mathrm{SL}_2(\mathbb{Z})$ indicated by Figure 2.1. First consider a contour $\Gamma$ within the one indicated, containing all zeros

within the original contour, such that $P(z, j(z))$ is non-zero on this contour. Let $0 < \epsilon < \min_{z\in\Gamma}|P(z, j(z))|$. Then by Rouché's theorem, the number of zeros of $P(z, j(z)) + \epsilon e^{i\theta}$ within $\Gamma$ is equal to that of $P(z, j(z))$, for any $\theta$. Choose $\theta$ so that $P_\epsilon(z, j(z)) := P(z, j(z)) + \epsilon e^{i\theta}$ is non-zero on the original contour — this is possible by the discreteness of zeros of analytic functions.

We now bound the winding number of $P_\epsilon(z, j(z))$ on this boundary. The contour is chosen to be dominated by the $q^{-1}$ term of $j$ near the cusps. Let $j(it) = J(t)$. We will refer to elements of $\mathrm{SL}_2(\mathbb{Z})$ by $g$, with entries

$$\begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{pmatrix}$$

acting on $z$ by

$$g(z) = \frac{\tilde{a}z + \tilde{b}}{\tilde{c}z + \tilde{d}}.$$

The contour is composed of lines or curves near the cusps, and images of the imaginary axis under the action of some $g$.

**Copies of the imaginary axis $i\mathbb{R}_{\geq 1}$:** Here we have

$$\begin{aligned}
\mathrm{Im}(P_\epsilon\left(g(it), j\left(g(it)\right)\right)) &= \mathrm{Im}\left(P_\epsilon\left(\frac{\tilde{a}it + \tilde{b}}{\tilde{c}it + \tilde{d}}, j\left(it\right)\right)\right) \\
&= \mathrm{Im}\left(P_\epsilon\left(\frac{(\tilde{a}it + \tilde{b})(-\tilde{c}it + \tilde{d})}{\tilde{c}^2t^2 + \tilde{d}^2}, j\left(it\right)\right)\right) \\
&= \left(\frac{1}{\tilde{c}^2t^2 + \tilde{d}^2}\right)^d Q(t, J(t))
\end{aligned}$$

for some real $Q$, as $j(it)$ is real. $Q(X, Y)$ is of degree $\leq 2d$ in $X$ and $\leq d$ in $Y$, and $Q(t, J(t))$ has the same number of zeros as $\mathrm{Im}(P_\epsilon\left(g(it), j\left(g(it)\right)\right))$.

As $J(t)$ is 1–1 from $(1, \infty)$ to $(1728, \infty)$, letting $x = J(t)$, the number of zeros of $Q(t, J(t))$ is equal to that of $Q(J^{-1}(x), x)$ on $x > 1728$, and as $\sqrt{1 - \frac{1728}{x}}$ is 1–1 on $x > 1728$, letting $y = \sqrt{1 - \frac{1728}{x}}$, the number of zeros of $Q(J^{-1}(x), x)$ is equal to that of

$$Q\left(J^{-1}\left(\frac{1728}{1-y^2}\right), \frac{1728}{1-y^2}\right) = \left(\frac{1}{1-y^2}\right)^d R\left(J^{-1}\left(\frac{1728}{1-y^2}\right), y\right)$$

on $(0, 1)$, where $R(X, Y)$ is of degree $\leq 2d$ in $X$ and $\leq 2d$ in $Y$. Now using the expression for $J^{-1}$ in terms of the hypergeometric series ${}_2F_1$, we have that the number of zeros of $\mathrm{Im}(P(g(it), j(g(it))))$ is bounded by that of

$$R\left(\frac{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} + \frac{y}{2}\right)}{{}_2F_1\left(\frac{1}{6}, \frac{5}{6}, 1, \frac{1}{2} - \frac{y}{2}\right)}, y\right)$$

for $y \in (0,1)$. Now by Lemma 2.4, this is a Pfaffian function of order 9 and degree $(6, 6d)$. By Theorem 2.3, the number of zeros is then bounded by $2^{76}d^{10}$, so by Lemma 2.2, on this copy of the imaginary axis, we have

$$|(P_\epsilon(z, j(z)), \Gamma)| \leq 2^{76}d^{10}.$$

**Copies of** $\mathrm{Im}(z) = Y$**:** For these sections of the boundary, we have

$$P_\epsilon(g(x+iY), j(g(x+iY))) = \left(\frac{1}{\tilde{c}(x+iY)+\tilde{d}}\right)^d Q(x+iY, j(x+iY)).$$

For sufficiently large $Y$, the argument of

$$\left(\frac{1}{\tilde{c}(x+iY)+\tilde{d}}\right)^d$$

is constrained between $-\frac{d\pi}{2} - 0.01$ and $-\frac{d\pi}{2} + 0.01$, so its contribution to the winding number of $P_\epsilon(g(x+iY), j(g(x+iY)))$ is bounded by 0.02. We now consider the remaining term in $Q$.

Letting $j(z)^l h(z)$ be the term with the largest power of $j$ occurring in $Q(z, j(z))$, where $h(z)$ is its coefficient over the polynomials in $z$, let the degree of $h$ be $n$. For sufficiently large $Y$, $j(x+iY)^l(iY)^n$ is the dominant term in $Q(x+iY, j(x+iY))$, i.e. $|j(x+iY)^l(iY)^n| > 2|Q(x+iY, j(x+iY)) - j(x+iY)^l(iY)^n|$. In the same way the $e^{-2\pi i(x+iY)}$ term in the $q$-expansion of $j$ dominates $j(x+iY)$ when $Y$ is large, and so as $Y \to \infty$,

$$\frac{\left|Q(x+iY, j(x+iY)) - e^{-2l\pi i(x+iY)}(iY)^n\right|}{|e^{-2l\pi i(x+iY)}(iY)^n|} \to 0.$$

So taking $f(z) = (iY)^n e^{-2l\pi iz}$ and $g(z) = Q(z, j(z)) - f(z)$, for sufficiently large $Y$, on the contour $\Gamma = \left[-\frac{1}{2} + iY, \frac{1}{2} + iY\right]$, Lemma 2.1 gives the bound

$$\begin{aligned}
|(Q(z, j(z)), \Gamma)| &\leq \left|\left((iY)^n e^{-2l\pi iz}, \Gamma\right)\right| + \frac{1}{6} \\
&= l + \frac{1}{6} \\
&\leq d + \frac{1}{6}.
\end{aligned}$$

We also take $Y$ sufficiently large to ensure all zeros in the fundamental domain (and its copies) are interior to the contour, which is possible by virtue of the dominating term in $P(z, j(z))$ at the various cusps (or by the o-minimality of $j$ implying there are only finitely many zeros in a fundamental domain). Finally, the total winding number is bounded in absolute value by $8 \cdot 2^{76}d^{10} + 10d + 2 \leq 2^{80}d^{10}$, as there are 8 copies of the imaginary axis, and 10 copies or half-copies of the line $(-1/2 + iY, 1/2 + iY)$. $\square$

## 2.3 Polynomials in $z$ and Weierstrass elliptic functions

In this section we apply another theorem of Khovanskii to obtain a zero bound. First we define the following:

**Definition 2.4.** *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a smooth function. We say that $x \in \mathbb{R}^n$ is* non-degenerate *if the matrix*

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) & \dots & \frac{\partial F_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1}(x) & \dots & \frac{\partial F_m}{\partial x_n}(x) \end{pmatrix}$$

*has rank $m$. A* non-degenerate level set *of a function $F : \mathbb{R}^n \to \mathbb{R}$ is a level set $F^{-1}(c)$ such that all points $x \in F^{-1}(c)$ are non-degenerate.*

We now state the theorem of Khovanskii.

**Theorem 2.6** (§2.3, Theorem 2, [48], c.f. Lemma 2.1, [57])**.** *Let $G : \mathbb{R}^{n+1} \to \mathbb{R}^1$ be a smooth function with nondegenerate level set $M^n$. Let $F : \mathbb{R}^{n+1} \to \mathbb{R}^n$ be a smooth proper map, and $\tilde{F} : M^n \to \mathbb{R}^n$ its restriction to $M^n$. Let, further, $\hat{J}$ be any smooth function on $\mathbb{R}^n$ that coincides on $M^n$ with the Jacobian $J$ of the map $(F, G) : \mathbb{R}^{n+1} \to \mathbb{R}^n \times \mathbb{R}^1$. Under these conditions the following holds: the maximum number of nondegenerate preimages of any point in the range of of the map $\tilde{F} : M^n \to \mathbb{R}^n$ is bounded by that of the map $(F, \hat{J}) : \mathbb{R}^{n+1} \to \mathbb{R}^n \times \mathbb{R}^1$.*

We apply this to a polynomial in $x$ and $\wp^{-1}(x)$, where $\wp$ is a Weierstrass $\wp$-function, and use the argument principle to bound the zeros of $P(z, \wp(z))$ in its fundamental domain. We consider $\wp$ with lattices of the form $\langle 1, i\tau \rangle$, $\tau > 0$ real. The bound on the number of zeros follows in a similar way to §2.3 Theorem 1 of [48] — we proceed in this manner to give a better bound than simply applying Theorem 2.3.

First we have the differential equation satisfied by $\wp$,

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3,$$

and $\tilde{\wp}(x) := \wp(ix)$ satisfies the equation

$$\tilde{\wp}'(x)^2 = -4\tilde{\wp}(z)^3 + g_2\tilde{\wp}(z) + g_3,$$

where $g_2, g_3$ are the Weierstrass invariants of $\wp$, which are real as $\wp$ is associated to the lattice $\langle 1, i\tau\rangle$. For the derivatives of the inverses of $\wp(x)$ and $\tilde{\wp}(x)$, we have the expressions

$$\frac{d}{dx}\wp^{-1}(x) = \frac{1}{\wp'(\wp^{-1}(x))} = \frac{1}{\sqrt{4x^3 - g_2x - g_3}},$$
$$\frac{d}{dx}\tilde{\wp}^{-1}(x) = \frac{1}{\tilde{\wp}'(\tilde{\wp}^{-1}(x))} = \frac{1}{\sqrt{-4x^3 + g_2x + g_3}}$$

where the sign of the square root depends on the branch of the inverse of $\wp$ or $\tilde{\wp}$ which is under consideration. Note that both expressions will be real in the domains under consideration.

**Proposition 2.1.** *Let $\wp$ have lattice $\langle 1, i\tau\rangle$, where $\tau > 0$ is real, and $P(X, Y)$ be a complex polynomial of degree at most $d$ in either variable. Then the number of zeros of $\mathrm{Im}(P(z, \wp(z)))$ on the open line $(\beta, \beta + \gamma)$, $\gamma \in \{1, i\tau\}$ between two adjacent poles of $\wp$ is bounded by $4d^2 + 6d + 1$.*

*Proof.* By periodicity, $\wp(z) = \wp(z - \beta)$, so we may take a transformation of $z$ to consider the lines $(0, 1)$, and $(0, i\tau)$. As $\wp$ has lattice $\langle 1, i\tau\rangle$, and $\tau$ is real, it is real on these lines, and $\mathrm{Im}(P(z, \wp(z))) = Q(x, \wp(\gamma x))$, $\gamma \in \{1, i\}$ for some real polynomial $Q(X, Y)$ of degree at most $d$. The argument proceeds identically for either line, so we consider $(0, 1)$.

Let $x_0 \in (0, 1)$ be such that $\wp'(x_0) = 0$. Then $\wp(x)$ is 1–1 on the interval $(0, x_0)$, so the number of zeros of $Q(x, \wp(x))$ is equal to that of $Q(\wp^{-1}(x), x)$ on the interval $(\wp(x_0), \infty)$. W consider the system

$$Q(u, x) = 0$$
$$u - \wp^{-1}(x) = 0.$$

Letting $J$ be the Jacobian of the system $(Q, u - \wp^{-1}(x))$, by Theorem 2.6, the number of nondegenerate solutions the system is bounded by an upper bound of the number of nondegenerate preimages of any point in the range of the system

$$Q(u, x)$$
$$J(u, x).$$

Letting $Q_X(X, Y) = \frac{\partial}{\partial X}Q(X, Y)$, and similarly for $Q_Y(X, Y)$, $J(x, u)$ is given by

$$\frac{Q_X(u, x)}{\sqrt{4x^3 - g_2x - g_3}} + Q_Y(u, x).$$

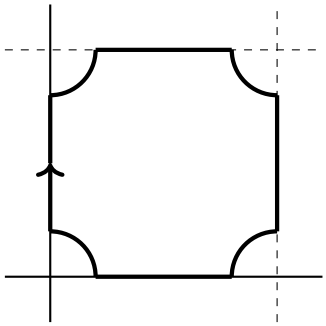

Figure 2.2: Our contour for $\wp$

Taking some point $(a, b)$ in the range of the system $(Q, J)$, we bound the number of nondegenerate preimages. If $J(x, u) = b$, then

$$J(u, x) = \frac{Q_X(u, x)}{\sqrt{4x^3 - g_2x - g_3}} + Q_Y(u, x) = b,$$

and this holds iff

$$Q_X(u, x) + (Q_Y(u, x) - b)\sqrt{4x^3 - g_2x - g_3} = 0,$$

and if this holds, then

$$\begin{aligned}(Q_X(u, x) + (Q_Y(u, x) - b)\sqrt{4x^3 - g_2x - g_3}) \\ \cdot (Q_X(u, x) - (Q_Y(u, x) - b)\sqrt{4x^3 - g_2x - g_3}) \\ = Q_X(u, x)^2 - (Q_Y(u, x) - b)^2(4x^3 - g_2x - g_3) = 0\end{aligned}$$

holds, so that the number of preimages of $(a, b)$ is bounded by the number of nondegenerate solutions of

$$\begin{aligned}Q(u, x) - a = 0 \\ Q_X(u, x)^2 - (Q_Y(u, x) - b)^2(4x^3 - g_2x - g_3) = 0,\end{aligned}$$

which by Bézout's theorem is bounded by $2d^2 + 3d$. The interval $(x_0, 1)$ is similar.

□

*Proof of Theorem 2.* Let $P(z) := P(z, \wp(z))$. We first take a box $B$ within the fundamental domain such that all zeros of $P$ in the interior of the fundamental domain lie within $B$. Next, define $\epsilon, \theta$ such that $0 < \epsilon < \min_{\partial B}|P|$, and $P_\epsilon(z) := P(z) + \epsilon e^{i\theta} \neq 0$ for $z \in \partial\mathcal{F}$. By Rouché's theorem, $P_\epsilon(z)$ has the same number of zeros in $B$ counting multiplicity as $P(z)$.

We now bound the number of zeros of $P_\epsilon$ within the contour $\Gamma$ given by the truncations of the lines on the boundary of $\mathcal{F}$ united with interior quarter-circles about the poles of $\wp(z)$, as indicated in Figure 2.2. The common radii of these circles is taken so that there are no zeros of $P_\epsilon$ in a disc of this radius about the poles, and such that they do not intersect $B$. The radii will be taken sufficiently small subject to this. As $B$ lies in the interior of $\Gamma$, a bound upon the number of zeros of $P_\epsilon$ within $\Gamma$ bounds the number within $B$, and so bounds that of $P(z, \wp(z))$ within the fundamental domain.

By the argument principle, the number of zeros of $P_\epsilon$ within $\Gamma$ is equal to

$$|(P_\epsilon(z), \Gamma)|,$$

which we now estimate. As $\wp(z)$ is real on the lines of $\Gamma$, considering for example the line $z = ix$, for $x$ real,

$$\mathrm{Im}(P_\epsilon(z)) = \mathrm{Im}(P(ix, \wp(ix))) + \epsilon \sin(\theta) = Q(x, \wp(ix)).$$

The number of zeros of which is, by Proposition [2.1], bounded by $4d^2 + 6d + 1$, and holds for any radius of quarter-circles. The other lines are similar, and so our bound for the contribution to the winding number of all four lines is $8d^2 + 12d + 6$ by Lemma 2.2.

On the quarter-circles, we may expand $P_\epsilon$ into its Laurent series — given a sufficiently small radius $\delta$, the term of smallest power will dominate. We let $\Gamma$ be the path $p + \delta e^{i\xi}$, where $\xi$ ranges over the appropriate interval of length $\pi/2$ in $[0, 2\pi]$, for the particular pole $p$, to be the interior quarter-circle. This path has length $\delta\pi/2$. Writing

$$P_\epsilon(z) = a_k(z-p)^k + \sum_{n=k+1}^{\infty} a_n(z-p)^n,$$

with $a_k \neq 0$, we have, as the radius $\delta \to 0$,

$$\frac{P_\epsilon(z) - a_k(z-p)^k}{a_k(z-p)^k} \to 0.$$

So for sufficiently small $\delta$, we have, letting $f(z) = a_k(z-p)^k$, and $g(z) = P_\epsilon(z) - f(z)$, by Lemma 2.1,

$$\begin{aligned} |(P_\epsilon(z), \Gamma)| &\leq \left|\left(a_k(z-p)^k, \Gamma\right)\right| + \frac{1}{6} \\ &\leq |k|/2 + \frac{1}{6} \\ &\leq d + \frac{1}{6}, \end{aligned}$$

where the last inequality follows from the fact that $\wp$ has poles of order 2, so $-2d \leq k \leq d$. As there are 4 quarter-circles about the poles of $\wp$, and 4 line segments, the absolute value of the winding number is bounded by $8d^2 + 16d + 7$, yielding the theorem. $\square$

# Chapter 3

# Point-counting via mild parameterization

In this chapter we give a bound for algebraic points of bounded height in a real number field on curves of the form $(\alpha t, j(it))$, where $j$ is the $j$-function, and a proof of the real six exponentials theorem, via mild parameterizations, introduced by Pila in [65]. For $\alpha$ algebraic, the algebraic points on the graph $(\alpha t, j(it))$ are exactly the quadratic irrationals, but for $\alpha$ transcendental, this provides a new result. We view this as a prototypical effective point counting result for the $j$-function.

Applying a mild parameterization to the $j$-function on the imaginary axis, and our zero-bound of the previous chapter, we obtain the following theorem.

**Theorem 3.1.** *Let $J(t) = j(it)$ and $\alpha > 0$. Then in a given number field $F$ of degree $f$, the number of algebraic points of height at most $T$ lying on the curve $(\alpha t, J(t))$ for $t > 1$ is $O\left((f^2 \log T)^{12+o(1)}\right)$.*

For the six exponentials theorem, as indicated in [67], by improving the mildness estimate for the parameterization $e^{\alpha(1-x^{-g})}$ of $x^\alpha$ on $(0, 1)$, we are able to sufficiently bound the number of algebraic points on the curve $(x, x^\alpha)$, for $\alpha$ irrational, to give a proof of the fol lowing theorem.

**Theorem 3.2** (Lang [51], Ramachandra [72])**.** *Let $(x_1, x_2)$ and $(y_1, y_2, y_3)$ be tuples of real numbers linearly independent over $\mathbb{Q}$. Then at least one of $e^{x_i y_j}$, for $i = 1, 2$, $j = 1, 2, 3$, is transcendental.*

We would note that approaching the *four* exponentials conjecture via mild parameterizations seems to have the same fundamental obstructions as the usual transcendece methods. The improvement of our mildness estimate of the function $x^\alpha$ also improves Pila's [67] bound for algebraic points in a given real number field

$F$ on the exponential-algebraic surface $\log x \log y = \log z$ from $O\left((\log T)^{44+\epsilon}\right)$ to $O\left((\log T)^{35+\epsilon}\right)$.

We note that by these methods one may prove bounds for algebraic points on the graphs of Riemann's Zeta function and Euler's Gamma function of $O\left((f^2 \log T)^{2+o(1)}\right)$. Up to the factor $O\left((\log T)^{o(1)}\right)$, these are of the same order as estimates which may be obtained by complex analytic methods (c.f. the results of Masser [58] and Besson [10], which consider restrictions to discs of radius 1). The method of mild parameterization is however restricted to counting points in a fixed field, whereas these results bound *all* points of bounded height and degree.

Also of interest are similar results of Boxall–Jones [18] for the restriction of certain analytic functions to compact disks, giving a bound of, for $C$ effectively computable from data of the function, $C\left(\log H\right)^{3+\epsilon}$ algebraic points of degree $d$ and height $H$.

In the following, the height of an algebraic number shall always be taken as the absolute multiplicative height $H$ (c.f. [16]), and we note its following properties for an algebraic number $\alpha$ of degree $f$:

$$H(\alpha_1 + \alpha_2) \leq 2H(\alpha_1)H(\alpha_2), \qquad H(\alpha^{-1}) = H(\alpha), \qquad |\alpha| \leq H(\alpha)^f.$$

## 3.1 Mildness

In the following definitions, $\mu \in \mathbb{N}^k$ for some $k$, $\mu! = \prod_{i=1}^k \mu_i!$, $|\mu| = \sum_{i=1}^k \mu_i$, and

$$\partial^\mu \theta = \frac{\partial^{|\mu|}\theta}{\partial x_1^{\mu_1} \cdots \partial x_m^{\mu_m}}.$$

**Definition 3.1.** *A function $\theta : (0,1)^k \to (0,1)^n$ is $(A,C)$-mild if, for each co-ordinate $\theta_i$ of $\theta$,*

$$\sup_{x \in (0,1)^k} |\partial^\mu \theta_i| \leq \mu!(A|\mu|^C)^{|\mu|}.$$

**Definition 3.2.** *Let $X \subset (0,1)^n$. Then $X$ has a $(J,A,C)$-mild parameterization if there exist $J$ $(A,C)$-mild functions $\theta_j : (0,1)^{\dim X} \to (0,1)^n$ such that*

$$X = \bigcup \theta_j \left((0,1)^{\dim X}\right).$$

In all our applications, $X$ will be a curve.

We now define some further quantities. Firstly, for any $Y \subset \mathbb{R}^n$, we define $Y^{\mathrm{alg}}$ to be the union of all connected semialgebraic subsets of $Y$ of positive dimension. Secondly, letting $F$ be a real number field, we define $Y(F,T)$ to be the set of points

of $F^n$ contained in $Y - Y^{\text{alg}}$ where the absolute multiplicative height of each ordinate is at most $T$.

The main theorem we use is Pila's following result on algebraic points on the transcendental part of a surface.

**Theorem 3.3** (Corollary 3.3, [67]). *Suppose $Y \subset (0,1)^n$ of dimension $k$ has a $(J, A, C)$-mild parameterization. Let $f$ be a positive integer, and $F \subset \mathbb{R}$ a number field of degree $f$ over $\mathbb{Q}$. Then $Y(F, T)$ is contained in at most*

$$Jc(k,n)^f A^{(k+1)(1+o(1))} (f \log T)^{C\left(\frac{n(k+1)}{n-k}\right)(1+o(1))}$$

*intersections of $Y$ with hypersurfaces (possibly reducible) of degree*

$$d = \left[ (f \log T)^{\frac{k}{n-k}} \right],$$

*where "$1 + o(1)$" is taken as $T \to \infty$ with implicit constants depending only on $k$ and $n$.*

We note that in the proof of Theorem 3.2 of [67], by taking $d$ (in the context of the proof of Theorem 3.2) to be $\left[ (f \log T)^{\frac{n}{n-k}} \right]$ as opposed to $\left[ (\log T)^{\frac{n}{n-k}} \right]$, the dependence of the above result in $f$ is reduced to

$$Jc(k,n) A^{(k+1)(1+o(1))} \left( f^2 \log T \right)^{C\left(\frac{n(k+1)}{n-k}\right)(1+o(1))}$$

intersections with hypersurfaces of degree

$$d = \left[ \left( f^2 \log T \right)^{\frac{k}{n-k}} \right],$$

and it is in this form we shall apply the theorem. We only apply this to curves in $\mathbb{R}^2$, so in the following, $n = 2$ and $k = 1$.

## 3.2 The $j$-function on the imaginary axis

In order to bound the number of algebraic points on the curve $(\alpha t, j(it))$, for $\alpha > 0$, $t > 1$, we first introduce a parameter to bring our curve into the box $(0,1)^2$. Letting

$$\eta = \begin{cases} 1 \text{ if } \alpha \geq 1, \\ \left( \left[ \frac{1}{\alpha} \right] + 1 \right)^{-1} \text{otherwise,} \end{cases}$$

we will consider points on the curve $\left( \frac{\eta t^g}{\alpha}, \frac{1}{j(it^{-g})} \right) \subset (0,1)^2$, $t \in (0,1)$. The algebraic points of this curve are, modulo the rational factor of $\eta$, the inverses of those on $(\alpha t, j(it))$ — an algebraic point $(\alpha t_0, j(it_0))$ will be witnessed by the corresponding algebraic point $\left( \frac{\eta t_1^g}{\alpha}, \frac{1}{j(it_1^{-g})} \right)$, with $t_1 = t_0^{-1/g}$.

**Lemma 3.1.** *Let $J(t) = j(it)$. Then the function*

$$t \mapsto \frac{1}{J(t^{-g})},\ t \in (0,1)$$

*is $\left(32g, 1 + \frac{1}{g}\right)$-mild.*

*Proof.* We will use some well known identities ([45], pg. 8) — with

$$L(t) = E_2(it), \qquad M(t) = E_4(it), \qquad N(t) = E_6(it),$$

where $E_2$, $E_4$, and $E_6$ are the usual Eisenstein series,

$$\frac{J'(t)}{J(t)} = 2\pi\frac{N(t)}{M(t)}, \qquad N' = \frac{LN - M}{2}, \qquad M' = \frac{LM - N}{3}, \qquad L' = \frac{L^2 - M}{12}.$$

We consider the $k$'th derivative, where we do not collect any terms, of $F(t) := J\left(t^{-g}\right)^{-1}$,

$$\frac{d^k F}{dt^k} = \sum_{i=1}^{m} \alpha_i \frac{L\left(t^{-g}\right)^{a_i} M\left(t^{-g}\right)^{b_i} N\left(t^{-g}\right)^{c_i}}{t^{d_i} J\left(t^{-g}\right)}.$$

At each successive derivative one term gives rise to at most 8 more terms, as differentiating the $L$, $M$, or $N$ factors gives rise to two terms each, and differentiating the $J^{-1}$ or $t^{-c}$ factors gives rise to one. By the relations between $J$, $L$, $M$, and $N$,

$$\begin{aligned} -k \leq b_i &\leq k, \\ 0 \leq a_i, c_i &\leq k, \\ d_i &\leq (g+1)k, \\ |\alpha_i| &\leq (2\pi)^k (g+1)^k k!, \\ m &\leq 8^k. \end{aligned}$$

For $t \in (0,1)$, $L$, $M$, and $N$ satisfy

$$\begin{aligned} 0 \leq L\left(t^{-g}\right) &\leq 1, \\ 1 \leq M\left(t^{-g}\right) &\leq 2, \\ 0 \leq N\left(t^{-g}\right) &\leq 1. \end{aligned}$$

Now by the $q$-expansion of $J$, for $t > 1$, $J(t) \geq e^{2\pi t}$, so that, combining these bounds, we have

$$\left|\frac{d^k F(t)}{dt^k}\right| \leq (64\pi)^k g^k k! \frac{1}{t^{(g+1)k}} e^{-2\pi t^{-g}}.$$

This bound has its maximum at

$$\left(\frac{2\pi g}{(g+1)k}\right)^{\frac{1}{g}},$$

so, for some $A$,

$$\sup_{x\in(0,1)}\left|\frac{d^kF}{dt^k}\right| \leq 32^k g^k k! k^{k\left(1+\frac{1}{g}\right)},$$

hence the map is $\left(32g, 1+\frac{1}{g}\right)$-mild. □

*Proof of Theorem 3.1.* Applying Theorem 3.3 with $g = [\log\log T]$, the algebraic points of height $\leq T$ are contained in

$$O\left(g^{2+o(1)}(f^2\log T)^{4+\frac{4}{g}+o(1)}\right) = O\left((f^2\log T)^{4+o(1)}\right)$$

intersections of $\left(\frac{\eta t^g}{\alpha}, \frac{1}{J(t^{-g})}\right)$ with hypersurfaces of degree

$$d = [f^2\log T].$$

Now as $J(t)\neq 0$ for $t>1$, for a polynomial $P(X,Y)$ of degree $D_1$ in $X$ and $D_2$ in $Y$, the zeros of $P\left(\frac{\eta t^g}{\alpha}, \frac{1}{J(t^{-g})}\right)$ for $t\in(0,1)$ are in 1–1 correspondence with those of

$$Q(t^g, J(t^{-g})) = J(t^{-g})^{D_2}P\left(\frac{\eta t^g}{\alpha}, \frac{1}{J(t^{-g})}\right), \quad t\in(0,1),$$

and as $t\mapsto t^{-1}$ is bijective from $(0,1)$ to $(1,\infty)$, these are in 1–1 correspondence with zeros of $Q(t^{-g}, J(t^g))$ for $t>1$. As $t\mapsto t^g$ is bijective for $t>1$, these are in 1–1 correspondence with zeros of

$$R(t, J(t)) = t^{D_1}Q(t^{-1}, J(t)), \quad t>1,$$

which is of degree $\leq d$. By Theorem 2.1 of the previous chapter, there are at most $2^{80}(f^2\log T)^{10}$ such zeros.

As we have observed, algebraic points on $(\alpha t, J(t))$, $t>1$ give rise to points on $\left(\frac{\eta t^g}{\alpha}, \frac{1}{J(t^{-g})}\right)$, $t\in(0,1)$ of heights bounded by $O(T)$, and consequently there are $O\left((f^2\log T)^{14+o(1)}\right)$ points of height $T$ in a given number field of degree $f$ on the graph of $(\alpha t, J(t))$ for $t>1$. □

## 3.3 The six exponentials theorem

For the six exponentials theorem, we first prove our aforementioned improvement of the mildness estimate for $x^\alpha$, $x \in (0,1)$.

**Lemma 3.2.** *Let $\alpha \geq 1$ and $g$ be a positive integer. Then*

$$x \mapsto \exp\left(\alpha\left(1-x^{-g}\right)\right), \quad x \in (0,1)$$

*is $\left(24\alpha g, \frac{1}{g}\right)$-mild.*

*Proof.* We consider the $k$th derivative, where we do not collect any terms,

$$\frac{d^k}{dx^k}\exp\left(\alpha\left(1-x^{-g}\right)\right) = \sum_i \frac{a_i \exp\left(\alpha\left(1-x^{-g}\right)\right)}{x^{(g+1)m_i+n_i}},$$

where $m_i$ is the number of times the exponential factor is differentiated, and $n_i$ is the number of times the $\frac{1}{x}$ factor is differentiated. For each $k$, the greatest power of $\frac{1}{x}$ occurring is $(g+1)k$, so the contribution to the coefficient $a_i$ from having repeatedly differentiated the $\frac{1}{x}$ factor is

$$\leq ((g+1)m_i + n_i)^{n_i},$$

and the contribution from differentiating the exponential factor is $(\alpha g)^{m_i}$. So we have

$$\begin{aligned} |a_i| &\leq 2^k \alpha^k g^k k^{n_i}, \\ m_i + n_i &= k. \end{aligned}$$

The maximum of $\frac{\exp\left(\alpha\left(1-x^{-g}\right)\right)}{x^{(g+1)m+n}}$ occurs when

$$\frac{\alpha g}{x^{g+1}} - \frac{(g+1)m+n}{x} = 0,$$

which is when

$$x = \frac{\alpha^{\frac{1}{g}}}{\left(\frac{1}{g}((g+1)m+n)\right)^{\frac{1}{g}}}.$$

So

$$\begin{aligned} \sup_{x\in(0,1)} \left|\frac{\exp\left(\alpha\left(1-x^{-g}\right)\right)}{x^{(g+1)m+n}}\right| &\leq \frac{\left(\frac{1}{g}((g+1)m+n)\right)^{\frac{(g+1)m+n}{g}}}{\alpha^{\frac{(g+1)m+n}{g}}} \\ &\leq 2^k k^{m+\frac{k}{g}} \end{aligned}$$

for some $A$. Finally, as from one term in the sum, taking its derivative produces two terms, and we are not collecting terms, there are $2^{k-1}$ terms, so we have the bound

$$\begin{aligned}\sup_{x\in(0,1)}\left|\frac{d^k}{dx^k}\exp\left(\alpha\left(1-x^{-g}\right)\right)\right| &\leq 8^k\alpha^k g^k k^k k^{\frac{k}{g}}\\ &\leq 24^k\alpha^k g^k k!k^{\frac{k}{g}},\end{aligned}$$

where we have used $k^k \leq 3^k k!$, hence the map is $\left(24\alpha g, \frac{1}{g}\right)$-mild. □

This also extends to the function of multiple variables $x_1, \ldots, x_n$

$$\exp\left(1-\frac{x_{b_1}\cdots x_{b_l}}{x_{b_{l+1}}\cdots x_{b_n}}\right),$$

by considering this a function of the variable $y = \frac{x_{b_1}\cdots x_{b_l}}{x_{b_{l+1}}\cdots x_{b_n}}$, analogously to Lemma 2.7 of [67].

We note that this lemma reduces the estimate for algebraic points of height $T$ in a given number field on $\log x \log y = \log z$ in [67] from $O\left((\log T)^{44+\epsilon}\right)$ to $O\left((\log T)^{35+\epsilon}\right)$, as it is proved there that there are $O\left((\log T)^{9C(1+o(1))+35}\right)$ such points, where $(A, C)$ is the mildness of $\exp(1-x^{-g})$. We now bound the number of algebraic points on $(x, x^\alpha)$.

**Lemma 3.3.** *Let $\alpha$ be a positive irrational number, and $Y \subset \mathbb{R}^2$ be the curve given by $(x, x^\alpha)$, $x > 0$. Then $\#Y(F,T) = O\left((f^2\log T)^{2+o(1)}\right)$.*

*Proof.* We first let $X = \{(x, x^\alpha) | x \in (0,1)\}$. If $(x, y)$ is a point in $Y$ such that $x, y > 1$, then this is witnessed by the point $\left(\frac{1}{x}, \frac{1}{y}\right) \in X$, so up to a factor of two we need only count the points on $X$. If $\alpha < 1$, then we can consider $\alpha' = \alpha + 1$, the algebraic points of which are the same up to a multiple of $x$ in the second coordinate, which only affects our estimates by a constant. By Lemma 3.2,

$$x \mapsto \left(\exp\left(1-x^{-g}\right), \exp\left(\alpha\left(1-x^{-g}\right)\right)\right)$$

is a $\left(24\alpha g, \frac{1}{g}\right)$-mild parameterization of $X$. We let $g = [\log\log T]$. As $\alpha$ is irrational, $X^{\text{alg}} = \emptyset$, so by Theorem 3.3, $X(F,T)$ is contained in

$$O\left(g^{2+o(1)}(f^2\log T)^{\frac{4}{g}+o(1)}\right) = O\left((f^2\log T)^{o(1)}\right)$$

intersections of $X$ with hypersurfaces of degree

$$d = \left[f^2\log T\right],$$

where $c$ is an absolute constant. Parameterizing $X$ by

$$t \mapsto \left(e^{-t}, e^{-\alpha t}\right),\ t \in (0, \infty),$$

as $\alpha$ is irrational, $e^{-t}$ and $e^{-\alpha t}$ are algebraically independent, so by 83.1 of [71], the number of real zeros of $P(e^{-t}, e^{-\alpha t})$ is at most $d^2$, so

$$\#X(F,T) = O\left((f^2 \log T)^{2+o(1)} g^{2+o(1)}\right).$$

And as $g = [\log \log T]$, we obtain

$$\#X(F,T) = O\left((f^2 \log T)^{2+o(1)}\right).$$

□

The proof of Theorem 3.2 then proceeds in the usual manner,

*Proof of Theorem 3.2.* As $x_1$ and $x_2$ are linearly independent over $\mathbb{Q}$, $x_2 = \alpha x_1$ for some real irrational $\alpha$, which may be taken to be positive without loss of generality. So $x_2 y_i = \alpha x_1 y_i$ for $1 \leq i \leq 3$. By the Lindemann–Weierstrass theorem, as $y_1, y_2, y_3$ are linearly independent over $\mathbb{Q}$, $e^{x_1 y_1}, e^{x_1 y_2}, e^{x_1 y_3}$ are multiplicatively independent, so the points $e^{n_1 x_1 y_1 + n_2 x_1 y_2 + n_3 x_1 y_3}$ are all distinct. If these are all algebraic, then up to height $T$ there are $\Omega((\log T)^3)$ such points. We suppose that $e^{x_1 y_1}$, $e^{\alpha x_1 y_1}$, $e^{x_1 y_2}$, $e^{\alpha x_1 y_2}$, $e^{x_1 y_3}$, $e^{\alpha x_1 y_3}$ are all algebraic. As the points

$$\left(e^{n_1 x_1 y_1 + n_2 x_1 y_2 + n_3 x_1 y_3}, e^{\alpha(n_1 x_1 y_1 + n_2 x_1 y_2 + n_3 x_1 y_3)}\right)$$

lie on the curve $Y = \{(x, x^\alpha) | x > 0\}$, taking

$$F = \mathbb{Q}\left(e^{x_1 y_1}, e^{x_1 y_2}, e^{x_1 y_3}, e^{x_2 y_1}, e^{x_2 y_2}, e^{x_2 y_3}\right),$$

$\#Y(F,T) = \Omega\left((\log T)^3\right)$. By Lemma 3.3, $\#Y(F,T) = O\left((f^2 \log T)^{2+o(1)}\right)$, which is violated by our lower bound for sufficiently large $T$. Hence at least one of $e^{x_1 y_1}$, $e^{\alpha x_1 y_1}$, $e^{x_1 y_2}$, $e^{\alpha x_1 y_2}$, $e^{x_1 y_3}$, $e^{\alpha x_1 y_3}$ must be transcendental. □

*Remark.* To prove the *four* exponentials conjecture in this manner, one would need a stronger bound of $O((\log T)^{2-\epsilon})$, which does not seem possible by the usual transcendence methods, nor a mildness argument. The methods of interpolation determinants and Siegel's lemma type arguments give a bound of $O((\log T)^2)$, and we note that one may prove the same by the method of mild parameterization if one adapts Pila's proof of Theorem 3.3 to this specific curve.

# Chapter 4

# Inversion of the $j$-function and testing complex multiplication

In this chapter we develop an algorithm to invert the $j$-function in quasilinear time, and give an application to testing whether an elliptic curve has complex multiplication.

Previous effective methods of calculating the inverse of $j$ generally pass through an elliptic curve representation — one can consider for example the elliptic curve

$$y^2 = x^3 - \frac{27j}{j-1728}x - \frac{27j}{j-1728},$$

which has $j$-invariant $j$, and apply an algorithm to determine its two periods. In the real case, the algorithm of [24], pg. 391 computes the periods via an arithmetic geometric mean, and in the complex case, there are methods based on the complex arithmetic-geometric mean of Cremona and Thongjunthug [27], or expressing the periods as elliptic integrals, which may be computed by the methods of Dupont [31].

However, it is considered of interest to study $j$ *without* passing through the elliptic curve representation, for example in the problem of Schneider — to prove directly that $j(\alpha)$ is transcendental for algebraic $\alpha$ of degree $\geq 3$. A previous algorithm of Alwaise [3] inverts $j$ directly by different methods to ours, but does not analyze the running time of their algorithm. The time complexity of our algorithm is essentially optimal, and of the same complexity as that of Cremona and Thongjunthug's and Dupont's algorithms.

Our method is similar to that of Labrande [50] for calculating Jacobi theta functions, which makes use of addition formulae between Jacobi's theta functions in order to reduce the argument to a fixed compact set. Here we repeatedly make use of the

modular polynomial

$$\Phi_2(X,Y) = X^3 + Y^3 - X^2Y^2 + 1488X^2Y + 1488XY^2 - 162000X^2 - 162000Y^2 \\ +40773375XY + 8748000000X + 8748000000Y - 157464000000000,$$

which has the property that the roots of $\Phi_2(j(\tau), z)$ are $j(2\tau)$, $j\left(\frac{\tau}{2}\right)$, and $j\left(\frac{\tau+1}{2}\right)$, to either compute $j(2^k\tau)$, the logarithm of which is a close approximation to $-2^{k+1}\pi\tau$ when $2^k\tau$ is sufficiently large, or to manipulate $\tau$ into a compact set to which Newton's method may be applied.

We define the *regulated precision* of an approximation $\tilde{\alpha}$ to $\alpha$ to be

$$\frac{|\alpha - \tilde{\alpha}|}{\max\{1, |\alpha|\}},$$

and denote by $M(P)$ the computational complexity of multiplication of two $P$-bit integers, which by a recent result of Harvey and Hoeven [38] may be taken to be $O(P \log P)$. We obtain the following theorem.

**Theorem 4.1.** *Suppose that $\tilde{j}$ is an approximation to $j(\tau)$, $\tau \in \mathcal{F}$, of regulated precision $2^{-P}$, with $P \geq 400$. Let $Q = P/6$ if $|\tau - i| \leq 2^{-30}$ or $\left|\tau - \frac{\pm 1 + i\sqrt{3}}{2}\right| \leq 2^{-30}$, and $Q = P - \max\{11 \log P, 100\}$ otherwise. Then we may obtain an approximation to $\tau$ of relative precision $2^{-Q}$ in time*

$$O(M(P) \log P).$$

We note that the $j$-function has two ramification points in its fundamental domain, which entail an unavoidable loss of precision in its inversion when $j$ is close to 0 or 1728, and that our algorithm is currently unimplemented.

We apply this algorithm to test for complex multiplication of elliptic curves. Given an approximation to the $j$-invariant of an elliptic curve and a bound upon its height and degree, we may invert it and determine if the inverse is a quadratic irrational, determining also the discriminant.

**Theorem 4.2.** *Suppose that $j$ is the $j$-invariant of an elliptic curve $E$, with $j$ of degree bounded by $d$ and absolute multiplicative height bounded by $H \geq e^e$. Then it may be determined from $d$, $H$, and an approximation to $j$ of regulated precision $2^{-300d^2 \log H(\log d + \log\log H)^2 - 200}$ whether $E$ has complex multiplication. If so, the associated discriminant may be determined in time*

$$O(M(T) \log T),$$

*where $T = d^2 \log H(\log d + \log\log H)^2$.*

Previous methods include that of Achter [2], based on reduction of elliptic curves at primes, which has an unconditional running time of $O(H^{cd})$, and assuming the Generalized Riemann Hypothesis a running time of $O(d^2(\log H)^2)$. Charles [22] gave the first unconditional polynomial time algorithm, based on Galois representations associated to torsion points, with running time $O(d^{c_1}(\log H)^{c_2})$, with an ineffective implicit constant. He also gave a probabilistic algorithm, also of polynomial running time. We note that our algorithm is currently unimplemented.

One may also apply our algorithm for the inversion of $j$ to detecting isogenies between two elliptic curves. This would have a running time $O(N \log N \log\log N)$, however the algorithm requires explicit bounds on the coefficients of modular polynomials $\Phi_N(X, Y)$, which are currently only available for isogenies of *prime* degree, due to Bröker and Sutherland [20].

## 4.1 Preliminaries

We will denote by $\mathcal{F}$ the usual fundamental domain of $j(z)$,

$$\mathcal{F} = \left\{ z \Big| -\frac{1}{2} < \mathrm{Re}(z) \leq \frac{1}{2}, |z| > 1 \right\} \cup \left\{ z \Big| 0 \leq \mathrm{Re}(z) \leq \frac{1}{2}, |z| = 1 \right\},$$

Throughout we will make use of the following results.

**Lemma 4.1** (Lemma 1 of [13]). *If $\tau \in \mathcal{F}$ then*

$$\left| j(\tau) - e^{-2\pi i \tau} \right| \leq 2079.$$

**Theorem 4.3** (Kantorovich, [47]). *Let $F : S(x_0, R) \subset X \to Y$ have a continuous Fréchet derivative in $\overline{S(x_0, r)}$. Moreover, let* (i) *the linear operation $\Gamma_0 = [F'(x_0)]^{-1}$ exist;* (ii) $\|\Gamma_0 F(x_0)\| \leq \eta$; (iii) $\|\Gamma_0 F''(x)\| \leq K$ $(x \in \overline{S(x_0, r)})$. *Now, if*

$$h = K\eta \leq \frac{1}{2}$$

*and*

$$r \geq \frac{1 - \sqrt{1 - 2h}}{h}\eta,$$

*then $F(x) = 0$ will have a solution $x^*$ to which the Newton method is convergent. Here,*

$$\|x^* - x_0\| \leq r.$$

*Furthermore, if for $h < \frac{1}{2}$,*

$$r < r_1 = \frac{1 + \sqrt{1 - 2h}}{h}\eta,$$

*or for* $h = \frac{1}{2}$

$$r \leq r_1,$$

*the solution* $x^*$ *will be unique in the sphere* $\overline{S(x_0, r)}$. *The speed of convergence is characterized by the inequality*

$$\|x^* - x_k\| \leq \frac{1}{2^k}(2h)^{2^k}\frac{\eta}{h}$$

*for* $k = 0, 1, 2, \ldots$.

We note that the condition

$$r \geq \frac{1 - \sqrt{1 - 2h}}{h}\eta$$

may be replaced with

$$r \geq 2\eta.$$

Finally, we will make use of the fact that in applying Newton's method, or the secant method, in order to obtain an approximation of precision $P$ bits, from a starting point which ensures convergence, if the computational complexity of evaluating the function in question to precision $P$ bits is superlinear, $O(f(P))$ say, then Newton's method, or the secant method, despite requiring $O(\log P)$ steps, may be performed in time $O(f(P))$, as opposed to $O(f(P)\log P)$. This is because at each step we may evaluate our function to only the precision required for the continuation of Newton's method, as opposed to full precision. Indeed, at each step the number we are calculating is a lower precision approximation to the final result. Details of Newton's method and this variable-precision approach may be found in [17], section 6.4.

## 4.2 Inversion of $j(z)$

Firstly, if $|j| \leq 2^{-P/2}$ or $|j - 1728| \leq 2^{-P/3}$, we will return $\tau = \frac{1+i\sqrt{3}}{2}$ or $\tau = i$ respectively. Otherwise, we split the fundamental domain of $j$ into 5 sections: $\mathrm{Im}(\tau) \geq 3$, $|\tau - i| \leq 2^{-31}$, $\left|\tau - \frac{1+\sqrt{3}}{2}\right| \leq 2^{-31}$, $\left|\tau - \frac{-1+\sqrt{3}}{2}\right| \leq 2^{-31}$ and the remaining compact subset of the fundamental domain. We may determine $\tau$ with sufficient precision by taking a low precision inverse via the expression for $j^{-1}$ in terms of Gaussian hypergeometric functions, as in Chapter 2. Letting $\alpha$ be a solution to

$$j(\tau) = \frac{1728}{4\alpha(1-\alpha)},$$

$\tau$ is equal to either

$$i\frac{{}_2F_1\left(\frac{1}{6},\frac{5}{6},1;1-\alpha\right)}{{}_2F_1\left(\frac{1}{6},\frac{5}{6},1;\alpha\right)}.$$

or the negative of its inverse. If $j$ is sufficiently large, or close to 0 or 1728, then we do not need to evaluate this in order to determine which section of the fundamental domain $\tau$ lies in, so we need only compute the above to a fixed precision in a compact set. As these points are bounded away from the zeros of ${}_2F_1\left(\frac{1}{6},\frac{5}{6},1;z\right)$, this takes only constant time.

### 4.2.1 Large $j$

Throughout this section we assume $\mathrm{Im}(\tau) \geq 3$, and will repeatedly make use of the consequent fact that $|j(\tau)| \geq 10^8$. We will make use of the modular polynomial $\Phi_2(X,Y)$ to obtain an approximation to $j(2\tau)$, and repeat the process until we have an approximation to $j(2^k\tau)$, where $k$ is sufficiently large. At this point taking the logarithm of $j(2^k\tau)$ gives a good approximation to $-2^{k+1}\pi\tau$, owing to the dominating $q^{-1}$ term of the $q$-series of $j$.

**Proposition 4.1.** *Let $j(\tau) = j$, and suppose that $\mathrm{Im}(\tau) \geq 3$ and $\tilde{j}$ is an approximation to $j$ of relative precision at least $2^{-P}$, with $P \geq 300$. Then the largest root, in absolute value, of $\Phi_2(\tilde{j},z)$ is an approximation to $j(2\tau)$ of relative precision at least $2^{-P+2}$.*

*Proof.* Let $f(z) = \Phi_2(j,z)$ and $g(z) = \Phi_2(\tilde{j},z)$. We first bound the coefficients of $f(z) - g(z)$. For the coefficient of $z^2$, we have, with $\tilde{j} = j + \delta$,

$$\begin{aligned}\left|-j^2 + 1488j - 162000 - (-(j+\delta)^2 + 1488(j+\delta) - 162000)\right| &= |2\delta j + \delta^2 + 1488\delta| \\ &\leq 2.1|\delta||j|,\end{aligned}$$

as $|\delta| \leq 2^{-300}|j|$ and $|j| \geq 10^8$. For the coefficient of $z$, we have

$$\begin{aligned}&\big|1488j^2 + 40773375j + 8748000000 \\ &\qquad - (1488(j+\delta)^2 + 40773375(j+\delta) + 8748000000)\big| \\ &= |2976\delta j + 1488\delta^2 + 40773375\delta| \\ &\leq 3210|\delta||j|.\end{aligned}$$

For the constant coefficient, we have

$$\begin{aligned}&\left|3\delta j^2 + 3\delta^2 j + \delta^3 - 16200\delta j - 16200\delta^2 + 8748000000\delta\right| \\ &\qquad \leq 3.4|\delta||j|^2.\end{aligned}$$

We now bound the values of $g(j(\tau'))$ for $\tau' \in \left\{2\tau, \frac{\tau}{2}, \frac{\tau+1}{2}\right\}$. For $j(2\tau)$, as $|j(2\tau)| \leq 1.02|j|^2$,

$$\begin{aligned}|g(j(2\tau))| &= |g(j(2\tau)) - f(j(2\tau))| \\ &\leq 2.1|\delta||j|(1.02|j|^2)^2 + 3210|\delta||j| \cdot 1.02|j|^2 + 3.4|\delta||j|^2 \\ &\leq 2.2|\delta||j|^5.\end{aligned}$$

For $g(j(\tau'))$, $\tau' \in \left\{\frac{\tau}{2}, \frac{\tau+1}{2}\right\}$, firstly, the absolute values of $j\left(\frac{\tau}{2}\right)$ and $j\left(\frac{\tau+1}{2}\right)$ are bounded above and below by $e^{2\pi \mathrm{Im}(\tau)/2} \pm 2079$, and the absolute value of $j$ is bounded above and below by $e^{2\pi \mathrm{Im}(\tau)} \pm 2079$. As $\mathrm{Im}(\tau) \geq 3$, these yield

$$0.83|j|^{1/2} \leq \left|j\left(\frac{\tau+1}{2}\right)\right|, \left|j\left(\frac{\tau}{2}\right)\right| \leq 1.17|j|^{1/2}, \tag{4.1}$$

so

$$\begin{aligned}|g(j(\tau'))| &= |g(j(\tau')) - f(j(\tau'))| \\ &\leq 2.1|\delta||j|(1.17|j|^{1/2})^2 + 3210|\delta||j|(1.17|j|^{1/2}) + 3.4|\delta||j|^2 \\ &\leq 20|\delta||j|^2.\end{aligned}$$

Let $\beta_0, \beta_1, \beta_2$ be the roots of $g(z)$. Then for $\beta_i$ the closest root of $g$ to $j(2\tau)$,

$$|j(2\tau) - \beta_i| \leq \left(2.2|\delta||j|^5\right)^{1/3} \leq \left(2.2 \cdot 2^{-300}|j|^6\right)^{1/3} \leq 10^{-29}|j|^2,$$

and we let $\beta_0$ be the closest root of $g$ to $j(2\tau)$. As $0.98|j|^2 \leq j(2\tau) \leq 1.02|j|^2$, $0.97|j|^2 \leq |\beta_0| \leq 1.03|j|^2$. For $\tau' = \frac{\tau}{2}$, with $\beta_i$ the closest root of $g$ to $j(\tau')$,

$$|j(\tau') - \beta_i| \leq \left(20|\delta||j|^2\right)^{1/3} \leq \left(20 \cdot 2^{-300}|j|^3\right)^{1/3} \leq 10^{-29}|j|,$$

and in particular,

$$|\beta_i| \leq 10^{-29}|j| + 1.17|j|^{1/2} < 0.97|j|^2 \leq |\beta_0|, \tag{4.2}$$

so $\beta_i \neq \beta_0$. Let this root of $g$ be $\beta_1$. Now we may improve the bound on $|j(\tau') - \beta_1|$,

$$|j(\tau') - \beta_1|^2 \leq \frac{20|\delta||j|^2}{|j(\tau') - \beta_0|} \leq \frac{20|\delta||j|^2}{0.98|j|^2 - 1.17|j|^{1/2}} \leq 30 \cdot 2^{-300}|j|,$$

so

$$|j(\tau') - \beta_1| \leq 10^{-88}|j|^{1/2},$$

and consequently

$$0.82|j|^{1/2} \leq |\beta_1| \leq 1.18|j|^{1/2}.$$

We now bound $\beta_2$. By our bound on the difference between the constant coefficients of $f$ and $g$, the constant coefficient $-\beta_0\beta_1\beta_2$ of $g$ is bounded in absolute value by

$$\left| j(2\tau) j\left(\frac{\tau}{2}\right) j\left(\frac{\tau+1}{2}\right) \right| + 3.4\delta|j|^2 \leq 1.4|j|^3 + 0.01|j|^3 \leq 1.5|j|^3,$$

so that

$$|\beta_2| \leq \frac{1.5|j|^3}{0.97|j|^2 \cdot 0.82|j|^{1/2}} \leq 2|j|^{1/2}.$$

Returning to $g(j(2\tau))$, we now bound below the terms $|j(2\tau)-\beta_i|$ for $i = 1, 2$ in order to improve our inequality for $|j(2\tau) - \beta_0|$. We now have, for $i = 1, 2$,

$$\begin{aligned} |j(2\tau) - \beta_i| &\geq |j(2\tau)| - 2|j|^{1/2} \\ &\geq 0.97|j|^2. \end{aligned}$$

This now improves our bound on the difference of $\beta_0$ to $j(2\tau)$,

$$\begin{aligned} |j(2\tau) - \beta_0| &\leq \frac{2.2\delta|j|^5}{|j(2\tau) - \beta_1||j(2\tau) - \beta_2|} \\ &\leq \frac{2.2\delta|j|^5}{(0.97|j|^2)^2} \\ &\leq 2.4\delta|j|. \end{aligned}$$

So the relative precision of $\beta_0$ as an approximation to $j(2\tau)$ is bounded by

$$\frac{|j(2\tau) - \beta_0|}{|j(2\tau)|} \leq \frac{2.4|\delta||j|}{0.97|j|^2} \leq 2.5\frac{|\delta|}{|j|} \leq 2^{-P+2}.$$

□

**Proposition 4.2.** *If* $\mathrm{Im}(\tau) \geq 3$ *and* $\tilde{j}$ *is an approximation to* $j(\tau)$ *of relative precision* $2^{-P}$*, where* $P$ *is at least* 300*, applying Newton's method to* $\Phi_2(\tilde{j}, z)$*, with starting point*

$$\tilde{j}^2 - 2 \cdot 744\tilde{j} - 2 \cdot 196884 + 744^2 + 744,$$

*will obtain an approximation to* $j(2\tau)$ *of relative precision* $2^{-P+3}$ *after at most* $[2 \log P]$ *steps of Newton iteration.*

*Proof.* We first give a rough approximation to $\beta_0$. Writing

$$j(\tau) = e^{-2\pi i\tau} + 744 + 196884e^{2\pi i\tau} + f(\tau),$$

we have

$$\begin{aligned} j(\tau)^2 - 2 \cdot 744j(\tau) + 2 \cdot 19688 + 2 \cdot 744^2 + 744 \\ = e^{-4\pi i\tau} + 744 + 196884^2 e^{4\pi i\tau} \\ + 2(196884e^{2\pi i\tau} + e^{-2\pi i\tau})f(\tau) + f(\tau)^2. \end{aligned}$$

As $\mathrm{Im}(\tau) \geq 3$, $f$ is maximized when $\mathrm{Re}(\tau) = 0$, and $f(\mathrm{Im}(\tau))$ is decreasing in $\mathrm{Im}(\tau)$,

$$\begin{aligned}
&|j(2\tau) - (j(\tau)^2 - 744j(\tau) + 19688 - 744^2 + 744)| \\
&\quad \leq (196884 + 196884^2)e^{-4\pi\mathrm{Im}(\tau)} + |f(\mathrm{Im}(2\tau))| \\
&\qquad + 2(196884e^{2\pi\mathrm{Im}(\tau)} + e^{2\pi\mathrm{Im}(\tau)})|f(\mathrm{Im}(\tau))| + |f(\mathrm{Im}(\tau))|^2 \\
&\quad \leq 10^{-7} + |f(6)| + 10^6|f(3)| + |f(3)|^2 \\
&\quad \leq 0.4
\end{aligned}$$

We now let

$$\gamma_0 = \tilde{j}^2 - 744\tilde{j} + 19688 - 744^2 + 744.$$

This will be our starting point for Newton iteration to find $\beta_0$. We now bound the terms appearing in Kantorovich's theorem. Firstly,

$$\begin{aligned}
\Phi_2(\tilde{j}, z) = z^3 &+ (-\tilde{j}^2 + 1488\tilde{j} - 162000)z^2 \\
&+ (1488\tilde{j}^2 + 40773375\tilde{j} + 8748000000)z \\
&+ \tilde{j}^3 - 162000\tilde{j}^2 + 8748000000\tilde{j} - 157464000000000.
\end{aligned}$$

Now it is clear, as $|j - \tilde{j}| \leq 2^{-300}|j|$, that $|\gamma_0 - j(2\tau)| \leq 0.4 + \frac{|j|^2}{2^{90}}$, and so as $0.98|j|^2 \leq |j(2\tau)| \leq 1.02|j|^2$, we have $0.979|j|^2 \leq |\gamma_0| \leq 1.021|j|^2$. We will take the $r$ of Kantorovich's theorem to be $0.009|j|^2$, and give bounds for the disc $|\gamma - \gamma_0| \leq 0.009|j|^2$. For $\gamma$ in this disc, $0.97|j|^2 \leq |\gamma| \leq 1.03|j|^2$ and in addition $|\tilde{j}| \leq 1.01|j|$, so for an upper bound on the first derivative, we have

$$\begin{aligned}
|\Phi_2'(\tilde{j}, \gamma)| &\leq 3|\gamma|^2 + 2(|\tilde{j}|^2 + 1488|\tilde{j}| + 162000)|\gamma_0| \\
&\quad + (1488|\tilde{j}|^2 + 40773375|\tilde{j}| + 8748000000) \\
&\leq 3.2|j|^4 + 2.2|j|^4 + 10^{-7}|j|^4 \\
&\leq 5.5|j|^4,
\end{aligned}$$

and for a lower bound, we have

$$\begin{aligned}
|\Phi_2'(\tilde{j}, \gamma)| &\geq 3|\gamma|^2 - 2(|\tilde{j}|^2 + 1488|\tilde{j}| + 162000)|\gamma_0| \\
&\quad - (1488|\tilde{j}|^2 + 40773375|\tilde{j}| + 8748000000) \\
&\geq 2.82|j|^4 - 2.1|j|^4 - 10^{-7}|j|^4 \\
&\geq 0.71|j|^4.
\end{aligned}$$

For our upper bound on the function at $\gamma_0$, as $|\beta_1|, |\beta_2| \leq 2|j|^{1/2}$, and as $|\beta_0 - j(2\tau)| \leq 2^{-298}|j|^2$ we have $|\gamma_0 - \beta_0| \leq 0.4 + 2^{-298}|j|^2$, so

$$\begin{aligned}|\Phi_2(\tilde{j}, \gamma_0)| &= |\gamma_0 - \beta_0||\gamma_0 - \beta_1||\gamma_0 - \beta_2| \\ &\leq (0.4 + 2^{-289}|j|^2)(1.03|j|^2 + 2|j|^{1/2})^2 \\ &\leq 2^{-54}|j|^6.\end{aligned}$$

For the second derivative, we have for an upper bound

$$\begin{aligned}|\Phi_2''(\tilde{j}, \gamma)| &\leq 6|\gamma| + 2|\tilde{j}|^2 + 1488|\tilde{j}| + 162000 \\ &\leq 8.3|j|^2,\end{aligned}$$

and for a lower bound,

$$\begin{aligned}|\Phi_2''(\tilde{j}, \gamma_0)| &\geq 6|\gamma| - 2|\tilde{j}|^2 - 1488|\tilde{j}| - 162000 \\ &\geq 3.8|j|^2.\end{aligned}$$

Now, by Theorem 4.3, as

$$\frac{|\Phi_2(\tilde{j}, \gamma_0)||\Phi_2''(\tilde{j}, \gamma)|}{|\Phi_2'(\tilde{j}, \gamma)|^2} \leq \frac{2^{-54}|j|^6 \cdot 8.3|j|^2}{(0.71|j|^4)^2} \leq 2^{-49} < \frac{1}{2},$$

and

$$\begin{aligned}2\eta = 2\frac{|\Phi_2(\tilde{j}, \gamma_0)|}{|\Phi_2'(\tilde{j}, \gamma_0)|} &\leq \frac{2^{-54}|j|^6}{0.71|j|^4} \leq 0.001|j|^2, \\ r &= 0.009|j|^2,\end{aligned}$$

Newton's method will converge to the root $\beta_0$, with a rate of convergence

$$|\gamma_k - \beta_0| \leq \frac{1}{2^k} 2^{-48 \cdot 2^k} \frac{|\Phi_2'(\tilde{j}, \gamma_0)|}{|\Phi_2''(\tilde{j}, \gamma_0)|} \leq \frac{1}{2^k} 2^{-48 \cdot 2^k} \frac{5.5|j|^4}{3.8|j|^2} \leq 2^{-48 \cdot 2^k}|j|^2$$

for $k \geq 1$. In particular, as $|\beta_0| \geq 0.97|j|^2$, when $k \geq [2 \log P]$, $\frac{|\gamma_k - \beta_0|}{|\beta_0|} \leq 2^{-P}$. Now as, by Proposition 4.1, $\beta_0$ is an approximation to $j(2\tau)$ of relative precision $2^{-P+2}$, after $[2 \log P]$ steps, $\gamma_k$ will be an approximation to $j(2\tau)$ of relative precision $2^{-P+3}$. $\square$

**Lemma 4.2.** *Suppose that $|j(\tau)| \geq 2^{P+12}$, and $\tilde{j}$ is an approximation to $j(\tau)$ of relative precision at least $2^{-P}$. Then*

$$-\frac{\log \tilde{j}}{2\pi}$$

*is an approximation to $\tau$ of absolute precision $2^{-P}$.*

*Proof.* Let $j(\tau) + \delta = \tilde{j}$. As

$$|j(\tau) - e^{-2\pi i\tau}| \leq 2079,$$

we have

$$\log(\tilde{j}) = -2\pi\tau + \log\left(1 + \frac{\delta + 2079\theta}{j(\tau)}\right),$$

where $|\theta| \leq 1$. So as $|\log(1+z)| \leq 1.1|z|$ when $|z| \leq 0.05$,

$$|\log \tilde{j} + 2\pi\tau| \leq 1.1 \cdot 2^{-P} + 0.6 \cdot 2^{-P} \leq 1.7 \cdot 2^{-P}.$$

So that

$$\left|\tau - \left(-\frac{\log \tilde{j}}{2\pi}\right)\right| \leq \frac{1.7}{2\pi} 2^{-P} \leq 2^{-P}.$$

□

Now the algorithm to invert $j$ when $\mathrm{Im}(\tau) \geq 3$ proceeds as follows: if $|j| \leq 2^{P+12}$, first iteratively compute approximations to $j(2^k\tau)$ by Newton's method applied to $\Phi_2(\tilde{j}, z)$, up to $k = \left\lceil 2\log\left(\frac{P+12}{\mathrm{Im}(\tau)}\right)\right\rceil + 1$. Then calculate the logarithm of $j(2^k\tau)$ to relative precision $2^{-P-2}$ (note that $|\tau| \geq 1$), and divide by $-2\pi$, where we have calculated $2\pi$ to relative precision $2^{-P-2}$. This process entails a loss of precision of at most $5\left\lceil 2\log\left(\frac{P+12}{\mathrm{Im}(\tau)}\right)\right\rceil + 2$, which is bounded by $11\log P$ when $P \geq 400$. The precision at all applications of Newton's method is at least $2^{-300}$, so our assumptions on the precision of our approximations in the propositions of this section are satisfied at each application.

The computational complexity of Newton's method applied at each step to these polynomials is $O(M(P))$, as at each step of Newton's method one may use only the required precision to obtain the next approximant, so that only the last step is performed at full precision.

The computational complexity of the complex logarithm and computing $\pi$ are $O(M(P)\log P)$, so if $|j| \leq 2^{P+12}$, the algorithm has time complexity

$$O\left(M(P)\left[2\log\left(\frac{P+12}{\mathrm{Im}(\tau)}\right) + 1\right]\right) = O\left(M(P)\log P\right),$$

and if $|j| \geq 2^{P+12}$, time complexity

$$O\left(M(P)\log P\right),$$

where the implicit constants are not too large and could be made effective.

### 4.2.2 Near 1728 and 0

When $\tau$ is close to either $i$ or $\frac{\pm 1+i\sqrt{3}}{2}$, we will make use of the modular polynomial $\Phi_2(X,Y)$ to obtain an approximation to one of $j(2\tau)$, $j\left(\frac{\tau}{2}\right)$, or $j\left(\frac{\tau+1}{2}\right)$. The $\mathrm{SL}_2(\mathbb{Z})$-equivalent elements of $\mathcal{F}$ to either $2\tau$, $\frac{\tau}{2}$, or $\frac{\tau+1}{2}$ will lie in the aforementioned compact set, to which we may apply Newton's method. We carry out the analysis only for $i$ and $\frac{1+i\sqrt{3}}{2}$, as when $\left|\tau - \frac{-1+i\sqrt{3}}{2}\right| \leq 2^{-31}$, the $\mathrm{SL}_2(\mathbb{Z})$-equivalent $\tau+1$ satisfies $\left|(\tau+1) - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-31}$.

**Lemma 4.3.** *If* $|\delta| \leq 2^{-30}$ *then*

$$\left| j(i+\delta) - 1728 - \frac{j^{(2)}(i)}{2}\delta^2 \right| \leq 0.07|\delta|^2,$$

*and*

$$\left| j\left(\frac{1+i\sqrt{3}}{2} + \delta\right) - \frac{j^{(3)}\left(\frac{1+i\sqrt{3}}{2}\right)}{3!}\delta^3 \right| \leq 0.07|\delta|^3.$$

*Proof.* We first bound the coefficients of the Taylor series of $j$ at $z = i$ and $z = \frac{1+i\sqrt{3}}{2}$. Firstly, by Theorem 1 of [19], the coefficient of $e^{2\pi i n\tau}$ in the $q$-expansion of $j$ is bounded by $4e^{4\pi\sqrt{n}}$, so the corresponding coefficient of the $k$'th derivative is bounded by $(2\pi n)^k e^{4\pi\sqrt{n}}$. Further, for $n \geq 1$, $e^{4\pi\sqrt{n}-\sqrt{3}\pi n} \leq 100e^{-\pi n}$, so we have the following bound on the derivatives at these two points,

$$\left|j^{(k)}(i)\right|, \left|j^{(k)}\left(\frac{1+i\sqrt{3}}{2}\right)\right| \leq (2\pi)^k e^{2\pi} + 744 + 400\sum_{n=1}^{\infty}(2\pi n)^k e^{-\pi n}.$$

We now bound the sum in this expression. Firstly, we have the identity

$$\sum_{n=1}^{\infty} e^{-\pi n x} = \frac{1}{e^{\pi x} - 1},$$

and so consider the derivatives of this function, expressing the derivative as sums of the form

$$\sum \alpha_i \frac{e^{a\pi x}}{(e^{\pi x}-1)^b},$$

where each term occurring in the expression for the $k$'th derivative is derived from taking the derivative of either the numerator or denominator of a term occurring in the $k-1$'th derivative, i.e. there is no collection of terms with $(a,b)$ equal. It is clear that there are at most $2^k$ terms, and, passing from one derivative to the next,

$|\alpha_i|$ may increase by at most $\pi \max\{a, b\}$, and that $a \leq b \leq k+1$, and $a \leq k$. So $|\alpha_i| \leq \pi^k (k+1)!$, and a bound for the whole expression evaluated at 1 is therefore

$$2^k \pi^k (k+1)! \frac{e^{k\pi}}{(e^\pi - 1)^k} \leq 14^k k!.$$

Now we have the bounds

$$\begin{aligned} |j^{(k)}(i)|, \left| j^{(k)} \left( \frac{1+i\sqrt{3}}{2} \right) \right| &\leq (2\pi)^k e^{2\pi} + 744 + 400 \cdot 14^k k! \\ &\leq 1700 \cdot 14^k k!. \end{aligned}$$

At $z = i$ and $z = \frac{1+i\sqrt{3}}{2}$, the Taylor series expansions for $j(z)$ are

$$\begin{aligned} j(z) &= 1728 + \sum_{n=2}^{\infty} c_n (z-i)^n, \\ j(z) &= \sum_{n=3}^{\infty} c'_n \left( z - \frac{1+i\sqrt{3}}{2} \right)^n. \end{aligned}$$

where $c_n$ and $c'_n$ are bounded in absolute value by $1400 \cdot 14^k$. We now bound the tails of the sums of the Taylor series,

$$\sum_{n=3}^{\infty} |c_n| |\delta|^n \leq |\delta|^3 \sum_{n=0}^{\infty} 1700 \cdot 14^{n+3} |\delta|^n,$$

and

$$\sum_{n=4}^{\infty} |c'_n| |\delta|^n \leq |\delta|^4 \sum_{n=0}^{\infty} 1700 \cdot 14^{n+4} |\delta|^n.$$

Now as $|\delta| \leq 2^{-30}$,

$$\sum_{n=0}^{\infty} 1700 \cdot 14^{n+4} |\delta|^n = \frac{1700 \cdot 14^4}{1 - 14|\delta|} \leq 7 \cdot 10^7,$$

so that

$$\begin{aligned} \left| j(i+\delta) - 1728 - \frac{j^{(2)}(i)}{2} \delta^2 \right| &\leq 2^{-30} \cdot 7 \cdot 10^7 |\delta|^2 \leq 0.07 |\delta|^2, \\ \left| j\left( \frac{1+i\sqrt{3}}{2} + \delta \right) - \frac{j^{(3)}\left( \frac{1+i\sqrt{3}}{2} \right)}{3!} \delta^3 \right| &\leq 2^{-30} \cdot 7 \cdot 10^7 |\delta|^3 \leq 0.07 |\delta|^3. \end{aligned}$$

□

Furthermore, we note that

$$\begin{aligned} 49600 &\leq |j^{(2)}(i)| \leq 49700, \\ 274000 &\leq \left| j^{(3)}\left(\frac{1+i\sqrt{3}}{2}\right)\right| \leq 275000, \end{aligned}$$

so if $|\tau - i| \leq 2^{-30}$, by the previous lemma, we have

$$2.4 \cdot 10^4 |\delta|^2 \leq |j - 1728| \leq 2.5 \cdot 10^4 |\delta|,$$

and if $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-30}$,

$$4.5 \cdot 10^4 |\delta|^3 \leq |j| \leq 4.6 \cdot 10^4 |\delta|^3,$$

from which we deduce the following lemma.

**Lemma 4.4.** *If $|\tau - i| \leq 2^{-30}$ and $P \geq 300$, then:*

1. *If $|j(\tau) - 1728| \leq 2^{-P}$, then $|\tau - i| \leq 2^{-P/2-7}$.*
2. *If $|j(\tau) - 1728| \geq 2^{-P}$, then $|\tau - i| \geq 2^{-P/2-8}$.*

*If $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-30}$ and $P \geq 300$, then:*

1. *If $|j| \leq 2^{-P}$, then $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-P/3-5}$.*
2. *If $|j| \geq 2^{-P}$, then $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \geq 2^{-P/3-6}$.*

**Lemma 4.5.** *Suppose that $|\delta| \leq 2^{-28}$. Then*

$$|j(2i+\delta) - j(2i) - j'(2i)\,\delta| \leq 0.2|\delta|,$$

*and*

$$\left|j\left(i\sqrt{3}+\delta\right) - j\left(i\sqrt{3}\right) - j'\left(i\sqrt{3}\right)\delta\right| \leq 0.2|\delta|.$$

*Proof.* We proceed similarly to the previous lemma — as $17e^{-2\pi n} \geq e^{4\pi n - 2\sqrt{3}\pi n}$,

$$|j^{(k)}(2i)|, |j^{(k)}(\sqrt{3}i)| \leq (2\pi)^k e^{4\pi} + 744 + 68\sum_{n=1}^{\infty}(2\pi n)^k e^{-2\pi n},$$

and a similar analysis to the previous lemma bounds the sum in this inequality by

$$(2\pi)^k e^{4\pi} + 744 + 68 \cdot 13^k k! \leq 300000 \cdot 13^k k!. \tag{4.3}$$

So when $|\delta| \leq 2^{-28}$,

$$\begin{aligned}|j\left(2i+\delta\right) - j(2i) - j'\left(2i\right)\delta| &\leq \sum_{n=2}^{\infty} 300000 \cdot 13^n \delta^n \\ &\leq 2^{-28} \cdot 300000 \cdot 13^2 |\delta| \sum_{n=0}^{\infty} 13^n 2^{-28n} \\ &\leq 0.2|\delta|.\end{aligned}$$

and as our bound for the terms of the Taylor series applies to both expansions, we similarly have

$$\left|j\left(i\sqrt{3}+\delta\right) - j\left(i\sqrt{3}\right) - j'\left(i\sqrt{3}\right)\delta\right| \leq 0.2|\delta|.$$

□

We now give two lemmas on the separation and closeness of the three preimages of roots of the modular polynomial $\Phi_2(j(\tau), z)$, for $\tau$ near $i$ and $\frac{1+i\sqrt{3}}{2}$.

**Lemma 4.6.** *Suppose that $\tau \in \mathcal{F}$. Then, if $\tau = i + \delta$,*

$$\begin{aligned}|2\tau - 2i| &\leq 2|\delta|, \\ \left|-\frac{2}{\tau} - 2i\right| &\leq 2|\delta|,\end{aligned}$$

*and if $\tau = \frac{1+i\sqrt{3}}{2} + \delta$,*

$$\begin{aligned}|(2\tau - 1) - \sqrt{3}i| &\leq 2|\delta|, \\ \left|\left(1 - \frac{2}{\tau}\right) - \sqrt{3}i\right| &\leq 2|\delta|, \\ \left|\left(-1 - \frac{2}{\tau - 1}\right) - \sqrt{3}i\right| &\leq 2|\delta|.\end{aligned}$$

*Proof.* First note that as $\tau \in \mathcal{F}$, $|\tau| \geq 1$. Let $2\tau = 2i + \delta_1$, and $-\frac{2}{\tau} = 2i + \delta_2$. For $\delta_1$, $|2\tau - 2i| = 2|\delta|$, and for $\delta_2$,

$$\begin{aligned}|\delta_2| = \left|-\frac{2}{\tau} - 2i\right| &= \left|\frac{-2 - 2i\tau}{\tau}\right| \\ &= \frac{|2i\delta|}{|\tau|} \\ &\leq 2|\delta|.\end{aligned}$$

If $\tau = \frac{1+i\sqrt{3}}{2} + \delta$, let $2\tau - 1 = \sqrt{3}i + \delta_1$, $\left(1 - \frac{2}{\tau}\right) = \sqrt{3}i + \delta_2$, and $\left(1 - \frac{2}{\tau - 1}\right) = \sqrt{3}i + \delta_3$. For $\delta_1$,

$$|\delta_1| = \left|2\tau - 1 - \sqrt{3}i\right| \leq 2|\delta|,$$

for $\delta_2$,

$$\begin{aligned} |\delta_2| = \left|\left(1 - \frac{2}{\tau}\right) - \sqrt{3}i\right| &= \left|\frac{\tau - 2 - \sqrt{3}i\tau}{\tau}\right| \\ &= \left|\frac{(1 - \sqrt{3}i)\delta}{\tau}\right| \\ &\leq 2|\delta|, \end{aligned}$$

and for $\delta_3$,

$$\begin{aligned} |\delta_3| = \left|\left(-1 - \frac{2}{\tau - 1}\right) - \sqrt{3}i\right| &= \left|\frac{-\tau + 1 - 2 - \sqrt{3}i\tau + \sqrt{3}i}{\tau}\right| \\ &= \frac{|(1 + \sqrt{3}i)\delta|}{|\tau - 1|} \\ &\leq 2|\delta|. \end{aligned}$$

□

**Lemma 4.7.** *If $\tau = i + \delta$, $\tau \in \mathcal{F}$, with $|\delta| \leq 2^{-30}$, then the distance between $2\tau$ and $-\frac{2}{\tau}$ is at least $3.99|\delta|$, and the distance between either of these and the point in $\mathcal{F}$ which is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to $\frac{\tau+1}{2}$ is at least $0.99$ in magnitude. If $\tau = \frac{1+i\sqrt{3}}{2} + \delta$, $\tau \in \mathcal{F}$, then the distance between any pair of the three points lying in $\mathcal{F}$ which are $\mathrm{SL}_2(\mathbb{Z})$-equivalent to $2\tau$, $\frac{\tau}{2}$, and $\frac{\tau+1}{2}$ is at least $3.46|\delta|$.*

*Proof.* Firstly, for the distance between $\tau$ and $-\frac{2}{\tau}$, we have

$$\begin{aligned} 2\tau + \frac{2}{\tau} &= \frac{2(\tau^2 + 1)}{\tau} \\ &= \frac{4i\delta + 2\delta^2}{\tau}. \end{aligned}$$

As $|\delta| \leq 2^{-30}$, $|\delta|^2 \leq 2^{-30}|\delta|$, and $|i + \delta| \leq 1 + 2^{-30}$, so

$$\left|2\tau - \left(-\frac{2}{\tau}\right)\right| \geq 3.99|\delta|.$$

Now for $-\frac{2}{\tau+1}+1$, we have

$$\begin{aligned}-\frac{2}{\tau+1}+1-i &= \frac{-2+\tau+1-i\tau-i}{\tau+1}\\ &= \frac{\delta-i\delta}{1+i+\delta}\end{aligned}$$

so that

$$\left|\left(-\frac{2}{\tau+1}+1\right)-i\right| \leq 2|\delta| \tag{4.4}$$

and so either $-\frac{2}{\tau+1}+1$ or $-\left(-\frac{2}{\tau+1}+1\right)^{-1}$ is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to $\frac{1+\tau}{2}$ and in $\mathcal{F}$, and its distance from $i$ is bounded by $\left|1+\frac{2|\delta|}{1-2|\delta|}\right| \leq 2^{-28}$. By Lemma 4.6, the distance of either $2\tau$ or $-\frac{2}{\tau}$ from $2i$ is at most $2|\delta| \leq 2\cdot 2^{-30}$, so both of these are separated from either $-\frac{2}{\tau+1}+1$ or $-\left(-\frac{2}{\tau+1}+1\right)^{-1}$ by a distance of at least 0.99.

Now for $\tau = \frac{1+i\sqrt{3}}{2}+\delta$, the $\mathrm{SL}_2(\mathbb{Z})$-equivalent points to $2\tau$, $\frac{\tau}{2}$, and $\frac{\tau+1}{2}$ are $2\tau-1$, $1-\frac{2}{\tau}$, and $-1-\frac{2}{\tau-1}$ respectively. For their differences, we have,

$$\begin{aligned}2\tau-1-\left(1-\frac{2}{\tau}\right) &= \frac{2(\tau^2-\tau+1)}{\tau}\\ &= \frac{i2\sqrt{3}\delta+2\delta^2}{\tau}.\end{aligned}$$

So as $|\tau| \leq 1+2^{-30}$,

$$\left|2\tau-1-\left(1-\frac{2}{\tau}\right)\right| \geq 3.46|\delta|$$

For the first and third points,

$$\begin{aligned}2\tau-1-\left(-1-\frac{2}{\tau-1}\right) &= \frac{2(\tau^2-\tau+1)}{\tau-1}\\ &= \frac{i2\sqrt{3}\delta+2\delta^2}{\tau-1}.\end{aligned}$$

As $|\tau-1| \leq 1+2^{-30}$, we have the lower bound

$$\left|2\tau-1-\left(-1-\frac{2}{\tau-1}\right)\right| \geq 3.46|\delta|.$$

For the second and third points,

$$\begin{aligned}\left(1-\frac{2}{\tau}\right)-\left(-1-\frac{2}{\tau-1}\right) &= \frac{2(\tau^2-\tau+1)}{\tau(\tau-1)}\\ &= \frac{i2\sqrt{3}\delta+2\delta^2}{\tau(\tau-1)}.\end{aligned}$$

So we obtain the lower bound

$$\left|\left(1-\frac{2}{\tau}\right)-\left(-1-\frac{2}{\tau-1}\right)\right| \geq 3.46|\delta|.$$

□

We now bound above and below the differences between $j$ evaluated at the points $j(2\tau)$, $j\left(\frac{\tau}{2}\right)$, and $j\left(\frac{\tau+1}{2}\right)$.

**Lemma 4.8.** *If $\tau = i + \delta$, with $|\delta| \leq 2^{-30}$, then*

$$7.18 \cdot 10^6 |\delta| \leq \left| j(2\tau) - j\left(\frac{\tau}{2}\right) \right| \leq 7.25 \cdot 10^6 |\delta|,$$

*and*

$$2.4 \cdot 10^5 \leq \left| j(2\tau) - j\left(\frac{\tau+1}{2}\right) \right|, \left| j\left(\frac{\tau}{2}\right) - j\left(\frac{\tau+1}{2}\right) \right| \leq 3.1 \cdot 10^5,$$

*and if $\tau = \frac{1+i\sqrt{3}}{2} + \delta$, with $|\delta| \leq 2^{-30}$, then for $\tau_1, \tau_2 \in \{2\tau, \frac{\tau}{2}, \frac{\tau+1}{2}\}$, $\tau_1 \neq \tau_2$,*

$$1.15 \cdot 10^6 |\delta| \leq |j(\tau_1) - j(\tau_2)| \leq 1.34 \cdot 10^6 |\delta|.$$

*Proof.* We first use the Taylor series expansion of $j(z)$ at $2i$. By Lemma 4.6, we may apply Lemma 4.5, and using also Lemma 4.7, we obtain the lower bound

$$\begin{aligned}
\left| j(2\tau) - j\left(-\frac{2}{\tau}\right) \right| &\geq |j(2i) + j'(2i)\delta_1 - j(2i) - j'(2i)\delta_2| - 0.4|\delta| \\
&= |j'(2i)|\,|\delta_1 - \delta_2| - 0.4|\delta| \\
&= |j'(2i)| \left| 2\tau + \frac{2}{\tau} \right| - 0.4|\delta| \\
&\geq 1.8 \cdot 10^6 \cdot 3.99|\delta| - 0.4|\delta| \\
&\geq 7.18 \cdot 10^6 |\delta|.
\end{aligned}$$

Similarly we have an upper bound of

$$\left| j(2\tau) - j\left(-\frac{2}{\tau}\right) \right| \leq 1.81 \cdot 10^6 \cdot 4|\delta| + 194|\delta| \leq 7.25 \cdot 10^6 |\delta|.$$

By Lemma 4.7, the $\mathrm{SL}_2(\mathbb{Z})$-equivalent point in $\mathcal{F}$ to $\frac{\tau+1}{2}$ is at most $2|\delta| \leq 2^{-29}$ from $i$, so

$$\left| j\left(\frac{\tau+1}{2}\right) \right| \leq e^{2.01\pi} + 2079 \leq e^{7.9},$$

and as by Lemma 4.6, $2\tau, -\frac{\tau}{2}$ are at most $2^{-29}$ from $2i$,

$$|j(2\tau)|, \left| j\left(\frac{\tau}{2}\right) \right| \geq e^{3.99\pi} - 2079 \geq e^{12.5},$$

so that

$$3.1 \cdot 10^5 \geq \left| j(2\tau) - j\left(\frac{\tau+1}{2}\right) \right|, \left| j\left(-\frac{\tau}{2}\right) - j\left(\frac{\tau+1}{2}\right) \right| \geq 2.4 \cdot 10^5.$$

Now using the Taylor series of $j(z)$ at $i\sqrt{3}$, by Lemma 4.6, we may apply Lemma 4.5, and using also Lemma 4.7, with $\tau_1$, $\tau_2$ any distinct pair of $2\tau$ ($\mathrm{SL}_2(\mathbb{Z})$-equivalent to $2\tau - 1$), $\frac{\tau}{2}$, ($\mathrm{SL}_2(\mathbb{Z})$-equivalent to $1 - \frac{2}{\tau}$), and $\frac{\tau+1}{2}$ ($\mathrm{SL}_2(\mathbb{Z})$-equivalent to $-1 - \frac{1}{\tau-1}$), with $\tau_j = i\sqrt{3} + \delta_j$, we obtain the following lower bound,

$$\begin{aligned}
|j(\tau_1) - j(\tau_2)| &\geq \left| j(\sqrt{3}i) + j'(\sqrt{3}i)\delta_1 - (j(\sqrt{3}i) + j'(\sqrt{3}i)\delta_2) \right| - 0.4|\delta| \\
&= |j'(\sqrt{3}i)||\tau_1 - \tau_2| - 0.4|\delta| \\
&\geq 334500 \cdot 3.46|\delta| - 0.4|\delta| \\
&\geq 1.15 \cdot 10^6 |\delta|.
\end{aligned}$$

For an upper bound we obtain

$$\begin{aligned}
|j(\tau_1) - j(\tau_2)| &\leq \left| j(\sqrt{3}i) + j'(\sqrt{3}i)\delta_1 - (j(\sqrt{3}i) + j'(\sqrt{3}i)\delta_2) \right| + 0.4|\delta| \\
&= |j'(\sqrt{3}i)||\tau_1 - \tau_2| + 0.4|\delta| \\
&\leq 334600 \cdot 4|\delta| + 0.4|\delta| \\
&\leq 1.34 \cdot 10^6 |\delta|.
\end{aligned}$$

□

#### 4.2.2.1 $j$ near 1728

We now bound the discrepancy between the roots of $\Phi_2(j, z)$ and $\Phi_2(\tilde{j}, z)$ when $|\tau - i| \leq 2^{-30}$ and $|\tilde{j} - 1728| \geq 2^{-P/3}$.

**Lemma 4.9.** *Suppose that $\tilde{j}$ is an approximation to $j(\tau)$ of relative precision $2^{-P}$, with $P \geq 300$, $|\tau - i| \leq 2^{-30}$, and $\left|\tilde{j} - 1728\right| \geq 2^{-P/3}$. Then the relative precision of any root of $\Phi_2(\tilde{j}, z)$ to the closest root of $\Phi_2(j, z)$ is at least $2^{-P/3+10}$*

*Proof.* Firstly, as in the proof of Lemma 4.1, with $f(z) = \Phi_2(j, z)$ and $g(z) = \Phi_2(\tilde{j}, z)$, and $\tilde{j} = j + \delta$, we will bound the size of the coefficients of $f - g$. We note that $|j| \geq 1700$. For $z^2$, we have

$$|2j\delta + \delta^2 + 1448\delta| \leq 3|\delta||j|,$$

for $z$, we have

$$|2976\delta j + 1488\delta^2 + 40773375\delta| \leq 30000|\delta||j|,$$

and for the constant term

$$\left|3\delta j^2 + 3\delta^2 j + \delta^3 - 16200\delta j - 16200\delta^2 + 8748000000\delta\right| \leq 3100|\delta||j|^2.$$

Now evaluating $g$ at $j(2\tau)$, we have, noting that $|j(2\tau)| \leq 0.1|j|^2$,

$$\begin{aligned}|g(j(2\tau))| &= |f(j(2\tau)) - g(j(2\tau))| \\ &\leq 3|\delta||j|(0.1|j|^2)^2 + 30000|\delta||j|(0.1|j|^2) + 3100|\delta||j|^2 \\ &\leq 0.04|j|^5\end{aligned}$$

Letting $\beta_0, \beta_1, \beta_2$ be the roots of $g$, where $\beta_0$ is the closest root of $g(z)$ to $j(2\tau)$,

$$\begin{aligned}|j(2\tau) - \beta_0||j(2\tau) - \beta_1||j(2\tau) - \beta_2| &\leq 0.04|\delta||j|^5 \\ |j(2\tau) - \beta_0| &\leq 10^5|\delta|^{1/3} \\ &\leq 1.2 \cdot 10^6 \cdot 2^{-P/3},\end{aligned}$$

and similarly for the nearest root of $g(z)$ to each of the roots of $f(z)$. As $|\tilde{j} - 1728| \geq 2^{-P/3}$, $|j - 1728| \geq 2^{-P/3} - 1800 \cdot 2^{-P} \geq 2^{-P/3-1}$, and by Lemma 4.4, $|\tau - i| \geq 2^{-P/6-9}$, so that by Lemma 4.8,

$$\left|j(2\tau) - j\left(\frac{\tau}{2}\right)\right| \geq 7.18 \cdot 10^7 \cdot 2^{-P/6-9}.$$

Now letting $\beta_i$ and $\beta_j$ be the nearest roots to $j(2\tau)$ and $j\left(\frac{\tau}{2}\right)$ respectively,

$$\begin{aligned}|\beta_i - \beta_j| &= \left|(\beta_i - j(2\tau)) - j(2\tau) - \left(\left(\beta_j - j\left(\frac{\tau}{2}\right)\right) - j\left(\frac{\tau}{2}\right)\right)\right| \\ &\geq \left|j(2\tau) - j\left(\frac{\tau}{2}\right)\right| - |j(2\tau) - \beta_i| - \left|j\left(\frac{\tau}{2}\right) - \beta_j\right| \\ &\geq 7.18 \cdot 10^7 \cdot 2^{-P/6-9} - 1.2 \cdot 10^6 \cdot 2^{-P/3} \\ &\geq 7.18 \cdot 10^7 \cdot 2^{-P/6-9} - 0.01 \cdot 2^{-P/6} \\ &\geq 1.4 \cdot 10^5 \cdot 2^{-P/6}.\end{aligned}$$

In particular, $\beta_i$ and $\beta_j$ are distinct, and given that, by Lemma 4.8, $\left|j(2\tau) - j\left(\frac{\tau+1}{2}\right)\right| \geq 2.4 \cdot 10^5 \geq 7.18 \cdot 10^7 \cdot 2^{-P/6-9}$, one may similarly deduce a separation of $2 \cdot 10^5$ between the closest root to $j\left(\frac{\tau+1}{2}\right)$ and either of the other two. So each root $j_i$ of $\Phi_2(j, z)$ has a unique associated closest root $\beta_i$ of $\Phi_2(\tilde{j}, z)$, which satisfies

$$|j_i - \beta_i| \leq 1.2 \cdot 10^6 \cdot 2^{-P/3}.$$

The relative precision of $\beta_i$ to its closest root is then bounded by

$$\frac{1.2 \cdot 10^6 \cdot 2^{-P/3}}{|j(2\tau)|}, \frac{1.2 \cdot 10^6 \cdot 2^{-P/3}}{|j\left(\frac{\tau}{2}\right)|} \leq 2^{-P/3+3}$$

if $\beta_i$ is the root closest to $j(2\tau)$ or $j\left(\frac{\tau}{2}\right)$, and

$$\frac{1.2 \cdot 10^6 \cdot 2^{-P/3}}{|j(\frac{\tau+1}{2})|} \leq 2^{-P/3+10}$$

for the root closest to $j\left(\frac{\tau+1}{2}\right)$. □

Finally we show that Newton's method may be used to obtain an approximation to $j(2\tau)$.

**Proposition 4.3.** *Let $j = j(\tau)$, where $|\tau - i| \leq 2^{-30}$, and $\tilde{j}$ be an approximation to $j$ of relative precision $2^{-P}$, $P \geq 300$, such that $|\tilde{j} - 1728| \geq 2^{-P/3}$. Let*

$$\tau_0 = i + \sqrt{\frac{\tilde{j} - 1728}{j^{(2)}(i)}},$$

*where the sign of the square root is chosen arbitrarily. Then Newton's method applied to $\Phi_2(\tilde{j}, z)$, with starting point $j(2\tau_0)$ will converge to either the root of $\Phi_2(\tilde{j}, z)$ closest to $j(2\tau)$, or the root of $\Phi_2(\tilde{j}, z)$ closest to $j\left(\frac{\tau}{2}\right)$, and after $[2 \log P]$ steps will produce an approximation either $j(2\tau)$ or $j\left(\frac{\tau}{2}\right)$ of relative precision $2^{-P/3+11}$.*

*Proof.* Firstly, we bound the difference between $\tau$ and $\tau_0$. We let $\tau = i + \delta$. By Lemma 4.3,

$$\left| j(\tau) - 1728 - \frac{j^{(2)}(i)}{2}\delta^2 \right| \leq 0.07|\delta|^2,$$

so that

$$\left| \sqrt{\frac{2(\tilde{j} - 1728)}{j^{(2)}(i)}} - \delta \right| \left| \sqrt{\frac{2(\tilde{j} - 1728)}{j^{(2)}(i)}} + \delta \right| \leq \frac{388}{j^{(2)}(i)}|\delta|^2 + \frac{2^{-P+1}|j|}{j^{(2)}(i)} \leq 3 \cdot 10^{-6}|\delta|^2.$$

Now we will show that, taking some branch of the square root above, we will obtain a good starting point for Newton's method. Let $\epsilon = \sqrt{\frac{2(\tilde{j}-1728)}{j^{(2)}(i)}}$, where the branch of the square root is arbitrary. Firstly, one of $|\epsilon - \delta|$ or $|\epsilon + \delta|$ is $\geq |\delta|$, and so the other is $\leq 3 \cdot 10^{-6}|\delta|$. In the first case, $|\tau_0 - \tau| \leq 3 \cdot 10^{-6}|\delta|$, and in the second case,

$$\left| \tau_0 - \left( -\frac{1}{\tau} \right) \right| = \left| \frac{-1 + i\delta + i\epsilon + 1 + \delta\epsilon}{\tau} \right| \leq |\delta + \epsilon| + |\delta\epsilon| \leq 3.1 \cdot 10^{-6}|\delta|.$$

Now as $|\delta| \leq 2^{-30}$, $j'(z)$ is bounded in absolute value by $1.81 \cdot 10^6$ between $2\tau_0$ and either of $2\tau$ or $-\frac{2}{\tau}$, and the distance from $2\tau_0$ to the closest of these is bounded by $6.2 \cdot 10^{-6}|\delta|$, so either

$$|j(2\tau_0) - j(2\tau)| \leq 12|\delta|$$

or

$$\left| j(2\tau_0) - j\left( -\frac{2}{\tau} \right) \right| \leq 12|\delta|.$$

Now we bound the terms occurring in Kantorovich's criterion. Let $\beta_0, \beta_1, \beta_2$ be the roots of $\Phi_2(\tilde{j}, z)$, with $\beta_0$ the closest root to $j(2\tau_0)$, and $\beta_1$ the other root near $j(2i)$.

Firstly, as $|j| \geq 2^{-P/3}$, by Lemma 4.8, the above, and Lemma 4.9, we have, for $2\tau_0$,

$$\begin{aligned}
|j(2\tau_0) - \beta_0| &\leq 12|\delta| + 2^{-P/3+10} \leq 12.1|\delta| \\
|j(2\tau_0) - \beta_1| &\leq 7.25 \cdot 10^6|\delta| + 2^{-P/3+10} \leq 7.26 \cdot 10^6|\delta| \\
|j(2\tau_0) - \beta_2| &\leq 3.1 \cdot 10^5 + 2^{-P/3+10} \leq 3.11 \cdot 10^5 \\
|j(2\tau_0) - \beta_1| &\geq 7.18 \cdot 10^6|\delta| - 2^{-P/3+10} \geq 7.17 \cdot 10^6|\delta| \\
|j(2\tau_0) - \beta_2| &\geq 2.4 \cdot 10^5 - 2^{-P/3+10} \geq 2.39 \cdot 10^5,
\end{aligned}$$

and similarly, for any $\tau'$ satisfying $|\tau' - 2\tau| \leq 35|\delta|$ (which we take to ensure the condition on $r$ is satisfied), we have, as $|j'(z)| \leq 1.81 \cdot 10^6$ between $2\tau_0$ and $\tau'$, and as $|\delta| \leq 2^{-30}$,

$$\begin{aligned}
|j(\tau') - \beta_0| &\leq 6.4 \cdot 10^7|\delta| + 12|\delta| \leq 0.06, \\
|j(\tau') - \beta_1| &\leq 6.4 \cdot 10^7|\delta| + 7.26 \cdot 10^6|\delta| \leq 0.06, \\
|j(\tau') - \beta_2| &\leq 6.4 \cdot 10^7|\delta| + 3.11 \cdot 10^5 \leq 3.2 \cdot 10^5.
\end{aligned}$$

For our bound on $\Phi_2(\tilde{j}, z)$ evaluated at $j(\tau')$, we have

$$\begin{aligned}
\left|\Phi_2(\tilde{j}, j(2\tau_0))\right| &= |j(\tau') - \beta_0||j(\tau') - \beta_1||j(\tau') - \beta_2| \\
&\leq 2.8 \cdot 10^{13}|\delta|^2.
\end{aligned}$$

For the first derivative, we have

$$\begin{aligned}
\left|\Phi_2'(\tilde{j}, j(2\tau_0))\right| &= (j(\tau') - \beta_0)(j(\tau') - \beta_1) + (j(\tau') - \beta_0)(j(\tau') - \beta_2) \\
&\quad + (j(\tau') - \beta_1)(j(\tau') - \beta_2) \\
&\geq |j(\tau') - \beta_1||j(\tau') - \beta_2| - |j(\tau') - \beta_0||j(\tau') - \beta_1| \\
&\quad - |j(\tau') - \beta_0||j(\tau') - \beta_2| \\
&\geq 7.17 \cdot 10^6 \cdot 2.39 \cdot 10^5|\delta| - 12.1 \cdot 7.26 \cdot 10^6|\delta|^2 \\
&\quad - 12.1 \cdot 3.11 \cdot 10^5|\delta| \\
&\geq 1.7 \cdot 10^{12}|\delta|,
\end{aligned}$$

and for the second derivative,

$$\begin{aligned}
\left|\Phi_2''(\tilde{j}, j(\tau'))\right| &\leq 2|j(\tau') - \beta_0| + 2|j(\tau') - \beta_1| + 2|j(\tau') - \beta_2| \\
&\leq 6.5 \cdot 10^5.
\end{aligned}$$

These now give

$$\frac{|\Phi_2(\tilde{j}, j(2\tau_0))||\Phi_2''(\tilde{j}, j(\tau'))|}{|\Phi_2'(\tilde{j}, j(2\tau_0))|^2} \leq \frac{2.8 \cdot 10^{13}|\delta|^2 \cdot 6.5 \cdot 10^5}{(1.7 \cdot 10^{12}|\delta|)^2} \leq 2^{-17} < \frac{1}{2},$$

and for the condition on $r$, we have

$$\begin{aligned} r &= 35|\delta|, \\ 2\eta &= 2\frac{|\Phi_2(\tilde{j}, j(2\tau_0))|}{|\Phi_2'(\tilde{j}, j(2\tau_0))|} \leq 2\frac{2.8 \cdot 10^{13}|\delta|^2}{1.7 \cdot 10^{12}|\delta|} \leq 34|\delta|, \end{aligned}$$

which ensures convergence. For the rate of convergence, we need a lower bound on the second derivative and an upper bound on the first derivative, which we have as follows,

$$\begin{aligned} \left|\Phi_2''(\tilde{j}, j(\tau'))\right| &\geq 2|j(\tau') - \beta_2| - 2|j(\tau') - \beta_0| - 2|j(\tau') - \beta_1| \\ &\geq 4.7 \cdot 10^5, \end{aligned}$$

and

$$\begin{aligned} \left|\Phi_2'(\tilde{j}, j(2\tau_0))\right| &\leq |j(2\tau_0) - \beta_1||j(2\tau_0) - \beta_2| + |j(2\tau_0) - \beta_0||j(2\tau_0) - \beta_1| \\ &\quad + |j(2\tau_0) - \beta_0||j(2\tau_0) - \beta_2| \\ &\leq 7.25 \cdot 10^6 \cdot 3.11 \cdot 10^5|\delta| + 12.1 \cdot 7.25 \cdot 10^6|\delta|^2 + 12.1 \cdot 3.11 \cdot 10^5|\delta| \\ &\leq 2.3 \cdot 10^{12}|\delta|, \end{aligned}$$

which gives a bound on the convergence, for $k \geq 1$, of

$$\frac{1}{2^k} 2^{-17 \cdot 2^k} \frac{|\Phi_2'(\tilde{j}, j(2\tau_0))|}{|\Phi_2''(\tilde{j}, j(\tau'))|} \leq 2^{-17 \cdot 2^k}.$$

So in order to obtain an absolute precision of $2^{-P}$, $[2 \log P]$ steps will suffice. By Lemma 4.9, this approximation to $\beta_0$ will then be an approximation to either $j(2\tau)$ or $j\left(-\frac{2}{\tau}\right)$ of relative precision $2^{-P/3+11}$. □

#### 4.2.2.2 $j$ near 0

We now bound the discrepancy between the roots of $\Phi_2(j, z)$ and $\Phi_2(\tilde{j}, z)$ when $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-31}$ and $|\tilde{j}| \geq 2^{-P/2}$.

**Lemma 4.10.** *Let $j = j(\tau)$, where $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-31}$, and suppose that $\tilde{j}$ is an approximation to $j$ of absolute precision $2^{-P}$, with $P \geq 300$, and $\left|\tilde{j}\right| \geq 2^{-P/2}$. Then the relative precision of any root of $\Phi_2(\tilde{j}, z)$ to its closest root of $\Phi_2(j, z)$ is at most $2^{-P/3+2}$, and the roots of $\Phi_2(\tilde{j}, z)$ are separated by at least $2^{-P/6+8}$.*

*Proof.* Firstly, as in the proof of the Lemma 4.9, with $f(z) = \Phi_2(j, z)$ and $g(z) = \Phi_2(\tilde{j}, z)$, we will bound the size of the coefficients of $f - g$. Let $\tilde{j} = j + \delta$. For the coefficient of $z^2$, we have

$$|2j\delta + \delta^2 + 1448\delta| \leq 1500|\delta|,$$

for the coefficient of $z$,

$$|2976\delta j + 1488\delta^2 + 40773375\delta| \leq 4.1 \cdot 10^7|\delta|,$$

and for the constant term

$$\left|3\delta j^2 + 3\delta^2 j + \delta^3 - 16200\delta j - 16200\delta^2 + 8748000000\delta\right| \leq 8.8 \cdot 10^9|\delta|.$$

Now evaluating $g(z)$ at $j(\tau')$, for $\tau' \in \{2\tau, \frac{\tau}{2}, \frac{\tau+1}{2}\}$, as $|j(\tau')| \leq 60000$, we have

$$\begin{aligned}|g(j(2\tau))| &\leq 60000^2 \cdot 1500|\delta| + 60000 \cdot 4.1 \cdot 10^7|\delta| + 60000 \cdot 8.8 \cdot 10^9|\delta| \\ &\leq 6 \cdot 10^{14}|\delta|.\end{aligned}$$

Letting $\beta_0, \beta_1, \beta_2$ be the roots of $g(z)$, we have, as $|\tilde{j}| \geq 2^{-P/2}$, which implies $|j| \geq 2^{-P/2-1}$,

$$\begin{aligned}|\beta_0 - j(\tau_i)||\beta_1 - j(\tau_i)||\beta_2 - j(\tau_i)| &\leq 6 \cdot 10^{14}|\delta| \\ &\leq 2^{-P+50}\end{aligned}$$

Now for any $\beta_i$, letting $j(\tau_j)$ be the nearest root of $f(z)$, we have

$$|\beta_i - j(\tau_j)| \leq 2^{-P/3+17}.$$

By Lemma 4.4, as $|j| \geq 2^{-P/2}$,

$$\left|\tau - \frac{1 + i\sqrt{3}}{2}\right| \geq 2^{-P/6-9},$$

which yields, by Lemma 4.8, for $i \neq j$,

$$|j(\tau_i) - j(\tau_j)| \geq 2^{-P/6+9}.$$

Similarly to the previous lemma, we now observe that, with $\beta_i$ the closest root to $j(\tau_i)$,

$$\begin{aligned}|\beta_i - \beta_j| &\geq |j(\tau_i) - j(\tau_j)| - 2 \cdot 2^{-P/3+17} \\ &\geq 2^{-P/6+9} - 2^{-P/3+17} \\ &\geq 2^{-P/6+8}\end{aligned}$$

and so $\beta_i$, $\beta_j$ are distinct. So each root has a unique closest associated root, of relative precision

$$\frac{2^{-P/3+17}}{|j(\tau_j)|} \leq 2^{-P/3+2}.$$

□

**Proposition 4.4.** *Let $j = j(\tau)$, where $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \leq 2^{-31}$, and $\tilde{j}$ be an approximation to $j$ of relative precision $2^{-P}$ such that $|\tilde{j}| \geq 2^{-P/2}$. Let*

$$\tau_0 = \frac{1+i\sqrt{3}}{2} + \sqrt[3]{\frac{6\tilde{j}}{j^{(3)}(i)}},$$

*where the branch of the cube root is chosen arbitrarily. Then Newton's method applied to $\Phi_2(\tilde{j}, z)$, with starting point $j(2\tau_0)$ will converge to either the root of $\Phi_2(\tilde{j}, z)$ closest to $j(2\tau)$, the root of $\Phi_2(\tilde{j}, z)$ closest to $j\left(\frac{\tau}{2}\right)$, or the root of $\Phi_2(\tilde{j}, z)$ closest to $j\left(\frac{\tau+1}{2}\right)$, and after $[2\log P]$ steps will produce an approximation to either $j(2\tau)$, $j\left(\frac{\tau}{2}\right)$, or $j\left(\frac{\tau+1}{2}\right)$ of relative precision $2^{-P/3+3}$.*

*Proof.* Let $\tau = \frac{1+i\sqrt{3}}{2} + \delta$. By Lemma 4.3,

$$\left| j(\tau) - \frac{j^{(3)}\left(\frac{1+i\sqrt{3}}{2}\right)}{6}\delta^3 \right| \leq 0.07|\delta|^3,$$

so that

$$\left| \frac{6\tilde{j}}{j^{(3)}\left(\frac{1+i\sqrt{3}}{2}\right)} - \delta^3 \right| \leq 1.6 \cdot 10^{-6}|\delta|^3.$$

We let $\epsilon = \sqrt[3]{\frac{6\tilde{j}}{j^{(3)}\left(\frac{1+i\sqrt{3}}{2}\right)}}$. Considering the geometry of the cube roots of $\delta^3$, the product of the two furthest from $\epsilon$ is at least $|\delta|^2$, so letting the closest be $\delta_0$, we have

$$|\epsilon - \delta_0| \leq 1.6 \cdot 10^{-6}|\delta|.$$

We now show that any branch of the cube root taken for $\epsilon$ will result $2\tau_0 - 1$ being very close to one of the $\mathrm{SL}_2(\mathbb{Z})$-equivalent elements of $\mathcal{F}$ to one of $2\tau$, $\frac{\tau}{2}$, $\frac{\tau+1}{2}$. We have

$$\begin{aligned}
\left| \frac{1+i\sqrt{3}}{2} + \epsilon - \left(-\frac{1}{\tau} + 1\right) \right| &= \frac{\left|\tau + i\sqrt{3}\tau + 2 + 2\epsilon\tau - 2\tau\right|}{2|\tau|} \\
&= \frac{\left|(1+i\sqrt{3})\epsilon - (1-i\sqrt{3})\delta + 2\epsilon\delta\right|}{|2\tau|} \\
&\leq |\epsilon - e^{4i\pi/3}\delta| + \frac{|\delta\epsilon|}{2},
\end{aligned}$$

and

$$\begin{aligned}\left|\frac{1+i\sqrt{3}}{2}+\epsilon-\left(-\frac{1}{\tau-1}\right)\right| &= \frac{\left|\tau-1+i\sqrt{3}(\tau-1)+2+2\epsilon(\tau-1)\right|}{2|\tau|}\\ &= \frac{\left|(1-i\sqrt{3})\epsilon-(1+i\sqrt{3})\delta-2\epsilon\delta\right|}{|2\tau|}\\ &\leq |\epsilon-e^{2i\pi/3}\delta|+\frac{|\delta\epsilon|}{2}.\end{aligned}$$

In particular, the difference between $2\tau_0-1$ and the closest of the $\mathrm{SL}_2(\mathbb{Z})$-equivalent elements of $\mathcal{F}$ to $2\tau, \frac{\tau}{2}, \frac{\tau+1}{2}$, which we let be $\tau_1$, and the others $\tau_2, \tau_3$, is bounded by

$$2\,|\epsilon-\delta_0|+|\delta\epsilon|\leq 3.3\cdot 10^{-6}|\delta|.$$

Now as the derivative of $j(z)$ is bounded in absolute value by $3.4\cdot 10^5$ between $\tau_0$ and $\tau_1$, we have

$$|j(2\tau_0)-j(\tau_1)|\leq 1.2|\delta|.$$

Now by the bound of Lemma 4.10 on the discrepancy of the roots of $\Phi_2(\tilde{j},z)$ and $\Phi_2(j,z)$, and the bounds on the distances between the distinct roots of Lemma 4.8, letting $\beta_0$ be the closest root of $\Phi_2(\tilde{j},z)$ to $j(2\tau_0)$, and $\beta_1,\beta_2$ the other two roots, we collect the relevant bounds for Kantorovich's criterion,

$$\begin{aligned}|j(2\tau_0)-\beta_0| &\leq 1.2|\delta|\\ |j(2\tau_0)-\beta_1|\,,|j(2\tau_0)-\beta_2| &\leq 1.35\cdot 10^6|\delta|\\ |j(2\tau_0)-\beta_1|\,,|j(2\tau_0)-\beta_2| &\geq 1.14\cdot 10^6|\delta|,\end{aligned}$$

and with $r=4|\delta|$, for any $\tau'$ such that $|\tau'-2\tau_0|\leq 4|\delta|$, we have, as the derivative of $j$ is bounded in absolute value by $3.4\cdot 10^5$ here,

$$\begin{aligned}|j(\tau')-\beta_0| &\leq 1.37\cdot 10^6|\delta|,\\ |j(\tau')-\beta_1|\,,|j(\tau')-\beta_2| &\leq 2.71\cdot 10^6|\delta|.\end{aligned}$$

Now we bound the various terms of Kantorovich's criterion. Firstly,

$$\begin{aligned}|\Phi_2(\tilde{j},j(2\tau_0))| &\leq 1.2|\delta|\cdot(1.35\cdot 10^6|\delta|)^2\\ &\leq 2.2\cdot 10^{12}|\delta|^3.\end{aligned}$$

For the first derivative, we have

$$\begin{aligned}|\Phi_2'(\tilde{j},j(2\tau_0))| &\geq (1.14\cdot 10^6|\delta|)^2-2\cdot 1.2|\delta|\cdot 1.35\cdot 10^6|\delta|\\ &\geq 1.2\cdot 10^{12}|\delta|^2,\end{aligned}$$

and for the second derivative,

$$\begin{aligned}|\Phi_2''(\tilde{j}, j(\tau'))| &\leq 2 \cdot 1.37 \cdot 10^6|\delta| + 4 \cdot 2.71 \cdot 10^6|\delta| \\ &\leq 1.36 \cdot 10^7|\delta|,\end{aligned}$$

so that we have

$$\frac{|\Phi_2(\tilde{j}, j(2\tau_0))||\Phi_2''(\tilde{j}, j(2\tau_0))|}{|\Phi_2'(\tilde{j}, j(2\tau_0))|^2} \leq \frac{2.2 \cdot 10^{12}|\delta|^3 \cdot 1.36 \cdot 10^7|\delta|}{(1.2 \cdot 10^{12}|\delta|^2)^2} \leq 2^{-15} < \frac{1}{2},$$

and for the condition on $r$ we have

$$\begin{aligned}r &= 4|\delta| \\ 2\eta &= 2\frac{|\Phi_2(\tilde{j}, j(2\tau_0))|}{|\Phi_2'(\tilde{j}, j(2\tau_0))|} \leq 3.8|\delta|,\end{aligned}$$

ensuring convergence. For the rate of convergence, we have the upper bound on the first derivative as follows,

$$\left|\Phi_2'(\tilde{j}, j(2\tau_0))\right| \leq 1.4 \cdot 10^{12}|\delta|^2.$$

For the lower bound on the second derivative, we have that

$$|\Phi_2''(\tilde{j}, j(\tau'))| = |6j(\tau') - 2(\beta_0 + \beta_1 + \beta_2)|.$$

Now the second term in the absolute value is the coefficient of $z^2$ of $\Phi_2(\tilde{j}, z)$, which is equal to

$$-\tilde{j}^2 + 1488\tilde{j} - 162000.$$

By Lemma 4.3, $|j| \leq \left(\frac{\left|j\left(\frac{1+i\sqrt{3}}{2}\right)\right|}{6} + 0.07\right)|\delta|^3$, and as $|\tilde{j} - j| \leq 2^{-P} \leq |j|^2$, $|\tilde{j}| \leq 1.01|j|$, so

$$|-\tilde{j}^2 + 1488\tilde{j}| \leq 4.13 \cdot 10^8|\delta|^3 \leq 0.1|\delta|.$$

In particular, $-(\beta_1 + \beta_2 + \beta_3) = 162000 + \theta_1$, where $|\theta_1| \leq 0.1|\delta|$. As $|\delta| \leq 2^{-31}$, $|\tau' - 2\tau_0| \leq 4|\delta|$, and $|2\tau_0 - 1 - i\sqrt{3}| \leq (2 + 3.3 \cdot 10^{-6})|\delta|$,

$$|\tau' - 1 - i\sqrt{3}| \leq 6.1|\delta| \leq 2^{-28},$$

and hence by Lemma 4.5,

$$|j(\tau') - j(i\sqrt{3})| \geq |j'(i\sqrt{3})\delta| - 1.3|\delta| \geq 334000|\delta|.$$

In particular, as $j(i\sqrt{3}) = 54000$, $j(2\tau_0) = 54000 + \theta_2$, where $|\theta_2| \geq 334000|\delta|$. Now combining these bounds, we have

$$\begin{aligned}|6j(\tau') - 2(\beta_0 + \beta_1 + \beta_2)| &= |324000 + 6\theta_2 - 324000 + 2\theta_1| \\ &\geq 6|\theta_2| - 2|\theta_1| \\ &\geq 6 \cdot 334000|\delta| - 0.2|\delta| \\ &\geq 2 \cdot 10^6|\delta|.\end{aligned}$$

Returning to the convergence, we have

$$\frac{|\Phi_2'(\tilde{j}, j(2\tau_0))|}{|\Phi_2''(\tilde{j}, j(\tau'))|} \leq \frac{1.4 \cdot 10^{12}|\delta|^2}{2 \cdot 10^6|\delta|} \leq 2^{-10},$$

so a bound for the rate of convergence is

$$2^{-15 \cdot 2^k}.$$

In particular, to obtain an absolute precision of $2^{-P}$, $[2 \log P]$ steps will suffice. The approximant obtained will then, by Lemma 4.10, be an approximation of one of $j(2\tau), j\left(\frac{\tau}{2}\right)$, or $j\left(\frac{\tau+1}{2}\right)$, of relative precision at least $2^{-P/3+3}$. □

#### 4.2.2.3 Running time

The time required to obtain an approximation to one of $j(2\tau), j\left(\frac{\tau}{2}\right)$, or $j\left(\frac{\tau+1}{2}\right)$ is, as in the first section, by using a variable-precision Newton iteration, $O(M(P))$, and the obtained value may then be used as an input to Newton's method on the compact set described in the subsequent section. In order to determine which of $j(2\tau), j\left(\frac{\tau}{2}\right), j\left(\frac{\tau+1}{2}\right)$ was computed above, after obtaining an inverse $\tau_0$, we compute $j(2\tau_0)$ and $j\left(\frac{\tau_0}{2}\right)$. We may then return the argument corresponding to whichever was closest to our initial approximation $\tilde{j}$ to $j$, after applying elements of $\mathrm{SL}_2(\mathbb{Z})$ to move it to the fundamental domain (which takes $O(M(P))$ time). In particular we will obtain an approximation of precision at least $2^{-P/3+12}$.

### 4.2.3 Newton iteration on the compact set

For $\tau$ such that $\tau \in \mathcal{F}$, $\mathrm{Im}(\tau) \leq 3.1$, $|\tau - i| \geq 2^{-32}$, $\left|\tau - \frac{1+i\sqrt{3}}{2}\right| \geq 2^{-32}$, we make use of an algorithm due to Dupont, [31] for the quasilinear evaluation of $j(\tau)$ to relative precision, in order to invert $j(z)$ by the secant method. As we are considering a compact set, there is some fixed precision our starting points for the secant method

may be in order to obtain convergence for any $\tau$. We compute the low precision inverses of $j$ by the formula

$$\tau_0 = i\frac{{}_2F_1\left(\frac{1}{6},\frac{5}{6},1,\frac{1}{2}+\frac{1}{2}\sqrt{1-\frac{1728}{\tilde{j}}}\right)}{{}_2F_1\left(\frac{1}{6},\frac{5}{6},1,\frac{1}{2}-\frac{1}{2}\sqrt{1-\frac{1728}{\tilde{j}}}\right)},$$
$$\tau_1 = \tau_0 - \frac{j(\tau_0)-j}{j'(\tau_0)},$$

and we note that as $j$ is bounded (absolutely) away from 0 and 1728, the arguments of the Gaussian hypergeometric functions are bounded away from the branch points 0 and 1, and $j'$ is bounded (absolutely) away from 0, so computation of the two points to a fixed precision takes a fixed amount of time. Upon a failure of convergence (or slow convergence), we may simply increase the precision of our starting points and repeat the process — as there is a uniform bound on the required precision, this requires only constant time. As the secant method converges doubly exponentially (with exponent $-c\left(\frac{1+\sqrt{5}}{2}\right)^k$, as opposed to the Newton method's $-c2^k$), and the evaluation of $j(z)$ via the algorithm of [31] takes $O(M(P)\log P)$ time, the computational complexity of obtaining an approximation of precision $2^{-P}$ by a variable-precision iteration requires time $O(M(P)\log P)$.

A similar analysis to that of Lemma 4.3 gives a bound of

$$\left|j^k(z_0)\right| \leq 3\cdot 10^8 \cdot 14^k k!$$

for $z_0$ in our compact set, so that if $z_1 = z_0 + |\delta|$, with $|\delta| \leq 2^{-150}$,

$$\begin{aligned}|j(z_0)+j'(z_0)\delta - j(z_1)| &\leq 3\cdot 10^8 \sum_{n=2}^{\infty}|\delta|^n 14^n \\ &\leq 3\cdot 10^8 \cdot 14^2 \cdot 2^{-150}\cdot|\delta|\sum_{n=0}^{\infty}2^{-150n}14^n \\ &\leq 5\cdot 10^{-35}|\delta|.\end{aligned}$$

It may be verified that $|j'(z)| \geq 10^{-19}$ in our compact set, so

$$|j(z_0)-j(z_1)| \geq |j'(z_0)||\delta| - 5\cdot 10^{-35}|\delta| \geq 10^{-20}|\delta|,$$

and in particular once we have computed by the secant method $\tau^*$ such that $|\tilde{j} - j(\tau^*)| \leq 2^{-P}$, $|j - j(\tau^*)| \leq 3\cdot 10^8 \cdot 2^{-P}$, and so, with $j(\tau) = j$, and $\tau^* = \tau + \delta$, we have

$$|\delta| \leq 3\cdot 10^8 \cdot 10^{20}\cdot 2^{-P} \leq 2^{-P+100}.$$

### 4.2.4 $j$ very close to 0 or 1728

Now if $|\tilde{j}| \leq 2^{-P/2}$, or $|\tilde{j} - 1728| \leq 2^{-P/3}$, we simply return $\frac{1+i\sqrt{3}}{2}$ or $i$ respectively, and this will, by Lemma 4.4, be an approximation to the inverse of $j$ of absolute, and as $|\tau| \geq 1$, relative precision $2^{-P/6}$.

## 4.3 Testing complex multiplication

Our algorithm to test complex multiplication rests on the following facts,

1. If $j$ is a singular modulus, then $j = j(\tau)$ for some quadratic irrational $\tau$;
2. The Mahler measure of $j(\tau)$ increases as the discriminant of $\tau$ increases;
3. Consequently, from a bound on the height and degree of $j$, we can bound the discriminant of any quadratic irrational $\tau$ for which it is possible that $j = j(\tau)$, so that our candidate preimages are finite and discrete;
4. By inverting $j$ to $z_0$ with sufficient precision we can determine which quadratic irrational $\tau$ it may be that $j = j(\tau)$;
5. Whether or not $j = j(\tau)$, both $j$ *and* $j(\tau)$ are algebraic, of bounded height and degree, so if $|j - j(\tau)|$ is sufficiently small, then in fact $j = j(\tau)$;
6. If the preimage $z_0$ of $j$ and a quadratic irrational $\tau$ are sufficiently close, then $j$ and $j(\tau)$ will be close also.

Firstly, given the input of the $j$-invariant of an elliptic curve $E$, the degree $d$ of $j$, and a bound on the absolute multiplicative height (c.f. [16]) $H \geq e^e$ of $j$, we bound the maximum discriminant $D$ of a quadratic irrational $\tau$ for which $j$ may equal $j(\tau)$, if $E$ were to have complex multiplication. Here we will denote by $M(—)$ the Mahler measure of an algebraic number, and note that by assumption $M(j) \leq H^d$. We first consider $\tau$ with $|D| \geq 16$.

Fixing a quadratic irrational $\tau$, with minimal polynomial $az^2 + bz + c$, its discriminant is $D = b^2 - 4ac$. Letting $\tau_i$ be the set of quadratic irrationals of discriminant $D$ in $\mathcal{F}$, by the theory of complex multiplication [26], the minimal polynomial of $j(\tau)$ is

$$\prod_i (z - j(\tau_i)).$$

In particular, considering the quadratic irrational which is the root in $\mathbb{H}$ of $z^2 - D/4$ if $D$ is even, and of $z^2 + z - (D-1)/4$ if $D$ is odd, one of the conjugates of $j(\tau)$ is

$j\left(\frac{D+\sqrt{D}}{2}\right)$, which corresponds to the principal form of discriminant $D$ — we can use this conjugate to bound below the Mahler measure of $j(\tau)$. As $|D| \geq 16$, $M(j(\tau)) \geq e^{\pi\sqrt{|D|}} - 2079 \geq e^{3.13\sqrt{|D|}}$. In particular, as $M(j) \leq H^d$,

$$|D| \leq \frac{d^2(\log H)^2}{9.7}.$$

We next bound the degree of $j(z)$ at a quadratic irrational of discriminant $D$.

**Lemma 4.11.** *Let $\tau$ be an imaginary quadratic irrational of discriminant $D$. Then*

$$[\mathbb{Q}(j(\tau)) : \mathbb{Q}] \leq \frac{1}{\pi}\sqrt{|D|}(2 + \log|D|).$$

*Proof.* This is an immediate consequence of Proposition 2.2, [63]. □

We now bound above the Mahler measure of $j(\tau)$ for $\tau$ a quadratic irrational.

**Lemma 4.12.** *Let $\tau$ be an imaginary quadratic irrational of discriminant $D$. Then*

$$M(j(\tau)) \leq e^{5.9\pi\sqrt{|D|}(\log|D|)^2}.$$

*Proof.* By the theory of complex multiplication, the conjugates of $j(\tau)$ are

$$j\left(\frac{-b+\sqrt{D}}{2a}\right),$$

where $ax^2 + bxy + cy^2$ is a reduced binary quadratic form of discriminant $D$. As $|j(z)| \leq e^{2\pi\mathrm{Im}(z)} + 2079 \leq 9.1e^{2\pi\mathrm{Im}(z)}$, we have the following upper bound,

$$M(j(\tau)) \leq 9.1^{h(D)} \exp\left(2\pi\sqrt{|D|} \sum_{\substack{(a,b,c) \\ \text{reduced}}} \frac{1}{a}\right).$$

The number of times each $a$ may occur is bounded by the number of solutions of $b^2 \equiv d(2a)$, which is bounded by twice the number of distinct divisors $r(a)$ of $a$. By [7], we have the bound

$$A(x) := \sum_{1 \leq a \leq x} 2r(a) \leq 2x\log x + 0.4x + 2x^{1/2}.$$

By Abel's summation formula and the above bound,

$$\begin{aligned}\sum_{1 \leq a \leq h(D)} \frac{2r(a)}{a} &= \frac{A(h(D))}{h(D)} + \int_1^{h(D)} \frac{A(y)}{y^2} dy \\ &\leq (\log h(D))^2 + 2.4\log(h(D)) + 4.9 \\ &\leq 2.6\log|D|^2,\end{aligned}$$

yielding a bound of

$$M(j(\tau)) \leq 9.1^{h(D)} e^{5.2\pi\sqrt{|D|}(\log|D|)^2} \leq e^{5.9\pi\sqrt{|D|}(\log|D|)^2}.$$

□

We can now apply this bound to give a lower bound on the difference of $j$ and any $j(\tau)$, where $\tau$ is a quadratic irrational of discriminant $D$, if $j \neq j(\tau)$. Firstly, the *length* of a polynomial is the sum of the absolute values of its coefficients. We make use of the following inequality, usually termed Liouville's inequality.

**Proposition** (Lemma 3 of [61], for example)**.** *Let $\alpha_1, \ldots, \alpha_n$ be algebraic numbers, of exact degrees $d_i$, $d = [\mathbb{Q}(\alpha_1, \ldots, \alpha_n) : \mathbb{Q}]$. Let $P(X_1, \ldots, X_n)$ be a polynomial of length $L$ and degrees $N_i$ in each variable. Then if $P(\alpha_1, \ldots, \alpha_n) \neq 0$,*

$$|P(\alpha_1, \ldots, \alpha_n)| \geq L^{1-d} \prod_{i=1}^{n} M(\alpha_i)^{\frac{dN_i}{d_i}}.$$

In particular, for $n = 2$,

$$|P(\alpha_1, \alpha_2)| \geq L^{-d_1 d_2} M(\alpha_1)^{d_2 N_1} M(\alpha_2)^{d_1 N_2}.$$

**Lemma 4.13.** *Let $j$ be an algebraic number of degree $\leq d$ and absolute multiplicative height $\leq H$, with $H \geq e^e$. Suppose that $j \neq j(\tau)$, and $\tau$ is an imaginary quadratic irrational of discriminant $D$ satisfying*

$$16 \leq |D| \leq \frac{d^2 (\log H)^2}{9.7}.$$

*Then*

$$|j - j(\tau)| \geq \exp\left(-30 d^2 \log H (\log d + \log\log H)^2\right).$$

*Proof.* By Lemmas 4.11, 4.12, and our assumptions on $j$ and $\tau$, we have the following parameters for Liouville's inequality, with $n = 2$,

$$\begin{aligned} M(j) &\leq H^d, \\ M(j(\tau)) &\leq e^{5.9\pi\sqrt{|D|}(\log|D|)^2}, \\ d_1 = [\mathbb{Q}(j) : \mathbb{Q}] &\leq d, \\ d_2 = [\mathbb{Q}(j(\tau)) : \mathbb{Q}] &\leq \frac{1}{\pi}\sqrt{|D|}(2 + \log|D|), \\ L &= 2, \\ N_1, N_2 &= 1, \end{aligned}$$

which yields

$$|j - j(\tau)| \geq 2^{-\frac{d}{\pi}\sqrt{|D|}(2+\log|D|)} H^{-\frac{d}{\pi}\sqrt{|D|}(2+\log|D|)} e^{5.9\pi d\sqrt{|D|}(\log|D|)^2},$$

and substituting our bound on $|D|$ of

$$|D| \leq \frac{d^2(\log H)^2}{9.7},$$

and collecting terms, we obtain the lower bound

$$\exp\left(-30d^2 \log H(\log d + \log\log H)^2\right).$$

□

Now we can give our criterion to determine if $j$ is a singular modulus with preimage $\tau$.

**Proposition 4.5.** *Let $j$ be an algebraic number of degree $\leq d$, and absolute multiplicative height $\leq H$, with $H \geq e^e$. Then from an approximation $\tilde{j}$ to $j$ of regulated precision*

$$2^{-300d^2 \log H(\log d + \log\log H)^2 - 200},$$

*and an imaginary quadratic irrational $\tau$ of discriminant $D$ satisfying*

$$16 \leq |D| \leq \frac{d^2(\log H)^2}{9.7},$$

*we may determine whether $j = j(\tau)$ by inverting $j$ to $z_0$ with our algorithm, and testing if*

$$|z_0 - \tau| \leq \exp\left(-31d^2 \log H(\log d + \log\log H)^2 - 21\right).$$

*Proof.* Letting $j = j(z_0)$, where $z_0 \in \mathcal{F}$, as $|j| \leq H^d$ and $|j(z_0)| \geq e^{2\pi \mathrm{Im}(z_0)} - 2079$, we have $\mathrm{Im}(z_0) \leq \frac{d\log H}{2\pi} + 2$. By differentiating once the $q$-expansion of $j$, and taking an upper bound, (note that the sum of the absolute values of the terms of the tail is decreasing in $\mathrm{Im}(\tau)$), we have, as $d\log H \geq 0$,

$$\|j'\|_{B(z_0,1)} \leq 2\pi e^{d\log H + 6\pi} + 800 \leq e^{d\log H + 21},$$

where $B(z_0, 1)$ is the disc of radius 1 with centre $z_0$. So if $|z_0 - \tau| \leq 1$,

$$|j - j(\tau)| \leq |z_0 - \tau| e^{d\log H + 21}.$$

Combining this with Lemma 4.13, noting that $\log d + \log\log H \geq 1$, if

$$|z_0 - \tau| \leq \exp\left(-31d^2 \log H(\log d + \log\log H)^2 - 21\right),$$

then $j = j(\tau)$, and so $j$ is a singular modulus. Our method of inverting $j$ from an input of regulated precision $2^{-P}$ obtains its inverse with relative precision at least $2^{-P/6}$. As $|z_0| \leq \frac{d \log H}{2\pi} + 2$, in order to obtain the inverse $z_0$ with absolute precision $2^{-Q}$, we will need an input of relative precision $2^{-6Q-2\log(d \log H+2)}$. In particular, with an input of relative precision

$$2^{-300d^2 \log H(\log d+\log\log H)^2-200},$$

our computed approximation $\tilde{z_0}$ to $z_0$ will have sufficient precision to determine whether

$$|z_0 - \tau|$$

is sufficiently small. □

All that now remains is to determine which $\tau$ we should be testing. In order to obtain our candidate $\tau$, we develop the continued fractions of the real part and the square of the imaginary part of our computed inverse $z_0$. As the discriminants under consideration are bounded by $|D|$, it may easily be shown by means of Liouville's inequality that the distance between the real parts of any two such quadratic irrationals is at least

$$19^{-1}d^{-4}(\log H)^{-4},$$

and the distance between any two of the squares of the imaginary parts is at least

$$19^{-2}d^{-8}(\log H)^{-8}.$$

So we develop the continued fractions of the real part of $z_0$ and the square of the imaginary part of $z_0$ until their difference from the convergents is less than

$$19^{-3}d^{-8}(\log H)^{-8}.$$

Once this holds of the convergents, they will form the only possible quadratic irrational inverse of $j$. If the height of the real convergent $c_r$ is greater than $\frac{d^2(\log H)^2}{9.7}$, or the height of the square of the imaginary convergent $c_i$ is greater than $\frac{d^4(\log H)^4}{90}$, or the square root of $c_i$ is a rational number, then we may conclude that $j$ is not a singular modulus. Otherwise, letting $\tau = c_r + i\sqrt{c_i}$, we may test if $|z_0 - \tau|$ satisfies our condition for $j = j(\tau)$. If so, then $j$ is a singular modulus, and otherwise $j$ is not a singular modulus for $\tau$ of discriminant $|D| \geq 16$. It is clear that the computational

complexity of the calculation of the convergents and other operations in this algorithm is dominated by that of inverting $j$, so letting $T = d^2 \log H (\log d + \log\log H)^2$ we obtain a running time of

$$O(M(T)\log T).$$

For testing discriminants with $|D| \leq 16$, we may simply take the list of all $\tau$ of discriminant $\leq 16$, and test whether $j$ is equal to $j(\tau)$. The degrees of $j(\tau)$ for such $\tau$ are bounded by 2, and their Mahler measures are bounded by $3 \cdot 10^6$, so if $j \neq j(\tau)$,

$$|j - j(\tau)| \geq 2^{-2d} H^{-2d} \left(3 \cdot 10^6\right)^{-d} \geq \exp(-2d(33 + \log H)).$$

So with an approximation $\tilde{j}$ to $j$ of absolute precision $2^{-4d(33+\log H)-2}$, we will be able to establish whether $j = j(\tau)$. By the algorithm to compute $j$ of [31], which requires time $O(M(P)\log P)$ for a relative precision of $P$ bits, as $|j(\tau)| \leq 3 \cdot 10^6$, computing $j(\tau)$ for each of the $\tau$ to the required absolute precision is possible in time

$$O(M(d\log H)(\log d + \log\log H)),$$

which again is $O\left(M(T)\log T\right)$.

# Chapter 5

# Analytically computing the rational torsion of an elliptic curve in quasilinear time

In this chapter we present an algorithm to compute the rational torsion subgroup of an elliptic curve

$$E : y^2 = 4x^3 + Ax + B, \quad A, B \in \mathbb{Z}$$

based on the evaluation of its corresponding Weierstrass elliptic function at division points of the lattice of $E$.

The algorithm is based on that of Doud [28], which computed values of Weierstrass elliptic functions by a $q$-series which converges linearly. Here we replace this with a computation based on Jacobi's theta functions,

$$\wp(z,\tau) = \pi^2\theta_{00}(0;\tau)^2\theta_{10}(0;\tau)^2\frac{\theta_{01}(z;\tau)^2}{\theta_{11}(z;\tau)^2} - \frac{\pi^2}{3}\left(\theta_{00}(0;\tau)^4 + \theta_{01}(0;\tau)^4\right). \tag{5.1}$$

By a recent result Labrande [50], these functions may be computed to an absolute precision of $P$ bits in $O(M(P)\log P)$ time. It is noted there that their algorithm may be used to compute Weierstrass elliptic functions, but as it computes these to absolute precision, and the expression for $\wp$ we make use of involves a division by $\theta_{11}(z;\tau)$, it must be bounded away from zero in order to compute its reciprocal to a given precision. As we are computing $\wp$ at rational torsion points, we are able to furnish such a bound, and obtain the following theorem.

**Theorem 5.1.** *Let $E : y^2 = 4x^2 + Ax + B$ be a rational elliptic curve, with $A, B$ integers. Letting $C = \max\{|A|, |B|\}$, the torsion subgroup of $E(\mathbb{Q})$ may be determined in time*

$$O\left(M(\log C)\log\log C\right),$$

*where $M(P)$ is the time complexity of multiplication of numbers $P$ bits in length.*

It was recently shown by Harvey and Hoeven [38] that the time complexity of multiplication of two $P$-bit numbers is $O(P \log P)$, so this may be taken to be $O(\log C(\log\log C)^2)$. In comparison to our result, the algorithm of Doud [28] has a time complexity of $O\left(M(\log C)\log C\right)$. The Nagell-Lutz theorem yields an algorithm dependent on the complexity of factorizing of $\Delta$, which is expected to be subexponential in $C$. The method of division polynomials, which uses the fact that the torsion points of particular orders are all roots of certain explicit polynomials, has time complexity $O((\log C)^2)$, and there is a more recent algorithm of García-Selfa, Olalla, and Tornero [34] based on the fact that rational elliptic curves with torsion points of particular orders lie in certain one-parameter families of curves, the so-called Tate normal form, which also has a time complexity of $O((\log C)^2)$.

We will denote by $\mathcal{F}$ the standard fundamental domain of $j$,

$$\mathcal{F} = \left\{ z \Big| -\frac{1}{2} < \mathrm{Re}(z) \leq \frac{1}{2}, |z| > 1 \right\} \cup \left\{ z \Big| 0 \leq \mathrm{Re}(z) \leq \frac{1}{2}, |z| = 1 \right\}.$$

Given $\omega_1, \omega_2$, we will evaluate $\wp(z, \omega_1, \omega_2)$ via the following transformation: finding $g \in \mathrm{SL}_2(\mathbb{Z})$ such that $\tau := g\omega_1/\omega_2 \in \mathcal{F}$, we have

$$\wp\left(z, \omega_1, \omega_2\right) = \frac{\wp\left(\frac{z}{c\omega_1 + d\omega_2}, g\tau, 1\right)}{(c\omega_1 + d\omega_2)^2}. \tag{5.2}$$

## 5.1 Proof of the theorem

Jacobi's four theta functions (c.f. [21]) may be defined as follows:

$$\begin{aligned}
\theta_{00}(z;\tau) &= \sum_{n=-\infty}^{\infty} \exp(\pi i n^2 \tau + 2\pi i n z), \\
\theta_{01}(z;\tau) &= \theta_{00}\left(z + \frac{1}{2}; \tau\right), \\
\theta_{10}(z;\tau) &= \exp\left(\frac{\pi i \tau}{4} + \pi i z\right) \theta_{00}\left(z + \frac{\tau}{2}; \tau\right), \\
\theta_{11}(z;\tau) &= \exp\left(\frac{\pi i \tau}{4} + \pi i \left(z + \frac{1}{2}\right)\right) \theta_{00}\left(z + \frac{\tau}{2} + \frac{1}{2}; \tau\right).
\end{aligned}$$

We also make use of the following relation:

$$\theta_{00}(z + b\tau; \tau) = \exp\left(-\pi i b^2 \tau - 2\pi i b z\right) \theta_{00}(z;\tau). \tag{5.3}$$

The following lemma follows directly from the $q$-series for the Jacobi theta functions.

**Lemma 5.1.** *For $\tau \in \mathcal{F}$,*

$$\begin{aligned} 0.8 \leq |\theta_{00}(0;\tau)| &\leq 1.2, \\ 0.8 \leq |\theta_{01}(0;\tau)| &\leq 1.2, \\ \exp\left(-\frac{\pi \mathrm{Im}(\tau)}{4}\right) \leq |\theta_{10}(0;\tau)| &\leq 1.1, \end{aligned}$$

*and for $z$ in the Euclidean parallelogram defined by $0, \frac{\tau}{2}, 1, 1+\frac{\tau}{2}$,*

$$|\theta_{01}(z;\tau)| \leq 3.4.$$

We now give our first lemma to bound $|\theta_{11}(z;\tau)|$ below.

**Lemma 5.2.** *If $\tau \in \mathcal{F}$, and $z$ lies in the Euclidean parallelogram with vertices $0$, $\frac{\tau}{2}$, $1$, $1+\frac{\tau}{2}$, at least one of $\theta_{01}(z;\tau)$ or $\theta_{11}(z;\tau)$ is bounded below in absolute value by $0.47$.*

*Proof.* Firstly, if $-\frac{\mathrm{Im}(\tau)}{4} \leq \mathrm{Im}(z) \leq \frac{\mathrm{Im}(\tau)}{4}$, then

$$\begin{aligned} |\theta_{00}(z;\tau)|, |\theta_{01}(z;\tau)| &\geq 1 - 2\sum_{n=1}^{\infty} \exp\left(-\pi n^2 \mathrm{Im}(\tau) + 2\pi n \mathrm{Im}(z)\right) \\ &\geq 1 - 2\sum_{n=1}^{\infty} \exp\left(-\frac{\pi n^2}{2}\mathrm{Im}(\tau)\right) \\ &\geq 1 - 2\sum_{n=1}^{\infty} \exp\left(-\frac{\pi n^2 \sqrt{3}}{4}\right) \\ &\geq 0.47. \end{aligned}$$

If $\frac{\mathrm{Im}(\tau)}{4} \leq \mathrm{Im}(z) \leq \frac{\mathrm{Im}(\tau)}{2}$, then as

$$\begin{aligned} \theta_{11}(z;\tau) &= \exp\left(\frac{\pi i \tau}{4} + \pi i\left(z+\frac{1}{2}\right)\right)\theta_{00}\left(z+\frac{\tau}{2}+\frac{1}{2};\tau\right) \\ &= \exp\left(\frac{\pi i \tau}{4} - \pi i z - \frac{\pi i}{2}\right)\theta_{00}\left(z-\frac{\tau}{2}+\frac{1}{2};\tau\right), \end{aligned}$$

where the second line follows from equation 5.3, we have the lower bound

$$|\theta_{11}(z;\tau)| \geq 0.47.$$

□

As in [28], pg. 467, letting $x = \wp(z_0, \omega_1, \omega_2)$, if $(x, y)$ is a rational torsion point of $E$, then $(4x, 4y)$ is a rational torsion point of the isomorphic elliptic curve

$$y^2 = x^3 + 4Ax + 16B.$$

By the Nagell–Lutz theorem, $16y^2$ divides its discriminant, so $y^2 \leq |16^2(A^3 + 27B^2)|$. As $x$ is a root of the equation $4x^3 + Ax + B - y^2 = 0$,

$$|\wp(z_0, \omega_1, \omega_2)| \leq \max\{|A|, |B| + 16^2(|A^3 + 27B^2|)\},$$

and we will only consider those with $\wp(z_0, \omega_1, \omega_2) \neq 0$, as determining the torsion points $(0, y)$ is elementary in time $O(M(|B|)\log|B|)$. We let $C = \max\{|A|^3, 28|B|^2\}$, so that torsion points are bounded by $16^2C$, and the absolute value of the discriminant of $E$ is $\leq 2C$. The $j$-invariant of $E$ is $\frac{1728A^3}{\Delta}$, which is bounded in absolute value by $1728C$. By Lemma 1 of [13],

$$|j(\tau) - e^{-2\pi i\tau}| \leq 2079,$$

we obtain a bound in the imaginary part of $\tau$,

$$\mathrm{Im}(\tau) \leq \log C + 8.1,$$

with which we may prove the following lemma.

**Lemma 5.3.** *$c\omega_1 + d\omega_2$ is bounded in absolute value below by*

$$\exp\left(-0.61 \log C - 4.9\right),$$

*and above by* 6.

*Proof.* Letting $\Delta$ be the discriminant of $E$ and noting that

$$\Delta = \Delta(\omega_1, \omega_2) = \Delta(c\omega_1 + d\omega_2, a\omega_1 + b\omega_2)$$

(as $g \in \mathrm{SL}_2(\mathbb{Z})$), by [21], pg. 69, we have the following expression for $(c\omega_1 + d\omega_2)^3$:

$$(c\omega_1 + d\omega_2)^3 = \frac{2\pi^3}{\Delta^{1/4}}\theta_{00}(0;\tau)^2\theta_{01}(0;\tau)^2\theta_{10}(0;\tau)^2.$$

Substituting in the bounds $|\Delta| \geq 1$ and $|\theta_{ij}(0;\tau)| \leq 1.2$, we obtain the upper bound 6, and as in [28] pg. 468, we have the lower bound

$$\begin{aligned} |c\omega_1 + d\omega_2| &\geq \frac{\pi}{2|\Delta|^{1/12}} e^{-\pi \mathrm{Im}(\tau)/6} \\ &\geq \exp\left(-0.61 \log C - 4.9\right). \end{aligned}$$

□

We now bound $|\theta_{11}(z;\tau)|$ below at rational torsion points.

**Lemma 5.4.** *Let $(\wp(z_1,\omega_1,\omega_2),y)$ be a rational torsion point of $E$, and $z_0$ an element in the fundamental domain of $\wp(z,\tau)$ equivalent to $\frac{z_1}{c\omega_1+d\omega_2}$ modulo the lattice $\langle\tau,1\rangle$ and negation, which satisfies $-1/2 \leq \mathrm{Re}(z_0) \leq 1/2$ and $0 \leq \mathrm{Im}(z_0) \leq \mathrm{Im}(\tau)/2$. Then*

$$|\theta_{11}(z_0;\tau)| \geq \exp\left(-1.9\log C - 18\right).$$

*Proof.* Firstly, by the relations $\wp(z\pm\tau) = \wp(z\pm 1) = \wp(-z) = \wp(z)$, $\wp(z_0,\tau)$ is equal to $\wp\left(\frac{z_1}{c\omega_1+d\omega_2},\tau\right)$. By Lemma 5.2 either there is a lower bound on $|\theta_{11}(z_0;\tau)|$ of 0.47, in which case we are done, or we have a lower bound on $|\theta_{01}(z_0,\tau)|$ of 0.47, from which we may deduce, by equations 5.1 and 5.2, and the bounds of Lemma 5.1 and Lemma 5.3,

$$\begin{aligned}|\theta_{11}(z_0;\tau)|^2 &\geq \frac{\pi^2|\theta_{00}(0;\tau)|^2|\theta_{10}(0;\tau)|^2}{16^2C|c\omega_1+d\omega_2|^2 + \frac{\pi^2}{3}\left(|\theta_{00}(0;\tau)|^4 + |\theta_{01}(0;\tau)|^4\right)} \cdot 0.47^2 \\ &\geq \exp\left(-3.8\log C - 34.6\right),\end{aligned}$$

which yields the lemma. □

These various bounds now allow us to calculate $\wp(z_1,\omega_1,\omega_2)$ to the required precision — one may calculate the following result.

**Proposition 5.1.** *Let $z_0$ be the element in the fundamental domain of $\wp(z,\tau)$ equivalent to $\frac{z_1}{c\omega_1+d\omega_2}$ with $0 \leq \mathrm{Im}(z) \leq \frac{\mathrm{Im}(\tau)}{2}$ modulo the lattice generated by 1 and $\tau$, and the relation $\wp(-z) = \wp(z)$. Suppose that*

$$|\theta_{11}(z_0;\tau)| \geq \exp\left(-1.9\log C - 18\right).$$

*Then calculating approximations to $\pi$, $\theta_{00}(0;\tau)$, $\theta_{10}(0;\tau)$, $\theta_{01}(z_0;\tau)$, $\theta_{11}(z_0;\tau)$, and $c\omega_1+d\omega_2$ of absolute precision*

$$\exp(-30\log C - 300)$$

*allows the calculation of $\wp(z_1,\omega_1,\omega_2)$ to absolute precision $1/8$ via the expression*

$$\wp(z_1,\omega_1,\omega_2) = \frac{1}{(c\omega_1+d\omega_2)^2}\left(\pi^2\theta_{00}(0;\tau)^2\theta_{10}(0;\tau)^2\frac{\theta_{01}(z_0;\tau)^2}{\theta_{11}(z_0;\tau)^2} - \frac{\pi^2}{3}\left(\theta_{00}(0;\tau)^4 + \theta_{01}(0;\tau)^4\right)\right).$$

*Proof of the theorem.* We first calculate the various parameters $\omega_1, \omega_2$, $\tau, c, d$, and $\pi$, from which we may calculate $z_0$, and thence $\wp(z_1, \omega_1, \omega_2)$.

The calculation of $\pi$ to the required precision may be performed in time $O(M(\log C) \log\log C)$ by the Gauss–Legendre algorithm. Likewise the required approximations to $\omega_1$ and $\omega_2$ may be performed in time $O(M(\log C) \log\log C)$ by Algorithm 7.4.7 of [24], pg. 391.

For the calculation of $g$ and $\tau$, as the $\omega_1, \omega_2$ computed satisfy $\mathrm{Re}(\omega_1/\omega_2) = 0$ or $1/2$, if it is equal to zero, and does not already lie in the fundamental domain, then we may take $-\omega_2/\omega_1$ as our $\tau$, and otherwise, if $\mathrm{Im}(\omega_1/\omega_2) \geq 1/2$, then $\omega_1/\omega_2$ lies in one of four copies of $\mathcal{F}$ under $\mathrm{SL}_2(\mathbb{Z})$, and finding $g$ takes time $O(M(\log C))$. Letting $\tau_0 = \omega_1/\omega_2$, the process of taking the negative of the inverse of $\tau_0$ if $|\mathrm{Re}(\tau_0)| \leq 1/2$, and otherwise replacing $\tau_0$ with $\tau_0 - [\mathrm{Re}(\tau_0) + 1/2]$, and then updating $g$ as appropriate will find $g$ such that $g\tau_0 \in \mathcal{F}$. On inverting $\tau_0$ when $\mathrm{Im}(\tau) \leq 1/2$, $\mathrm{Im}(-1/\tau_0) \geq 2\mathrm{Im}(\tau_0)$, so the required number of steps is $O(-\log \mathrm{Im}(\tau_0))$. The modular discriminant $\Delta(\tau)$ decays exponentially as $\tau_0$ moves towards the cusp at $1/2$, and by Proposition 5.2 of [68] the entries of $g$ grow only polynomially. So $-\log \mathrm{Im}(\tau_0) = O(\log\log C)$, and the time required for finding $g$ is then $O(M(\log C) \log\log C)$. We note that if a $\tau_0$ we calculate is sufficiently close to the boundary of the fundamental domain (within the precision), we will not be able to decide whether $\tau_0$ is within the standard fundamental domain — at this point we would output $\tau$ and $g$ corresponding to $\tau_0$. However, such a $\tau$ would be sufficiently close to the fundamental domain that this will not affect our computations.

In order to calculate $z_0$, we first compute $\frac{z_1}{c\omega_1 + d\omega_2}$. By the relations

$$\wp(z_0 \pm 1, \tau) = \wp(z_0 \pm \tau, \tau) = \wp(z_0, \tau),$$
$$\wp(-z_0, \tau) = \wp(z_0, \tau),$$

we may then compute $z_0$ satisfying the conditions of [50]. We now consider the running time of these calculations. Firstly, $z_1$ is bounded in absolute value by $|\omega_1 + \omega_2|$. Algorithm 7.4.7 of [24] computes $\omega_1$ and $\omega_2$ as a constant divided by $\mathrm{AGM}(\alpha, \beta)$ (the arithmetic-geometric mean), where $\alpha$ and $\beta$ are positive, and bounded below in terms of the root separation of our elliptic curve. For example, in the disconnected case, letting $e_1 > e_2 > e_3$ be the (real) roots of $4x^3 + Ax + B$,

$$\omega_2 = \frac{2\pi i}{\mathrm{AGM}(\sqrt{e_1 - e_3}, \sqrt{e_2 - e_3})}.$$

The absolute value of this is then bounded by $\exp(c \log C)$ by standard root separation results (one could even use Liouville's inequality). The absolute value of $\omega_1$ is bounded

similarly, as are $\omega_1, \omega_2$ in the connected case. Furthermore, by Lemma 5.3, $c\omega_1 + d\omega_2$ is bounded below in absolute value by $\exp(-0.61 \log C - 4.9)$, so $\left|\frac{z_1}{c\omega_1 + d\omega_2}\right| \leq \exp(c' \log C)$. We may therefore calculate $z_0$ to the required absolute precision in time $O(M(\log C))$.

The algorithm of [50] computes $\theta_{ij}(z;\tau)$, for $\tau \in \mathcal{F}$, $|\mathrm{Re}(z)| \leq 1/2$, and $0 \leq \mathrm{Im}(z) \leq \frac{\mathrm{Im}(\tau)}{2}$, to an absolute precision of $P$ bits in time $O(M(P) \log P)$. We note that if $|\theta_{11}(z_0;\tau)|$ is smaller than the bound of the hypothesis of Proposition 5.1, then by Lemma 5.4, the point is not one of rational torsion, and may be rejected, and that the calculation of the theta terms may be done in time $O(M(\log C) \log \log C)$.

By the isomorphism of $E : y^2 = 4x^3 + Ax + B$ with $E' : y^2 = x^3 + 4Ax + 16B$, if $(x, y)$ is a torsion point of $E$, then $(4x, 4y)$ is a torsion point of $E'$, and so has integer coordinates, so a precision of $1/8$ is sufficient to determine the unique candidate torsion point. Having calculated the $x$-coordinate, and rejecting the point if $|x| \geq 16^2 C$, the potential $y$-coordinate may be computed to precision $1/8$ by the extraction of the square root of $4x^3 + Ax + B$, in time $O(M(\log C) \log \log C)$. If the corresponding integer point $(4x, 4y)$ lies on $E'$, we then may test whether the point is of finite order, for which we need only test up to order 12, by Mazur's theorem [60], and reject the point if it is not of finite order. This takes time $O(M(\log C))$.

As in [28], the basis obtained by the algorithm of [24], pg. 391 has the property that multiples of $\omega_1$ correspond to the infinite component of $E$, and translates of these by $\omega_2/2$ correspond to the bounded component of $E$, if it exists. So we determine whether the division points $\omega_1/n$, $\omega_1/n + \omega_2/2$, $n \leq 12$ indeed correspond to points of rational torsion, and a finite computation will yield the group structure (including here the points $(0, y)$), and this requires a constant number of steps of time complexity $O(M(\log C) \log \log C)$. $\square$

# Chapter 6

# An algorithm to combine linear congruences

In this chapter we describe an improvement to algorithms for determining numbers satisfying congruence conditions modulo primes. The first automatic method to combine linear congruences was a mechanical device built by D.H. Lehmer, described in [55], made of a common axle turned by a motor, with bike chains with coprime numbers of links hanging from the axle, with each link carrying a pin or not depending on the congruence conditions modulo the length of its chain. The pin actuates a circuit breaker when it reaches the top of the common axle, and if all of the links at the top have pins, then the circuit powering the motor is cut, and the number satisfying the congruence conditions can be retrieved. There were further mechanical devices made, and when electronic computers became available, algorithms commonly split the domain under consideration into congruence classes. For example if the modular tests are $\left(\frac{n}{p}\right) = 1$ for $p$ prime and $\geq 3$, then numbers may be tested by considering separately the two congruence classes $1 + 15n$ and $4 + 15n$, speeding up the computation $15/2$ times.

Bernstein [8] introduced a new method, doubly-focused enumeration, which makes use of the Chinese remainder theorem to pre-sieve by many primes. In order to test $n < H$, we take two products of primes $P_1, P_2$, $P_1 P_2 \geq H$, where this product is taken much larger than $H$. Then the numbers $a_1$ modulo $P_1$ which satisfy the modular conditions modulo each prime dividing $P_1$ are precomputed, and similarly for $P_2$. By the Chinese remainder theorem, every number up to $P_1 P_2$ which satisfies the modular conditions will be equal to $a_1 n_1 + a_2 n_2 \mod P_1 P_2$ for some precomputed $a_1, a_2$, and $n_1, n_2$ associated to the Chinese remainder theorem for $P_1$ and $P_2$. This allows for a much greater speedup than separating into congruence classes — indeed it is testing, as opposed to many numbers in various congruence classes, one number

in each admissible congruence class modulo $P_1P_2$ which is $< H$. As an example, if the modular tests are again $\left(\frac{n}{p}\right) = 1$ for primes $p \geq 3$, and we wish to test numbers up to $10^{11}$, we can take $P_1$ to be the product $3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$, and $P_2$ the product $23 \cdot 29 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47$, then $3 \cdot 10^{17} \approx P_1P_2 > 10^{11}$, and the number of numbers which remain to be tested is roughly $10^{11} \cdot \prod_{i=2}^{14} \frac{p_i - 1}{2p_i} \approx 1.1 \cdot 10^6$, a speedup of roughly 60000 times. Our algorithm is based on the doubly-focused enumeration algorithm of Sorenson [81].

Our novelty is to test, as opposed to each number individually, a bit-vector $C$ corresponding to many numbers at once: letting $a_b$ correspond to the $b$'th bit of $C$, if $v_{i,j}$ is a bit-vector where the $b$'th bit is 1 if $\left(\frac{j - a_0 + a_b}{p_i}\right) \neq 1$, and zero otherwise, then bitwise-ORing $C$ with $v_{i,a_0 \mod p_i}$ will simultaneously test all $a_b$, and each test with a $p_i$ will eliminate (by setting its corresponding bit to 1) roughly $1/2$ of the remaining $a_b$ in one hardware operation. With for example 512 bits, this will require roughly 12 tests to eliminate 512 numbers, as opposed to the average of roughly two tests to eliminate any particular number tested directly with $\left(\frac{n}{p_i}\right)$ — a speedup of around 42 times, however practical concerns do limit this to a certain extent. We test our algorithm against that of Sorenson [81], and achieve a roughly 50 times speedup with $B = 512$ on an Intel Xeon Gold 6140 (2017) compared to an AMD Opteron 8354 (2007). As their code is private, we are not able to perform this comparison on the same hardware, so some of this improvement is due to increased per-core performance. Benchmark comparisons between these two processors indicate a roughly 50% improvement in performance per thread for the Intel Xeon Gold 6140 over the AMD Opteron 8354.

Unfortunately our algorithm only allows for a small amount of "double focusing". The product $P_1P_2$ may be taken up to around 100 times larger than the size of the search window before this starts to become counterproductive. As such our algorithm is closer to a very efficient singly focused algorithm than Bernstein's doubly focused algorithm. It would be interesting to try to improve the amount of double focusing possible with our algorithm.

We will apply this algorithm in the subsequent chapter to give a new bound for negative discriminants with one class of binary quadratic forms in each genus.

We note the following application to finding pries in patterns, specifically Cunningham chains, recently obtained from a variation of this algorithm. Cunningham chains of the first kind are chains of primes $p_i$ such that $p_{i+1} = 2p_i + 1$, and Cunningham chains of the second kind are chains of primes $p_i$ such that $p_{i+1} = 2p_i - 1$.

**Theorem 6.1.** *The smallest Cunningham chain of the first kind of length 18 begins at*

$$p = 776208125345634679522109.$$

*the smallest Cunningham chain of the second kind of length 17 begins at*

$$p = 8058978143544782179441.$$

*The smallest and second smallest Cunningham chains of the second kind of length 18 begin at*

$$p = 773377687199227745520601$$

*and*

$$p = 868594057941691288178401$$

*respectively.*

Our Cunningham chain of the first kind of length 18 is the first to be found, and our chains of the second kind of lengths 17 and 18 are not the first to be found of this length and type, but are the smallest. We note that finding the *smallest* chain of a given length is harder than finding *any* chain of that length, for example the previous smallest known Cunningham chain of the second kind of length 18 began at

$$117302256313688977973793781,$$

found by Jaroslaw Wroblewski, which is a 27 digit number, as opposed to our chains, which begin with numbers of 24 digits. This type of algorithm is highly effective at exhaustively finding chains, as long chains yield dense modular conditions (for example, this approach has little value for finding primes in arithmetic progression, as the common difference of length $k$ chains must be divisible by $\prod_{p<k} p$, which yields only one inadmissible value for the first value of such a chain for primes $< k$). In comparison to the method of Sorenson and Webster [82] (which also implements a bit-packing method of a different type), our algorithm is roughly 1000 times faster.

## 6.1 The algorithm

Let $T(a, p^\alpha)$ hold if and only if $a$ lies in a particular set of congruence classes modulo $p^\alpha$, and $T(a, Q) = \bigwedge_{p^\alpha \| Q} T(a, p^\alpha)$. We fix $B$ to be a number corresponding to the bit-length of some hardware variable which may be readily bitwise-ORed. Further let $H_1 < H_2$ be the two positive numbers between which we wish to test our modular conditions.

1. We first take coprime products of prime powers $P_1, \ldots, P_k$, and for each $1 \leq i \leq k$, determine and list those $0 \leq n < P_i$ such that $T(n, P_i)$ holds, and let $P = \prod_{i=1}^{k} P_i$.

2. Compute the parameters of the Chinese remainder theorem $M_i$ such that if $a \equiv n_i \mod P_i$, then $a \equiv \sum_{i=1}^{k} n_i M_i \frac{P}{P_i} \mod P$. These $M_i$ satisfy, for some $m_i$, $M_i \frac{P}{P_i} + m_i P_i = 1$. In particular, if $a$ satisfies $T(a, P_i)$, then as the $P_j$ are coprime, $a \equiv n_i M_i \frac{P}{P_i} \equiv n_i \mod P_i$, so if $a$ satisfies $\bigwedge_{i=1}^{k} T(a, P_i)$, the values of $a \mod P_i$ must be among our previously computed admissible values mod each $P_i$, so every $0 \leq a < P$ satisfying $\bigwedge_{i=1}^{k} T(a, P_i)$ is equal to $\sum_{i=1}^{k} n_i M_i \frac{P}{P_i} \mod P$, for some $n_i$, $1 \leq i \leq k$ which are elements of our precomputed admissible values modulo $P_i$.

3. Compute, for each $P_i$, $n_j M_i$, and sort these into an increasing list $L_i$, letting the length of each be $l_i$. We denote the $j$th entry of the $i$th list by $r_{i,j}$, indexed from $j = 0$. For $L_1$, append to this list copies of $r_{1,l_1}$ so that $l_1$ is divisible by $B$.

4. For two further sets of $k', k''$ coprime products of prime powers $Q_1, \ldots, Q_{k'}$, $R_1, \ldots, R_{k''}$, we compute the $0 \leq n < Q_i$ such that $T(n, Q_i)$ holds, and similarly for $R_i$ — we will use these to test the remaining values mod $P$. The $Q_i$ are distinguished from the $R_i$ as we will perform some precomputation to speed up the testing by $Q_i$.

5. Compute, for each $1 \leq i \leq k$ and $0 \leq j < l_i$, and each $Q_m$, $1 \leq m \leq k'$, the number $a$ such that $0 \leq a < Q_m$ and $a \equiv r_{i,j} \mod Q_m$, and set $Q_m(r_{i,j}) = a$. Also compute $Q_m(P)$ similarly.

6. Begin loop over $0 \leq n_1 < l_1/B$.

7. Now we implement the key novelty of this algorithm. For each $Q_i$ we compute, for each $0 \leq j < Q_i$, a special bit-vector $v_{i,j}$ of length $B$, initialized to 0, as follows: for each $0 \leq b < B$, we compute the modular value of $j + r_{1,Bn_1+b} - r_{1,Bn_1}$, and if $T(j + r_{1,Bn_1+b} - r_{1,Bn_1}, Q_i)$ does not hold, we set the $b$th bit of $v_{i,j}$ to be 1, and similarly with $R_i$, denoting the corresponding bit-vectors $w_{i,j}$. As this is done in the outer loop of our algorithm, it does not contribute significantly to the running time.

8. Begin loop over $0 \leq n_2 < l_2, \ldots, 0 \leq n_{k-1} < l_{k-1}$.

9. Compute $r_{1,Bn_1} + \sum_{i=2}^{k-1} r_{i,n_i}$ modulo each $Q_j$ — let these be $q_j$.

10. Compute $A_1 = r_{1,Bn_1} + \sum_{i=2}^{k-1} r_{i,n_i}$ and $A_2 = r_{1,Bn_1+B-1} + \sum_{i=2}^{k-1} r_{i,n_i}$, and then compute $\epsilon = \max\{m | A_1 - mP \geq 0\}$. Then compute three ranges $[e_1, e_2]$, $[f_1, f_2]$, and $[g, l_k]$ such that:

    (a) $e_1$ is minimal such that $A_2 + r_{k,e_1} \geq \epsilon P + H_1$, and $e_2$ is maximal such that $A_1 + r_{k,e_2} \leq \epsilon P + H_2$;

    (b) $f_1$ is minimal such that $A_2 + r_{k,f_1} \geq (\epsilon + 1)P + H_1$, and $f_2$ is maximal such that $A_1 + r_{k,f_2} \leq (\epsilon + 1)P + H_2$;

    (c) $g$ is minimal such that $A_2 + r_{k,g} \geq (\epsilon + 2)P + H_1$ if such a $g$ exists (usually there is no such $g$).

11. For $e_1 \leq n_k \leq e_2$: let $C$ be a bit-vector initialized to 0. We denote the bit-vector of length $B$ with all entries 1 by $\mathbf{1}$.

    (a) For each $Q_i$, calculate $a_i \equiv q_i + Q_i(r_{n_k}) - \epsilon Q_i(P) \mod Q_i$, and bitwise-OR $v_{i,a_i}$ with $C$, and if $C = \mathbf{1}$, restart the loop with $n_k + 1$.

    (b) Calculate $A = r_{1,Bn_1} + \sum_{i=2}^{k} r_{i,n_i}$. We note that our precomputation of the $q_i$ avoided the necessity of computing $A$ and its values modulo each $Q_i$ to test with $Q_i$.

    (c) For each $R_i$, calculate $a_i \equiv A - \epsilon P \mod R_i$, and bitwise-OR $w_{i,a_i}$ with $C$, and if $C = \mathbf{1}$, restart the loop with $n_k + 1$. We note that owing the generally large number of $R_i$ we wish to test with, and the large sizes of the lists $L_j$, it is infeasible memory-wise to precompute the values of $r_{j,j'} \mod R_i$ in order to speed up this step by avoiding the modulus operation applied to the large number $A - \epsilon P$.

    (d) If we have reached this step, then $C \neq \mathbf{1}$, so there are some $r_{n_1 B+b} + \sum_{i=2}^{k} r_{i,n_i} - \epsilon P$ which have survived all testing. For $0 \leq b < B$, we test if the $b$'th bit of $C$ is equal to zero — this corresponds to a number which has survived testing. We then compute $a = r_{n_1 B+b} + \sum_{i=2}^{k} r_{i,n_i} - \epsilon P$ — if $a \leq H_1$ or $a \geq H_2$, then it is an extraneous value and may be discarded. Otherwise, if $a$ satisfies all further tests we wish to perform, we output $a$ as a number which has survived testing, and store it in an array to be saved to disk.

12. For $f_1 \le n_k \le f_2$: Perform the steps of the loop over $e_1 \le n_k \le e_2$ with $\epsilon$ replaced by $\epsilon + 1$.

13. For $g \le n_k \le l_k$: Perform the steps of the loop over $e_1 \le n_k \le e_2$ with $\epsilon$ replaced by $\epsilon + 2$, if $g$ exists.

*Demonstration of validity.* Suppose that $0 \le H_1 \le a \le H_2 < P$ and $T(a, P)$ holds. Then $a = r_{1,Bn_1+b} + \sum_{i=2}^{k} r_{i,n_i} - \epsilon' P$ for some $b$, $\epsilon'$, and $(n_1, \ldots n_k)$. Let $A_1 = r_{1,Bn_1} + \sum_{i=2}^{k-1} r_{i,n_i}$ and $A_2 = r_{1,Bn_1+B-1} + \sum_{i=2}^{k-1} r_{i,n_i}$, and $\epsilon = \max\{m | A_1 - mP > 0\}$. As $r_{k,n_k} \le P$ and $|r_{1,Bn_1} - r_{1,Bn_1+b}| \le P$, $\epsilon \le \epsilon' \le \epsilon + 2$. If $\epsilon' = \epsilon$, as the $r_{i,j}$ are increasing in $j$,

$$H_1 + \epsilon P \le a + \epsilon P \le A_2 + r_{k,n_k}, \text{ and}$$
$$A_1 + r_{k,n_k} \le a + \epsilon P \le H_2 + \epsilon P,$$

so $e_1 \le n_k \le e_2$. If $\epsilon' = \epsilon + 1$, then

$$H_1 + (\epsilon + 1)P \le a + (\epsilon + 1)P \le A_2 + r_{k,n_k}, \text{ and}$$
$$A_1 + r_{k,n_k} \le a + (\epsilon + 1)P \le H_2 + (\epsilon + 1)P,$$

so $f_1 \le n_k \le f_2$. If $\epsilon' = \epsilon + 2$, then

$$H_1 + (\epsilon + 2)P \le a + (\epsilon + 2)P \le A_2 + r_{k,n_k},$$

so $g \le n_k \le l_k$. We now show that the iteration of loops $(n_1, \ldots n_k)$ does indeed test $a$ in a valid manner. Supposing $\epsilon' = \epsilon$, for step (a), the $a_i$ calculated is $a_i = A_1 + r_{k,n_k} - \epsilon P \mod Q_i$, so as $a = A_1 - r_{1,Bn_1} + r_{1,Bn_1+b} + r_{k,n_k} - \epsilon P$, letting $j = a_i$, the value of $a$ mod $Q_i$ is $a_i - r_{Bn} + r_{Bn+b}$, so (by step 7) the $b$'th bit of $v_{i,j}$ is 1 if $T(a, Q_i)$ does not hold, and 0 if it does hold. So if $a$ fails the test mod $Q_i$, the $b$'th bit is set to one, and as all operations on $C$ are bitwise-ORs, it will remain 1 through the remainder of step (a) and through step (c) — then in step (d), the $b$'th bit of $C$ is one, so $a$ is not computed, nor is it output as a survivor. If $a$ passes all tests mod the $Q_i$, the $b$'th bit of $C$ will remain 0 through step (a). Step (c) is similar, as are the cases $\epsilon' = \epsilon + 1$ and $\epsilon' = \epsilon + 2$. Step (d) computes $a$ itself if the $b$'th bit of $C$ has remained 0, and performs any remaining tests, so $a$ is output if it passes all tests, and not output if it fails any of the further tests, or tests modulo the $Q_i$ or $R_i$. □

*Remark.* Each bitwise OR simultaneously tests $B$ candidate numbers $0 \le a < P$, and on average will eliminate a candidate with the probability

$$\frac{\#\{T(a, Q_i) \text{ does not hold } | 0 \le a < Q_i\}}{Q_i},$$

and similarly for the $R_i$. For example, in the case of pseudosquares (the $m$th pseudosquare problem is to find the minimal non-square $N$ such that $N \equiv 1 \mod 8$ and $\left(\frac{N}{p_i}\right) = 1$ for $2 \leq i \leq m$), each test will eliminate roughly half of the candidates surviving up to that point, so in order to eliminate all $B$ candidates, we require $O(\log B)$ tests on average. However using a larger $B$ does increase the size of the intervals $[e_1, e_2]$, $[f_1, f_2]$, and $[g, l_k]$. Tests on the pseudosquare problem show a roughly 33–50% speedup using $B = 64$ as opposed to $B = 32$, and a speedup of around 4 times using $B = 512$ over $B = 64$, making use of the 512-bit vector units available on modern processors, however these speedups seem sensitive to the parameters of the problem.

*Remark.* Increasing the sizes of the $P_i$ will reduce the number of numbers which must be tested in proportion to $P$, (for pseudosquares by roughly 1/2 for each prime used), but when $H_2 - H_1$ is too small compared to $P$, each iteration of the loops will test very few numbers, so the overhead of precomputations will increase, and the size of the ranges $[e_1, e_2]$, $[f_1, f_2]$, and $[g, l_k]$ do not decrease in exact proportion with $H_2 - H_1$.

To test the speed of this algorithm, we compare it to that of [81], which tested the range $[5 \cdot 10^{24}, 10^{25}]$, making use of doubly-focused enumeration to determine the largest currently known minimal pseudosquares. We take the following parameters: $H_1 = 5 \cdot 10^{24}$, $H_2 = 10^{25}$, $B = 512$,

$$\begin{aligned} P_1 &= 2^3 \cdot 3 \cdot 11 \cdot 19 \cdot 31 \cdot 43 \cdot 59 \cdot 71, & P_2 &= 5 \cdot 13 \cdot 23 \cdot 37 \cdot 47 \cdot 61, \\ P_3 &= 7 \cdot 17 \cdot 29 \cdot 41 \cdot 53 \cdot 67, & Q_1 &= 73 \cdot 79, \\ Q_2 &= 83 \cdot 89, & Q_3 &= 97 \cdot 101, \\ Q_4 &= 103 \cdot 107, & Q_5 &= 109 \cdot 113, \\ R_{i-30} &= p_i, \end{aligned}$$

where $p_i$ is the $i$'th prime, for $i = 31 \ldots 73$. Modulo $P_1$, there are 14387625 admissible modular values, and performing 20 parallelized iterations of the loop in $n_1$ takes an average of around 2200 seconds per iteration, so that as the full algorithm requires $[14387625/512] + 1 = 28101$ iterations of the outer loop, we estimate that the full computation would take roughly 2 core-years, in comparison to that of [81], which took, in 2009, roughly a core-century.

# Chapter 7

# Discriminants with one class of binary quadratic forms in each genus

A discriminant $d$ has one class of binary quadratic forms in each genus if and only if each class of binary quadratic forms of discriminant $d$ is ambiguous — its square in the class group is the principal form. In particular, such a $d$ has no binary quadratic forms $ax^2 + bxy + cy^2$, with $0 < b < a$ and $a$ coprime to $d$, for then $ax^2 + bxy + cy^2$ and $ax^2 - bxy + cy^2$ are inequivalent binary quadratic forms of discriminant $d$ lying in the same genus, as the number $a$ is properly represented by both forms, and so the values of the generic characters at numbers represented by the forms are the same. Gauss' principal genus theorem states that a discriminant $d$ has $2^{\omega(d)-1}$ genera, so in the case of one class per genus, the class number of $d$ is $2^{\omega(d)-1}$.

Discriminants with one class of binary quadratic forms in each genus are connected to Euler's idoneal numbers. These are numbers $n$ such that if $m$ is an odd number properly and uniquely represented by the binary quadratic form $x^2 + ny^2$, and $m$ is coprime to $n$, then $m$ is prime. Euler used these forms to produce large primes, and found 65 such numbers $n$, conjecturing that his list is complete, the largest being 1848.

By Gauss' theory of genera, those even negative discriminants $-4n$ which have one class of binary quadratic form in each genus yield idoneal numbers $n$, and the opposite implication was established by Grube [37]. Euler's form $x^2 + ny^2$ is then the principal form of the discriminant $-4n$. Gauss conjectured that if a negative discriminant has one class of binary quadratic forms in each genus, it has at most 32 genera, i.e. it has at most 6 distinct prime factors and its class number is at most 32. By Watkins' [86] determination of all discriminants of class number $\leq 100$, this

conjecture is equivalent to $-7392$ being the largest discriminant with one class of binary quadratic forms in each genus. We note that Euler's conjecture that 1848 is the largest idoneal number corresponds to $-7392$ being the largest *even* discriminant with one class of binary quadratic forms in each genus. Their finiteness was proved by Chowla [23], and Weinberger [88] proved by means of Tatuzawa's theorem that only one further such *fundamental* discriminant may exist, and that no further exist under the Generalized Riemann Hypothesis. It is further known that if $2^3 \parallel d$, and $d$ is a fundamental discriminant with one class per genus, then $4d$ will be a non-fundamental discriminant with one class per genus, and this is the only situation a further non-fundamental discriminant could have one class of binary quadratic forms in each genus. A useful survey with a historical overview may be found in [46]. Applying our algorithm to combine linear congruences yields the following theorem.

**Theorem 7.1.** *There are no negative fundamental discriminants $d$ with one class of binary quadratic forms in each genus satisfying $7392 < |d| \leq 10^{21}$.*

Subsequently we consider two conditions on the prime factors of $d$, and prove the following two theorems.

**Theorem 7.2.** *There are no negative fundamental discriminants $d$ such that $2, 3, 5 \nmid d$ with one class of binary quadratic forms in each genus satisfying $7392 < |d| \leq 10^{5500}$.*

**Theorem 7.3.** *Suppose that $d$ is a negative fundamental discriminant and $|d| > 7392$, and let $P$ be its largest prime factor. Then $d$ has at least two classes of binary quadratic forms in each genus if*

$$P \geq 5 \cdot 10^{15} |d|^{1/2} (\log |d|)^5 \log\log |d|.$$

We note that extending Theorem 7.2 to the weaker condition of $2, 3 \nmid d$ is certainly feasible, and we believe that with further work (and computational resources) this approach seems likely to be able to provide large general lower bounds, with no divisibility criteria, for the one class per genus problem.

## 7.1 Sieving small discriminants

Firstly, if $\left(\frac{d}{p}\right) = 1$ for an odd prime $p$, then $0 \not\equiv d \equiv a_1^2 \equiv a_2^2 \mod p$ for some $a_1 \equiv -a_2 \mod p$. If $d \equiv 1 \mod 4$, if $a_1$ is odd, then $d - a_1^2 \equiv 0 \mod 4$ and $d - a_1^2 \equiv 0 \mod p$, and if $a_1$ is even, then as $p$ is odd, then $a_2 = -a_1 + p$ is odd, so $d - a_2^2 \equiv 0 \mod 4$ and $d - a_2^2 \equiv 0 \mod p$. If $d \equiv 0 \mod 4$, then as one of

$a_1, a_2$ is even, $d - a_i^2 \equiv 0 \mod 4$ and $d - a_i^2 \equiv 0 \mod p$ for the even $a_i$. Thus in either case there exists $k$ such that $d = a^2 - 4kp$, so there exists a binary quadratic form $(p, a, k)$ of discriminant $d$, and if $p \leq \sqrt{|d|/3}$, then this form is reduced. If $d$ has only one class of binary quadratic forms in each genus, the only reduced binary quadratic forms of discriminant $d$ are of the form $(a, a, c)$, $(a, b, a)$, or $(a, 0, c)$. As $0 < a < p$, unless $p = k$, $d$ has at least two binary quadratic forms in each genus. If $p = k$, then $|d| \leq 4p^2$. So to eliminate possible fundamental discriminants $d$ with one class of binary quadratic forms in each genus, we apply our algorithm with the tests $\left(\frac{d}{p_i}\right) \neq 1$ for $2 \leq i \leq 169$. Then if $|d| \geq 4.1 \cdot 10^7 > 4p_{169}^2$ and $d$ fails a test, there will exist a non-ambiguous binary quadratic form of discriminant $d$. We apply our algorithm to combine linear congruences with the following parameters:

$$
\begin{array}{lll}
B = 64, & H_1 = 0, & H_2 = 10^{21}, \\
P_1 = 3 \cdot 11 \cdot 19 \cdot 31 \cdot 43 \cdot 59, & P_2 = 5 \cdot 13 \cdot 23 \cdot 37 \cdot 47, & P_3 = 7 \cdot 17 \cdot 29 \cdot 41 \cdot 53, \\
Q_1 = 61 \cdot 67, & Q_2 = 71 \cdot 73, & Q_3 = 79 \cdot 83, \\
Q_4 = 89 \cdot 97, & Q_5 = 101 \cdot 103, & R_{i-27} = p_i,
\end{array}
$$

where $p_i$ is the $i$'th prime, for $i = 28, \ldots, 169$. No such discriminants were found, this computation taking about six core-months. The range $7392 < |d| \leq 4.1 \cdot 10^7$ may be tested directly, and none were found. This yields Theorem 7.1. We note that compared to the pseudosquare problem, as we are testing $\left(\frac{d}{p}\right) \neq 1$, as opposed to $\left(\frac{d}{p}\right) = 1$ in the case of pseudosquares, there are roughly $8 \prod_{i=2}^{n} \frac{p_i+1}{p_i-1} \approx 100$ times more admissible values of $d$ modulo $p_1 \cdot \cdots \cdot p_n$.

## 7.2 Discriminants indivisible by 2, 3, or 5

We consider here odd fundamental discriminants $d$ with one class of binary quadratic forms in each genus, so $d \equiv 1(4)$, and the possible forms occurring are as follows:

**No Forms** $(a, a, a)$**:** $-3a^2 = d$, so as $d$ is fundamental and odd, $d = -3$.

**No forms** $(a, 0, c)$**:** $-4ac = d$, but $d$ is odd.

**Forms** $(a, a, c)$**:** $a^2 - 4ac = d$, so $a|d$. Suppose $a|d$, then as $a \cdot \frac{d}{a} \equiv 1(4)$, $a - d/a \equiv 0(4)$, so letting $c = (a - d/a)/4$ when $a \leq c$ gives a reduced form. In particular we have such a form if $a \leq \sqrt{|d|/3}$.

**Forms** $(a,b,a)$**:** $b^2-4a^2=d$. Suppose $x|d$, $\sqrt{|d|/3}<x<\sqrt{|d|}$. $d\equiv 1(4)$, so $d/x+x\equiv 0(4)$. So $a=(|d|/x+x)/4$, $b=(|d|/x-x)/2$ satisfy $b^2-4a^2=d$ and $(|d|/x-x)/2<(|d|/x+x)/4$ iff $|d|<3x^2$. However, we note that a form $(a,b,a)$ is equivalent to the form $(2a-b,2a-b,a)$ under the transformation $f(x+y,-x)$, and that this form satisfies $2a-b=\frac{|d|/x+x}{2}-\frac{|d|/x-x}{2}=x\leq\sqrt{|d|}$.

In particular, taking the forms $(a,a,c)$, where $a$ runs over the divisors of $d$ which are $\leq\sqrt{|d|}$, we obtain $2^{\omega(d)-1}$ (not necessarily reduced) forms from distinct classes of discriminant $d$, i.e. all of them.

### 7.2.1 The main inequality

Here we follow Section 3 of [86]. Let $\chi_k$ be a real primitive character modulo $|k|$, with $(d,k)=1$. Then we have

$$\left(\frac{|k||d|^{\frac{1}{2}}}{2\pi}\right)^{s-\frac{1}{2}}\Gamma(s)L(s,\chi_k)L(s,\chi_{dk})=T(s)+T(1-s)+U(s)$$

where, denoting by $Q_d$ a set of binary quadratic forms of discriminant $d$, with one element from each class, $Q$ its elements, and $a$ the coefficient of $x^2$ in $Q$,

$$\begin{aligned}
T(s)&=\Gamma(s)\zeta(2s)P_k(s)A(s)\left(\frac{|k||d|^{\frac{1}{2}}}{2\pi}\right)^{s-\frac{1}{2}},\\
U(s)&=4\pi^{\frac{1}{2}}|k|^{-1}\sum_{Q\in Q_d}a^{-\frac{1}{2}}\sum_{n=1}^{\infty}K_{s-\frac{1}{2}}\left(\frac{\pi n|d|^{\frac{1}{2}}}{a|k|}\right)n^{s-\frac{1}{2}}\\
&\qquad\cdot\sum_{y|n}y^{1-2s}\mathrm{Re}\left(\sum_{j=1}^{|k|}\chi_k(Q(j,y))e\left(\frac{jn}{|k|y}\right)e\left(\frac{bn}{2a|k|}\right)\right),\\
P_k(s)&=\prod_{p|k}(1-p^{-2s}),\\
A(s)&=\sum_{Q\in Q_d}\chi_k(a)a^{-s}.
\end{aligned}$$

So noting that $T\left(\frac{1}{2}+ix\right)=\overline{T\left(\frac{1}{2}-ix\right)}$, evaluating the expression at a zero $s=\frac{1}{2}+it$ of $L(s,\chi_k)$ gives

$$\frac{T\left(\frac{1}{2}+it\right)}{\overline{T\left(\frac{1}{2}+it\right)}}=-1-\frac{U\left(\frac{1}{2}+it\right)}{\overline{T\left(\frac{1}{2}+it\right)}},$$

yielding the inequality

$$2|\cos\theta|=\sqrt{(\sin 2\theta)^2+(-1-\cos 2\theta)^2}\leq\frac{\left|U\left(\frac{1}{2}+it\right)\right|}{\left|T\left(\frac{1}{2}+it\right)\right|},$$

where $\theta = \arg\left(T\left(\frac{1}{2}+it\right)\right)$. Letting

$$\varphi = \arg\left(i\zeta(1+2it)\Gamma(1/2+it)P_k(1/2+it)\left(\frac{|k|}{2\pi}\right)^{it}\right),$$
$$F = \left|\zeta(1+2it)\Gamma(1/2+it)P_k(1/2+it)\right|,$$

which are quantities we may compute, the above inequality gives

$$\left|\sin\left(\frac{t}{2}\log|d| + \arg(A(1/2+it)) + \varphi\right)\right| \leq \frac{|U(1/2+it)|}{2|A(1/2+it)|F}. \tag{7.1}$$

This inequality excludes ranges of $d$. We will establish upper bounds on $|U(s)|$ and $|\arg(A(1/2+it))|$, and a lower bound on $|A(1/2+it)|$ in order to apply this inequality.

We define the important auxiliary quantity

$$B(s) = \prod_{p|d}\left(1+\chi_k(p)p^{-s}\right),$$

which will be used as an approximation to $A(s)$.

**Lemma 7.1.** *The term*

$$A\left(\frac{1}{2}+it\right) = \sum_{Q\in Q_d}\chi_k(a)a^{-1/2-it}$$

*is bounded below by*

$$|B(1/2+it)| - \frac{2^{\omega(d)-1}}{|d|^{1/4}}.$$

*Proof.* Firstly, we have the two expressions

$$A(s) = \sum_{\substack{p_{i_1}\cdots p_{i_k}\leq\sqrt{|d|}\\ p_{i_j}|d \text{ distinct}\\ l\leq\omega(d)}}\chi_k(p_{i_1})\cdots\chi_k(p_{i_l})(p_{i_1}\cdots p_{i_l})^{-s},$$
$$B(s) = \sum_{\substack{p_{i_j}|d \text{ distinct}\\ l\leq\omega(d)}}\chi_k(p_{i_1})\cdots\chi_k(p_{i_l})(p_{i_1}\cdots p_{i_l})^{-s},$$

so that the difference between this and $B(s)$ at $s=1/2+it$ is bounded by

$$\sum_{p_{i_1}\cdots p_{i_l}\geq\sqrt{|d|}}(p_{i_1}\cdots p_{i_l})^{-1/2} \leq \frac{2^{\omega(d)-1}}{|d|^{1/4}}.$$

□

Now to bound $\arg(A(1/2+it))$, we have the following lemma.

**Lemma 7.2.** *Let*

$$R_1 = \frac{2^{\omega(d)-1}}{|d|^{1/4}},$$
$$R_2 = \frac{2^{\omega(d)-2}\log|d|}{|d|^{1/4}},$$
$$C(s) = -\sum_{p|d} \frac{\chi(p)(\log p)p^{-s}}{1+\chi(p)p^{-s}}.$$

*Then* $\arg(A(1/2+it))$ *is bounded in absolute value by*

$$t\left(\sup_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|C\left(\frac{1}{2}+i\tau\right)\right| + \frac{R_2}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right)$$
$$\cdot\left(1-\frac{R_1}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right)^{-1}.$$

*Proof.* We first bound the derivative of $A(s)$,

$$A'(s) = -\sum_{Q\in Q_d}(\log a)\chi(a)a^{-s}$$

in terms of

$$B'(s) = -\sum_{p_{i_1}\cdots p_{i_k}|d}(\log p_{i_1}\cdots p_{i_k})\chi(p_{i_1}\cdots p_{i_k})p_{i_1}\cdots p_{i_k}^{-s}.$$

Similarly to the previous lemma, their difference is bounded, at $s=\frac{1}{2}+it$, by

$$\frac{2^{\omega(d)-2}\log|d|}{d^{1/4}}.$$

Now to bound $A'/A$, we first have, by the above, $|A'|\le|B'|+R_2$, and by Lemma 7.1, $|A|\ge|B|-R_1$, so that when $|B|\ge R_1$,

$$\left|\frac{A'}{A}\right| \le \frac{|B'|+R_2}{|B|-R_1} = \frac{|B'|}{|B|}\cdot\frac{|B|}{|B|-R_1} + \frac{R_2}{|B|-R_1} = \left(\left|\frac{B'}{B}\right|+\frac{R_2}{|B|}\right)\left(1-\frac{R_1}{|B|}\right)^{-1}.$$

We now give our expression for $\frac{B'}{B}$,

$$-\frac{\sum_{q|d}(\log q)\chi_k(q)q^{-s}\prod_{p|d,p\neq q}(1+\chi_k(p)p^{-s})}{\prod_{p|d}(1+\chi_k(p)p^{-s})} = -\sum_{p|d}\frac{\chi_k(p)(\log p)p^{-s}}{1+\chi_k(p)p^{-s}} =: C(s).$$

As the argument of $A\left(\frac{1}{2}\right)$ is 0, the argument of $A\left(\frac{1}{2}+it\right)$ is bounded by

$$\int_0^t \left|\frac{A'(1/2+iy)}{A(1/2+iy)}\right| dy.$$

Over this interval, we may bound $\left|\frac{A'(1/2+iy)}{A(1/2+iy)}\right|$ by

$$\left(\sup_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|C\left(\frac{1}{2}+i\tau\right)\right| + \frac{R_2}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right) \cdot\left(1-\frac{R_1}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right)^{-1},$$

and the lemma follows. □

#### 7.2.1.1 Computation of bounds on $A$ and its argument

To give a lower bound for $B(1/2+it)$, we take first a list of the absolute values of $\left(1+\chi_k(p)p^{-1/2+it}\right)$, for the first $\omega(d)$ primes under consideration (i.e. starting from 7). For subsequent primes $p_i$ we compare the largest absolute value in the list with that of $\left(1+\chi_k(p_i)p_i^{-1/2+it}\right)$, and if $\left(1+\chi_k(p_i)p_i^{-1/2+it}\right)$ is smaller, it replaces the largest value in the list. This process is continued until the largest value in the list is smaller than $\left(1-p_{i+1}^{-1/2}\right)$, as this is a lower bound for all subsequent primes, at which point we have a lower bound for $B(1/2+it)$.

In order to give an upper bound for $\frac{|B'|}{|B|}$, we take a list of values of

$$\left|\frac{(\log p)p^{-1/2-i\tau^*}}{1+\chi(p)p^{-1/2-i\tau^*}}\right|,$$

where $\tau^*=0$ if $\chi_k(p)=-1$, and $\tau^*=it$ if $\chi_k(p)=1$ — as $t\log p\le\pi$ this maximizes the term over the interval $\tau^*\in[0,t]$. Similarly to before, we replace the smallest element of our list by

$$\left|\frac{(\log p_i)p_i^{-1/2-i\tau^*}}{1+\chi(p_i)p_i^{-1/2-i\tau^*}}\right|$$

if the expression in $p_i$ is larger, until all elements of the list are larger than

$$\left|\frac{(\log p_{i+1})p_{i+1}^{-1/2}}{1-p_{i+1}^{-1/2}}\right|,$$

as this is an upper bound for all subsequent primes.

The bounds for $A(1/2+it)$ and its argument then follow from Lemma 7.1 and Lemma 7.2.

### 7.2.2 A computational bound on $U(1/2+it)$ for $k=-17923$

Here we consider the expression

$$U_Q(s) := 4\sqrt{\pi}|k|^{-1}a^{-1/2}\sum_{n=1}^{\infty} K_{s-1/2}\left(\frac{\pi n|d|^{1/2}}{a|k|}\right) n^{s-1/2} \cdot \sum_{y|n} y^{1-2s}\mathrm{Re}\left\{\sum_{j=1}^{|k|}\chi(Q(j,y))e\left(\frac{jn}{|k|y}\right)e\left(\frac{bn}{2ak}\right)\right\},$$

where $k=-17923$, and

$$K_v(x) = \int_1^{\infty} e^{-\frac{x}{2}(u+u^{-1})}(u^{-v-1}+u^{v-1})du$$

is the modified Bessel function of the second kind.

All computations of this section were performed by programs written using functions from Arb [43], an arbitrary-precision ball arithmetic library.

We note that for $x>0$ and $\mathrm{Re}(v)=0$, it is easily shown by that

$$|K_v(x)| \le |K_0(x)| \le \sqrt{\frac{\pi}{2x}}e^{-x}.$$

We let $\alpha = \frac{\sqrt{|d|}}{a}$, and split the above sum into four parts,

$$U_Q(s) = U_{Q,L}(s) + U_{Q,M}(s) + U_{Q,H}(s) + U_{Q,S}(s),$$

where:

1. $U_{Q,L}(s)$ is the sum up to $n=2|k|$,
2. $U_{Q,M}(s)$ is the sum from $n=2|k|+1$ to $2\cdot 10^6$ excluding $n$ divisible by $k^2$,
3. $U_{Q,H}(s)$ is the sum from $n=2\cdot 10^6+1$ to infinity excluding $n$ divisible by $k^2$,
4. $U_{Q,S}(s)$ is the sum over $n$ divisible by $k^2$.

The value of $t$ is given by

$$0.0309857994985 + \theta\cdot 10^{-13},$$

where $|\theta|\le 1$.

We will now bound these four terms for our forms $(a,a,c)$, $a\le\sqrt{|d|}$.

**Lemma 7.3.** *With the above definitions, and $k = -17923$, it may be computed that*

$$|U_{Q,L}(1/2+it)| \leq \begin{cases} 230|d|^{-1/4} \text{ if } \alpha \geq 1 \\ 107|d|^{-1/4} \text{ if } \alpha \geq 100 \\ 32|d|^{-1/4} \text{ if } \alpha \geq 1000 \end{cases}$$

*Proof.* We first precompute the values of

$$\sum_{j=1}^{|k|} \chi(j^2+c')e\left(\frac{jm}{|k|}\right)$$

for all $m, c' \not\equiv 0(|k|)$. As our binary quadratic forms are of the form $(a,a,c)$, the expression

$$\sum_{j=1}^{|k|} \chi(Q(j,y))e\left(\frac{jn}{|k|y}\right)$$

may be put in this form by completion of the square and substitution assuming $a^{-1}c \not\equiv 4^{-1}(|k|)$ and $a \not\equiv 0(|k|)$ — as we have assumed $(d,k)=1$, and $d = a^2 - 4ac$, neither of these cases occur.

We then compute the sum up to $2|k|$ for each $c' \not\equiv 0(|k|)$, and take their maximum. This will be done for various values of $\alpha = \sqrt{d}/a$. As

$$\begin{aligned}\sum_{j=1}^{|k|} \chi(aj^2+ajy+cy^2)e\left(\frac{jn}{|k|y}\right) &= \sum_{j=1}^{|k|} \chi(a(j+2^{-1}y)^2+(c-2^{-2})y^2)e\left(\frac{jn}{|k|y}\right) \\ &= \chi(a)\sum_{i=1}^{|k|} \chi(i^2+c'y^2)e\left(\frac{in}{|k|y}\right)e\left(-\frac{2^{-1}yn}{|k|y}\right),\end{aligned}$$

we have, as $2^{-1} \equiv 8962(|k|)$,

$$\begin{aligned}U_{Q,L}\left(\frac{1}{2}+it\right) = \frac{4\sqrt{\pi}\chi(a)\alpha^{1/2}}{|k||d|^{-1/4}} \sum_{n=1}^{2k} K_{it}\left(\frac{\alpha\pi n}{|k|}\right) n^{it} \\ \cdot \sum_{y|n} y^{-2it} \mathrm{Re}\left\{ e\left(-\frac{8962n}{|k|}\right) e\left(\frac{n}{2|k|}\right) \sum_{j=1}^{|k|} \chi(j^2+c'y^2)e\left(\frac{jn}{|k|y}\right)\right\}.\end{aligned}$$

We note that the final character sum in this expression is real. We will bound this expression by computing the maximum value over $c'$ of

$$\begin{aligned}C(\alpha) := \sum_{n=1}^{2|k|} K_{it}\left(\frac{\alpha\pi n}{|k|}\right) n^{it} \sum_{y|n} y^{-2it} \\ \cdot \left(\mathrm{Re}\left\{ e\left(-\frac{8962n}{|k|}\right) e\left(\frac{n}{2|k|}\right)\right\} \sum_{j=1}^{k} \chi(j^2+c'y^2)e\left(\frac{jn}{|k|y}\right)\right)\end{aligned}$$

for a sufficient number of values of $\alpha$, by computing its absolute value at various gridpoints, and using a bound on its derivative between these gridpoints. A first bound on the absolute value of the derivative with respect to $\alpha$ of this expression is as follows:

$$D(\alpha) := 2\sqrt{|k|}\sum_{n=1}^{\infty} d(n)\frac{\pi n}{|k|}\left|K'_{it}\left(\frac{\alpha\pi n}{|k|}\right)\right|,$$

where $d(n)$ is the number of divisors function. One may deduce this by taking the absolute value of the derivative, and using the bound of $2\sqrt{|k|}$ of Lemma 6 of [88] for the expression

$$\left|\sum_{j=1}^{|k|}\chi(Q(j,y))e\left(\frac{jm}{|k|}\right)\right|,$$

when $(y, m, |k|) = 1$, which holds in all cases here as $n \leq 2|k|$.

Using the differential relation

$$2e^{\pi i\nu}K'_{\nu}(x) = e^{\pi i(\nu+1)}K_{\nu+1}(x) + e^{\pi i(\nu-1)}K_{\nu-1}(x),$$

we have the following expression for $D(\alpha)$,

$$D(\alpha) = \frac{\pi}{\sqrt{|k|}}\sum_{n=1}^{2|k|} nd(n)\left|K_{it+1}\left(\frac{\alpha\pi n}{|k|}\right) + K_{it-1}\left(\frac{\alpha\pi n}{|k|}\right)\right|.$$

Now using the integral expression of the modified Bessel function $K_{\nu}$,

$$K_{\nu}(x) = \int_1^{\infty} e^{\frac{x}{2}(u+u^{-1})}(u^{\nu-1} + u^{-\nu-1})du,$$

the sum $|K_{it+1}(x) + K_{it-1}(x)|$ is bounded as follows:

$$\begin{aligned}
&\left|\int_1^{\infty} e^{\frac{x}{2}(u+u^{-1})}(u^{it} + u^{-it-2})du + \int_1^{\infty} e^{\frac{x}{2}(u+u^{-1})}(u^{it-2} + u^{-it})du\right| \\
&\quad = 2\left|\int_1^{\infty} e^{\frac{x}{2}(u+u^{-1})}\cos(t\log u)(u+u^{-2})du\right| \\
&\quad \leq 2\int_1^{\infty} e^{\frac{x}{2}(u+u^{-1})}(u+u^{-2})du \\
&\quad = 2K_1(x).
\end{aligned}$$

We now have a bound for $|D(\alpha)|$ of

$$|D(\alpha)| \leq \frac{2\pi}{\sqrt{|k|}}\sum_{n=1}^{2|k|} nd(n)K_1\left(\frac{\alpha\pi n}{|k|}\right) =: E(\alpha),$$

Which serves as our bound on the derivative of $C(\alpha)$. We note that this is decreasing in $\alpha$, so that for $h > 0$,

$$|C(\alpha + h)| \leq |C(\alpha)| + hE(\alpha).$$

We aim for a bound of $580000\alpha^{-1/2}$ on $C(\alpha)$ for $\alpha \in [1, 100]$, of $280000\alpha^{-1/2}$ for $\alpha \in [100, 1000]$, and of $80000\alpha^{-1/2}$ for $\alpha \in [1000, \infty)$. So, with $\alpha_0 = 1$, upon calculation of $C(\alpha_i)$ and $E(\alpha_i)$, we take $\alpha_{i+1}$ to be greatest number such that

$$|C(\alpha_i)| + (\alpha_{i+1} - \alpha_i)E(\alpha_i) \leq c\alpha_{i+1}^{-1/2},$$

where $c$ is 580000, 280000, or 80000 depending on whether $\alpha_{i+1}$ is in the range $[1, 100]$, $[100, 1000]$, or $[1000, \infty)$, and repeat the calculations for $\alpha_{i+1}$. We perform these calculations up to $\alpha_{i+1} = 10^9$, and verify the bounds of $c\alpha^{-1/2}$ on $C(\alpha)$. For $\alpha \geq 10^9$, letting

$$g(x) = x^{-1}e^{-x}(1 + \log(1 + x^{-1})),$$

the second to last equation in the proof of Lemma 9 [62] gives the bound of

$$\begin{aligned}
|U_Q(1/2 + it)| &\leq 4\left(\frac{\pi}{|k|}\right)^{1/2}\left(\prod_{p|k}(2 + 3p^{-3/2})\right)a^{-1/2}g\left(\frac{\pi|d|^{1/2}}{2a|k|}\right) \\
&\leq 3000\frac{a^{1/2}}{|d|^{1/2}} \\
&\leq 3000\alpha^{-1/2}|d|^{-1/4} \\
&\leq 0.1|d|^{1/4},
\end{aligned}$$

which yields

$$\left|4\pi k^{-1}\frac{\alpha^{1/2}}{|d|^{1/4}}C(\alpha)\right| \leq 0.1|d|^{-1/4}$$

for $\alpha \geq 10^9$. Now multiplying each of 580000, 280000, and 80000 by $4\sqrt{\pi}|k|^{-1}$ yields the lemma. □

We now computationally bound the sum over $2|k| \leq n \leq 2 \cdot 10^6$.

**Lemma 7.4.**

$$|U_{Q,M}(1/2 + it)| \leq 7|d|^{-1/4}.$$

*Proof.* By Lemma 6 of [88], the expression

$$\left|\sum_{j=1}^{|k|}\chi(Q(j,y))e\left(\frac{jn}{|k|y}\right)\right|$$

is bounded in absolute value by $2\sqrt{|k|}$ if $(y, n/y, |k|) = 1$, and by $|k|$ if $(y, n/y, |k|) = |k|$, so as $k^2 \nmid n$, as we have excluded these terms from the sum in $U_{Q,M}$, we take the bound $2\sqrt{|k|}$, giving

$$\left|a^{-1/2}\sum_{\substack{n=2|k|\\ k^2\nmid n}}^{2\cdot 10^6} K_{it}\left(\frac{\pi n|d|^{1/2}}{a|k|}\right)n^{it}\sum_{y|n}y^{-2t}\mathrm{Re}\left\{\sum_{j=1}^{|k|}\chi(Q(j,y))e\left(\frac{jn}{|k|y}\right)e\left(\frac{n}{2|k|}\right)\right\}\right|$$

$$\leq 2\sqrt{|k|}a^{-1/2}\sum_{n=2|k|}^{2\cdot 10^6} d(n)K_0\left(\frac{\pi n|d|^{1/2}}{a|k|}\right). \tag{7.2}$$

We compute the following bound on the first term when $\alpha = \frac{|d|^{1/2}}{a} = 1$:

$$\sum_{n=2p}^{2\cdot 10^6} d(n)K_0\left(\frac{\pi n}{|k|}\right) < 58.$$

As may easily be shown, $\sqrt{\alpha}K_0(\alpha\tau)$ is decreasing in $\alpha$ for $\tau \geq \pi$ and $\alpha \geq 1$, so this serves as a bound regardless of the ratio $\alpha = \sqrt{|d|}/a$. So as $a^{-1/2} = \sqrt{\alpha}|d|^{-1/4}$, we obtain

$$2\sqrt{|k|}a^{-1/2}\sum_{n=2|k|}^{2\cdot 10^6} d(n)K_0\left(\frac{\pi n\alpha}{|k|}\right) < 15530|d|^{-1/4},$$

and multiplying by $4\sqrt{\pi}|k|^{-1}$ yields the lemma. □

We now bound the sum for terms $n \geq 2 \cdot 10^6$, $k^2 \nmid n$.

**Lemma 7.5.**

$$|U_{Q,H}(1/2 + it)| \leq 0.001|d|^{-1/4}$$

*Proof.* We note again that

$$\sqrt{\alpha}K_0(\alpha\tau)$$

is decreasing in $\alpha$ for $\tau \geq \pi$ and $\alpha \geq 1$, and $0 < K_0(x) \leq \sqrt{\frac{\pi}{2x}}e^{-x}$. Substituting this into equation 7.2 of the previous lemma, we have the bound

$$\begin{aligned} a^{-1/2}\cdot 2\sqrt{|k|}\sum_{n=2\cdot 10^6}^{\infty} d(n)K_0\left(\frac{n\pi\alpha}{k}\right) &\leq \sqrt{\pi}|d|^{-1/4}\cdot 8\sqrt{|k|}\sum_{n=2\cdot 10^6}^{\infty}\frac{\sqrt{|k|}d(n)}{\sqrt{2\pi n}}e^{-\pi n/k}\\ &\leq |d|^{-1/4}\cdot 6|k|\sum_{n=2\cdot 10^6}^{\infty}\sqrt{n}e^{-\pi n/k}\\ &\leq 0.001|d|^{-1/4}, \end{aligned}$$

and multiplying by $4\sqrt{\pi}|k|^{-1}$ yields the lemma. □

**Lemma 7.6.**

$$|U_{Q,S}(1/2+s)| \leq 0.001|d|^{-1/4}.$$

*Proof.* The term

$$\left|\sum_{j=1}^{|k|} \chi(Q(j,y))e\left(\frac{jn}{|k|y}\right)\right|$$

is bounded by $|k|$, as $(n, n/y, k) = |k|$, so similarly to the previous two lemmas,

$$\begin{aligned}|d|^{-1/4}|k|\sum_{n=1}^{\infty} d(|k|^2 n)K_0(|k|\pi n) &\leq |d|^{-1/4}|k|^{5/2}\sum_{n=1}^{\infty}\sqrt{n}e^{-|k|\pi n} \\ &\leq 0.001|d|^{-1/4},\end{aligned}$$

and again multiplying by $4\sqrt{\pi}|k|^{-1}$, we obtain the lemma. □

From these lemmas we conclude the following.

**Proposition 7.1.** *Let $d$ be an odd negative fundamental discriminant having one class of binary quadratic forms in each genus such that $(d, 17923) = 1$, let $Q$ be a binary quadratic form $(a, a, c)$ of discriminant $d$, and let $\alpha = \frac{\sqrt{|d|}}{a}$. Then*

$$\left|U_Q\left(\frac{1}{2}+it\right)\right| \leq \begin{cases} 238d^{-1/4} \text{ if } \alpha \geq 1, \\ 115d^{-1/4} \text{ if } \alpha \geq 100, \\ 40d^{-1/4} \text{ if } \alpha \geq 1000. \end{cases}$$

We note that the equivalent bound of Montgomery and Weinberger [88] is roughly $8994|d|^{-1/4}$, and of Watkins [86] is roughly $672|d|^{-1/4}$, for $\alpha \geq \sqrt{3}$.

### 7.2.3 Preliminary bounds on $|d|$

Here we apply the inequality 7.1 to $d$ such that $2, 3, 5 \nmid d$, separating into various numbers of prime factors of $d$, using the bounds of Lemma 7.1, Lemma 7.2, and Proposition 7.1. The bounds obtained are listed in Table 7.1. We will now exclude any other possibility of number of factors of $d$ for $|d| \leq 10^{60}$. Using Robin's [75] bound on the number of distinct prime factors of $d$,

$$\omega(d) \leq \frac{\log|d|}{\log\log|d|} + 1.45743\frac{\log|d|}{(\log\log|d|)^2},$$

and applying the inequality 7.1, again together with Lemma 7.1, Lemma 7.2, and Proposition 7.1, yields that there are no $d$ such that $10^{46.8} \leq |d| \leq 10^{60}$. As the

Table 7.1: $|d| \leq 10^x$ for various $\omega(d)$

| $\omega(d)$ | Bound | $\omega(d)$ | Bound |
|---|---|---|---|
| 8 | 16.5 | 19 | 31.5 |
| 9 | 18.0 | 20 | 32.8 |
| 10 | 19.4 | 21 | 34.1 |
| 11 | 20.8 | 22 | 35.5 |
| 12 | 22.1 | 23 | 36.8 |
| 13 | 23.6 | 24 | 38.1 |
| 14 | 24.8 | 25 | 39.6 |
| 15 | 26.1 | 26 | 41.0 |
| 16 | 27.4 | 27 | 42.5 |
| 17 | 28.8 | 28 | 44.0 |
| 18 | 30.1 | 29 | 45.5 |

product of the smallest 29 primes starting from 7 is $\geq 10^{49}$, we need only consider $d$ with fewer than 29 prime factors, and we already have the lower bound of Theorem 7.1, $|d| \geq 10^{21}$, so we need only test for discriminants with between 12 and 28 prime factors. The products of primes starting from 7 of lengths 23–28 are greater than their respective bounds in Table 7.1, so we do not need to test these. We will use an algorithm to eliminate the possibility of $d$ having a particular number of prime factors, after we improve these bounds in a subsequent section. In particular, we have so far shown the following exclusion.

**Proposition 7.2.** *If $d$ is a negative fundamental discriminant with one class of binary quadratic forms in each genus such that $2, 3, 5 \nmid d$ and $|d| \leq 10^{60}$, $d$ has at least 12 and at most 22 prime factors.*

### 7.2.4 An application of the Berry–Esseen inequality

Here we bound the number of positive divisors of $d$ close to $\sqrt{|d|}$ in order to make use of Proposition 7.1. We make use of an explicit version of a quantitative central limit theorem due independently to Berry [9] and Esseen [32]. Their result bounds the difference between the cumulative distribution functions of the normal distribution and of a sum of $n$ independent variates.

We consider $d$ in the following manner: its prime divisors correspond to our variates, and its divisors correspond to sums of these variates. As our $d$ has a relatively large number of prime factors, we will have large number of variates — enough to yield a useful bound on the distribution of divisors. These divisors, and hence the

associated binary quadratic forms, will be distributed somewhat similarly to the normal distribution. This then will allow us to make use of our previous bounds on $Q$ which depend on $\alpha$.

We first define the variates $\xi_p$, for each $p$ dividing $d$, by

$$\mathbb{P}\left(\xi_p = \frac{\log p}{2\sigma}\right) = 1/2,$$
$$\mathbb{P}\left(\xi_p = -\frac{\log p}{2\sigma}\right) = 1/2,$$

where

$$\sigma = \sqrt{\sum_{p|d} \frac{(\log p)^2}{4}}.$$

These have means 0, variances $\frac{(\log p)^2}{4\sigma^2}$, sum of variances 1, and absolute third moments $\frac{(\log p)^3}{8\sigma^3}$. We let $\mathcal{N}(x)$ be the cumulative distribution function of the normal distribution, and $F(x)$ the cumulative distribution function of $\sum_{p|d} \xi_p$.

We now apply our explicit Berry–Esseen theorem, due to Shevstova: by Theorem 1 of [79],

$$\sup_{0\le x\le 1} |F(x) - \mathcal{N}(x)| \le 0.56 \left(\sum_{p|d} \frac{(\log p)^3}{8\sigma^3}\right) =: R.$$

Now $R$ is maximized when $p$ are chosen so that $p_1, \dots p_{\omega(d)-1}$ are the first primes starting from 7, and $p_{\omega(d)}$ is the largest prime such that $p_1 \cdot \dots \cdot p_{\omega(d)}$ is bounded by our bound $B_n$ on $d$ with $n$ divisors. In particular, we simply take $\frac{B_n}{7\cdot\dots q_{n+2}}$ in our calculation, where $q_i$ is the $i$'th prime, as this is an upper bound on $p_{\omega(d)}$. Now a particular value taken by the sum of these variates corresponds to a divisor of $d$ as follows:

$$\sum_{p|d} \xi_p + \frac{\log|d|}{2\sigma} = \frac{\log\left(\prod_{p|d,\xi_p>0} p\right)}{\sigma},$$

and so a divisor bounded by $\sqrt{|d|}/\alpha$ corresponds to a sum of the $\xi_p$ bounded by $-\log\alpha/\sigma$. Our lower bound is then given by

$$\mathbb{P}\left(\sum_{p|d} \xi_p \le \frac{-\log\alpha}{\sigma}\right) \ge \mathcal{N}\left(\frac{-\log\alpha}{\sigma}\right) - R,$$

from which we deduce an upper bound for the number of divisors between $\sqrt{|d|}/\alpha$ and $\sqrt{|d|}$. Taking $\alpha = 10, 100, 1000$, we may apply the estimates for $|U_Q(s)|$ of

Table 7.2: $|d| \leq 10^x$ for various $\omega(d)$

| $\omega(d)$ | Bound | $\omega(d)$ | Bound |
|---|---|---|---|
| 12 | 22.0 | 18 | 29.7 |
| 13 | 23.4 | 19 | 31.0 |
| 14 | 24.7 | 20 | 32.1 |
| 15 | 26.0 | 21 | 33.4 |
| 16 | 27.2 | 22 | 34.7 |
| 17 | 28.5 | — | — |

Proposition 7.1, as these divisors correspond to binary quadratic forms $(a, a, c)$ with particular ratios $\frac{\sqrt{|d|}}{\alpha}$.

The remainder terms involved in the estimation of $A(1/2+it)$ and $\arg(A(1/2+it))$ are also improved. As an example, for Lemma 7.1, if there are $\geq 0.2 \cdot 2^{\omega(d)-1}$ divisors $\leq \frac{\sqrt{|d|}}{10}$, then there are also $\geq 0.2 \cdot 2^{\omega(d)-1}$ divisors $\geq 10\sqrt{|d|}$, which then contribute $\leq \frac{0.2 \cdot 2^{\omega(d)-1}}{\sqrt{10}|d|^{1/4}}$ to the difference between $A(1/2+it)$ and $B(1/2+it)$, giving the bound

$$|A(1/2+it) - B(1/2+it)| \leq \frac{0.8 \cdot 2^{\omega(d)-1}}{|d|^{1/4}} + \frac{0.2 \cdot 2^{\omega(d)-1}}{\sqrt{10}|d|^{1/4}} \leq 0.87\frac{2^{\omega(d)-1}}{|d|^{1/4}},$$

and there is a similar improvement for Lemma 7.2. The new bounds derived from equation 7.1 after our application of the Berry–Esseen inequality are listed in Table 7.2.

### 7.2.5 The algorithm

Our algorithm to deal with these cases is straightforward — we simply take a sufficiently large list of primes, and then, for a particular number of prime factors, say $n$, have $n$ loops, where each loop runs over primes starting from the prime succeeding that of the previous loop, and at each step of the inner loop, we test whether the product $-p_{i_1} \cdot \cdots \cdot p_{i_n}$ is a non-zero quadratic residue modulo test primes $q$. A particular iteration of the loop is exited when the product of primes taken is larger than our bound on $|d|$ for the particular number of factors. As an example, at the beginning of the algorithm, we test $-p_1 \cdot \cdots \cdot p_{n-2} \cdot p_{n-1} \cdot p_n$, then $-p_1 \cdot \cdots \cdot p_{n-2} \cdot p_{n-1} \cdot p_{n+1}$, and so on until this product exceeds our bound, at which point we have finished the inner loop. The penultimate loop then increments, and we repeat the inner loop starting from $-p_1 \cdot \cdots \cdot p_{n-2} \cdot p_n \cdot p_{n+1}$. Running this with the bounds of Table 7.2 eliminated all possible discriminants in roughly 1000 core-hours.

Together with Proposition 7.2, this yields the following exclusion.

**Proposition 7.3.** *There are no negative fundamental discriminants $d$, $2, 3, 5 \nmid d$ with one class of binary quadratic forms satisfying $7392 < |d| \leq 10^{60}$.*

*Remark.* This algorithm is of course quite basic, and some improvements can be made. Firstly, as the algorithm is running, and it iterates through primes, we *know* some of the primes which divide $d$, and this allows us to dynamically calculate the bounds on $|A|$ and $\arg(A)$ in order to obtain better upper bounds via equation 7.1 to which we must test. We tested this in the case of 12 prime factors, recalculating upper bounds at each iteration of the loop corresponding to the 8th prime divisor, and this provided a roughly 6 times speedup.

A further improvement is to the handling of the last loop. We precompute the Jacobi symbols of all the primes under consideration with respect to 30 or so primes, and before the final loop begins, compute $\left(\frac{q_1 \cdots q_{n-1}}{p}\right)$ for these 30 test primes $p$, and using a bit-packing method similar to our algorithm to combine linear congruences can simultaneously test many $q_n$.

In testing, these improvements together speed up our algorithm by roughly 150 times, and we were able to eliminate the possibility of $d$ having 12 prime factors and being indivisible by 2 or 3, so we believe that with not too much further work the above result may be extended fully to $d$ indivisible by 2 or 3.

We have also since developed an improved manner of parallelization for this algorithm.

Indeed we expect that these speedups will be greater as the divisibility criteria on $d$ are relaxed, and hope that these improvements, perhaps together with dynamically applying the improvements obtained from the Berry–Esseen theorem, and more computational resources, should allow us to completely eliminate the possibility of small discriminants with no divisibility conditions. We very roughly estimate that with the above described algorithmic improvements the necessary computations for the general case (with no divisibility criteria) would require 100–1000 core-years.

### 7.2.6 An application of Watkins' inequality

Here, in order to obtain a larger range of exclusion, we apply Watkins' [86] general bounds for $U(s)$. However, as this applies only to expansions of $L(s, \chi_d)$ in terms of *reduced* binary quadratic forms, we must consider forms $(a, b, a)$ in addition to forms $(a, a, c)$. We derive the new appropriate bounds on $A(s)$ and $\arg A(s)$ for $A(s) = \sum_{Q \in Q_d} \chi_k(a) a^{-s}$.

**Lemma 7.7.** *The term*

$$A(1/2+it) = \sum \chi_k(a) a^{-1/2+it}$$

*is bounded below in absolute value by*

$$|B(1/2+it)| - \frac{2^{\omega(d)+1}}{|d|^{1/4}}.$$

*Proof.* Letting $G(s)$ be the sum $\sum \chi_k(a) a^{-s}$, where $a$ ranges over the minima of reduced binary quadratic forms of the form $(a,b,a)$, and recalling that these correspond to divisors of $d$ which lie between $\sqrt{|d|/3}$ and $\sqrt{|d|}$, we have the expressions

$$A(s) = \sum_{\substack{p_{i_1}\cdots p_{i_l} \leq \sqrt{|d|/3} \\ p_{i_j} | d \text{ distinct} \\ l \leq \omega(d)}} \chi_k(p_{i_1}) \cdots \chi_k(p_{i_l}) (p_{i_1} \cdots p_{i_l})^{-s} + G(1/2+it)$$

and

$$B(s) = \sum_{\substack{p_{i_j} | d \text{ distinct} \\ l \leq \omega(d)}} \chi_k(p_{i_1}) \cdots \chi_k(p_{i_l}) (p_{i_1} \cdots p_{i_l})^{-s}.$$

Their difference is bounded in absolute value by

$$\sum_{\substack{p_{i_1}\cdots p_{i_l} \geq \sqrt{|d|} \\ p_{i_j} | d \text{ distinct} \\ l \leq \omega(d)}} (p_{i_1} \cdots p_{i_l})^{-1/2} + \sum_{\substack{\sqrt{|d|} \geq p_{i_1}\cdots p_{i_l} \geq \sqrt{|d|/3} \\ p_{i_j} | d \text{ distinct} \\ l \leq \omega(d)}} (p_{i_1} \cdots p_{i_l})^{-1/2} + |G(1/2+it)|.$$

For reduced binary quadratic forms of the form $(a,b,a)$, as $d = b^2 - 4a^2$, $a$ is bounded below by $\sqrt{|d|}/2$. This gives an upper bound on the terms in $G(1/2+it)$ of $\frac{2^{1/2}}{|d|^{1/4}}$, so the difference is bounded by

$$\frac{2^{\omega(d)-1} + 3^{1/4} 2^{\omega(d)-1} + 2^{1/2} 2^{\omega(d)-1}}{|d|^{1/4}} \leq \frac{2^{\omega(d)+1}}{|d|^{1/4}}.$$

□

**Lemma 7.8.** *Let*

$$R_1 = \frac{2^{\omega(d)+1}}{|d|^{1/4}},$$

$$R_2 = \frac{2^{\omega(d)} \log|d|}{|d|^{1/4}},$$

$$C(s) = -\sum_{p|d} \frac{\chi(p)(\log p) p^{-s}}{1 + \chi(p) p^{-s}}.$$

*Then* $\arg(A(1/2+it))$ *is bounded in absolute value by*

$$t\left(\sup_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|C\left(\frac{1}{2}+i\tau\right)\right|+\frac{R_2}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right)\cdot\left(1-\frac{R_1}{\inf_{\tau\in\left[\frac{1}{2},\frac{1}{2}+it\right]}\left|B\left(\frac{1}{2}+i\tau\right)\right|}\right)^{-1}.$$

*Proof.* We first bound the derivative of $A(s)$,

$$A'(s)=-\sum_a(\log a)\chi(a)a^{-s}$$

in terms of

$$B'(s)=-\sum_{p_{i_1}\cdots p_{i_k}|d}(\log p_{i_1}\cdots p_{i_k})\chi(p_{i_1}\cdots p_{i_k})p_{i_1}\cdots p_{i_k}^{-s}.$$

Similarly to the previous lemma, their difference is bounded, at $s=\frac{1}{2}+it$, by

$$\frac{2^{\omega(d)}\log|d|}{|d|^{1/4}}.$$

Now to bound $A'/A$ and thence $|\arg(A(1/2+it))|$, the remainder of the lemma proceeds as in Lemma 7.2. □

We now make use of Lemma 10 and Lemma 12 of [86], which allow for $k$ not coprime to $d$. With $(|k|,|d|)=g$, this inequality gives

$$2|\sin\theta|\le\left(W+2^{\omega(d)-1}\frac{2\sqrt{2\pi|k|}}{\sqrt{g}|d|^{1/4}}e^{-\pi g/|k|}\right)|T(1/2+it)|^{-1},\tag{7.3}$$

where

$$\theta=\arg\left(g^{1/2-s}T(1/2+it)\right),$$
$$T(s)=\Gamma(s)\zeta(2s)P_k(s)A(s)\left(\frac{|k||d|^{\frac{1}{2}}}{2\pi}\right)^{s-\frac{1}{2}},$$

and, letting $A_{UB}$ be an upper bound for $A(s)$ on the line $\mathrm{Re}(s)=1/4$,

$$W=11\left(\frac{2\pi g}{|k|\sqrt{|d|}}\right)^{1/4}A_{UB}\prod_{p|k}\left(1+p^{-1/2}\right).$$

We now bound $A_{UB}$.

**Lemma 7.9.** *For all $y$ real,*

$$|A(1/4+iy)| \leq \prod_{i\leq\omega(d)} (1+p_i^{-1/4}) + \frac{2^{\omega(d)+1}}{|d|^{1/8}},$$

*where $p_i$ is the $i$th prime.*

*Proof.* By the same argument as Lemma 7.7, $|A(1/4+iy)| \leq |B(1/4+iy)| + \frac{2^{\omega(d)+1}}{|d|^{1/8}}$, and

$$|B(1/4+iy)| \leq \prod_{p|d} \left(1+p^{-1/4}\right) \leq \prod_{i\leq\omega(d)} \left(1+p_i^{-1/4}\right).$$

□

Lastly, we make use of the following bound on $\omega(d)$ of Robin [75]:

$$\omega(d) \leq \frac{\log|d|}{\log\log|d|} + 1.45743\frac{\log|d|}{(\log\log|d|)^2},$$

from which we may bound $|B(1/2+it)|$, $\left|\frac{B'(1/2+it)}{B(1/2+it)}\right|$, and $\prod_{i\leq\omega(d)} \left(1+p^{-1/4}\right)$.

We now apply these bounds in combination with Watkins' inequality 7.3, for $k=-115147$, with each of $(|k|,|d|) = g = 1, 113, 1019, 115147$, which eliminates the range $10^{54}$ to $10^{770}$ for $2,3,5 \nmid d$, and similarly with $k = -175990483 = -19\cdot1427\cdot6491$ for each possibility of $(|k|,|d|)$, which eliminates the range $10^{70}$ to $10^{5500}$, and obtain the following exclusion.

**Proposition 7.4.** *There are no negative fundamental discriminants $d$, $2,3,5 \nmid d$ with one class of binary quadratic forms in each genus satisfying $10^{54} \leq |d| \leq 10^{5500}$.*

This combined with Proposition 7.3 yields Theorem 7.2.

### 7.2.7 Small gaps between zeros of Riemann's zeta function

We note that from the (expected) existence of particularly small gaps between zeros of Riemann's zeta function, lower bounds for class numbers may be deduced by a similar method to the above by evaluating the relation

$$\zeta(s)L(s,\chi) = \sum_Q \zeta_Q(s)$$

at $1/2+i\gamma_1$ and $1/2+i\gamma_2$. A possibility for further work would be to make use of a large set of zeros, such as those of LMFDB [56], or to compute new zeros by for example the methods of Bober–Hiary [15], and obtain ranges of exclusion from explicit zero-gaps.

For example, a small gap of size $\frac{2\pi}{1000 \log T}$, with $\gamma_1, \gamma_2$ around $T = 10^{30}$ (indicated by Bober–Hiari as a feasible area to compute $\zeta(z)$) would exclude discriminants roughly from $10^{150}$ to $10^{10^5}$. Many such small gaps, the sizes of which do not lie in a close harmonic ratio, can then be combined to give larger ranges of exclusion, likely up to $10^{10^6}$ and beyond.

We further note that, if we supposed that there existed $\theta$, $\epsilon$, and $\delta_\epsilon$ such that for any $T$, there exist zeros of Riemann's zeta function $1/2 + i\gamma_1$, $1/2 + i\gamma_2$, $\gamma_1, \gamma_2 \in [T, T^{1+\delta}]$, such that

$$\frac{\theta - \epsilon}{\log T} \leq |\gamma_2 - \gamma_1| \leq \frac{\theta + \epsilon}{\log T},$$

then if $\epsilon$ and $\delta$ are sufficiently (absolutely) small, one could deduce both an "analytic" proof of the class number one problem, and also an effective upper bound for negative discriminants with one class of binary quadratic forms in each genus. In particular, it would be interesting to attempt to prove, assuming the Riemann Hypothesis, that there is always a gap sufficiently close to the average gap of $\frac{2\pi}{\log T}$, or perhaps other combinations of zeros. This seems as though it would be easier to obtain than proving the existence of small gaps, which even assuming the Riemann Hypothesis seems currently out of reach.

## 7.3 Discriminants with a large prime factor

We consider here a special case of the one class per genus problem, where the discriminants under consideration have a large prime factor. The main argument proceeds in a similar manner to Baker's solutions of the class number one and two problems (see for example [6]). We consider an expansion in terms of the Epstein zeta functions associated to binary quadratic forms of products of $L$-functions, and reduce this to a linear form in logarithms. In our case, the height of the coefficients in the linear form in logarithms remains small as the minima of the binary quadratic forms are all divisors of the discriminant. The large prime factor ensures that the minima are small in comparison to the discriminant, which ensures that the "remainder" terms of the Epstein zeta functions are sufficiently small. Watkins [87] indicates that the spectral methods employed for his proof of the class number one problem would yield the impossibility of a prime factor $P \gg |d|^{1-\alpha}$, for some positive $\alpha$, though these methods too are restricted to the situation of small minima.

### 7.3.1 The main equality

This follows from [6]. Let $k > 0$, and $d < 0$ be the discriminants of the fields $\mathbb{Q}(\sqrt{k})$, $\mathbb{Q}(\sqrt{d})$ respectively, and $\chi$, $\chi'$ be the Kronecker symbols $\left(\frac{k}{\cdot}\right)$, $\left(\frac{d}{\cdot}\right)$ respectively. Then assuming $(k, d) = 1$, we have the following,

$$L(1,\chi)L(1,\chi\chi') = \frac{1}{2}\sum_{Q}\left(\frac{\pi^2}{3}\frac{\chi(a)}{a}\prod_{p|k}\left(1-\frac{1}{p^2}\right)+\sum_{r=-\infty}^{\infty}A_{r,Q}e^{\pi irb/(ka)}\right),$$

where the sum over $Q$ runs over the set of inequivalent reduced binary quadratic forms of discriminant $d$, and where $a$ and $b$ are the first and second coefficients of $Q$. The following hold,

$$|A_{r,Q}| \leq \frac{4\pi}{\sqrt{|d|}}|r|e^{-\pi|r|\sqrt{|d|}/(ka)}$$

for $r \neq 0$, and

$$A_{0,Q} = \begin{cases} -\frac{4\pi}{k\sqrt{|d|}}\chi(a)\log p & k \text{ a prime power,} \\ 0 & \text{otherwise.} \end{cases}$$

By the Dirichlet class number formula,

$$L(1,\chi) = \frac{2h(k)\log\epsilon}{\sqrt{k}}, \qquad L(1,\chi\chi') = \frac{h(kd)\pi}{\sqrt{k|d|}},$$

where $\epsilon$ is the fundamental unit of $\mathbb{Q}(\sqrt{k})$, and $h(k)$, $h(kd)$ are the class numbers of the fields $\mathbb{Q}(\sqrt{k})$ and $\mathbb{Q}(\sqrt{kd})$ respectively.

### 7.3.2 Estimates for some quantities

We computed in Section 7.1 that there are no further discriminants with one class per genus such that $|d| \leq 10^{21}$. We use this fact implicitly to simplify the expressions in subsequent bounds, for example replacing a bound of $\log\log|d|$ by $0.09\log|d|$. Regarding the forms, for a reduced form $(a, b, a)$, $|d| = 4a^2 - b^2$, with $0 \leq b \leq a$. Consequently $a \geq \sqrt{|d|/4}$, and $(2a-b)|d$, so that $d$ has a factor between $\sqrt{|d|/4}$ and $\sqrt{|d|}$. However $P > 2|d|^{1/2}$, so all factors of $|d|$ are either $< \sqrt{|d|}/2$ or $> 2\sqrt{|d|}$. So all forms are of the form $(a, 0, c)$ or $(a, a, c)$, and so all occurring minima divide $d$.

In order to bound $\omega(d)$, we apply Robin's bound (Theorem 12, [75]) on the number of prime factors of a number:

$$\omega(n) \leq \frac{\log n}{\log\log n} + 1.45743\frac{\log n}{(\log\log n)^2}.$$

As $|d|/P \leq \sqrt{|d|}$, we have, for $|d| \geq 10^{21}$,

$$\omega(d) \leq \frac{\log\sqrt{|d|}}{\log\log\sqrt{|d|}} + 1.45743\frac{\log\sqrt{|d|}}{(\log\log\sqrt{|d|})^2} + 1 \leq \frac{\log|d|}{\log\log|d|}.$$

We will make use of this bound throughout the subsequent analysis.

#### 7.3.2.1 The auxiliary factors

Here we define the parameter $k$ which will be of importance to our analysis.

**Lemma 7.10.** *Let $k = q_1q_2$, where $q_1, q_2$ are the first two primes which do not divide $d$, and satisfy $q_1q_2 \equiv 1(4)$. Then*

1. $k \leq 1.62(\log|d|)^2$,
2. *The class number of $\mathbb{Q}(\sqrt{k})$ is bounded by $0.64\log|d|$,*
3. *The regulator of $\mathbb{Q}(\sqrt{k})$ is bounded by $1.69\log|d|\log\log|d|$,*
4. *The class number of $\mathbb{Q}(\sqrt{kd})$ is bounded by $1.01\sqrt{|d|}\log|d|$,*
5. *The height of $Q := \prod_{p|k}\left(1 - \frac{1}{p^2}\right)$ is bounded by $2.16(\log|d|)^4$.*

*Proof. 1.* Firstly, we have $q_1, q_2 \leq p_{\omega(d)+4}$, as taking the first three odd primes not dividing $d$, two of them will be equal modulo 4, so that their product is congruent to 1 modulo 4, and hence $k = q_1q_2$ is the discriminant of the real quadratic field $\mathbb{Q}(\sqrt{k})$. By results of Rosser and Schonfeld (see Corollary to Theorem 3, [77]), for $n \geq 6$,

$$p_n \leq n(\log n + \log\log n),$$

where $p_n$ is the $n$'th prime. So the $q_i$ are bounded by

$$q_i \leq p_{\omega(d)+4} \leq (\omega(d)+4)(\log(\omega(d)+4) + \log\log(\omega(d)+4)) \leq 1.27\log|d|,$$

and so $k$ is bounded by $k \leq 1.62(\log|d|)^2$.

*2.* By a result of Le (Theorem (a), [54]), as $k$ is square-free and $\equiv 1(4)$, the class number of the real quadratic field $\mathbb{Q}(\sqrt{k})$ is bounded by $\sqrt{k}/2 \leq 0.64\log|d|$.

*3.* Let $\epsilon$ be the fundamental unit of $\mathbb{Q}(\sqrt{k})$. By a result of Hua, [42], p.329,

$$\log\epsilon \leq \sqrt{k}\left(\frac{1}{2}\log(k) + 1\right),$$

which yields a bound of

$$\begin{aligned}\sqrt{k}\left(\frac{1}{2}\log(k)+1\right) &\leq 1.27\log|d|(\log(1.27\log|d|)+1)\\ &\leq 1.69\log|d|\log\log|d|.\end{aligned}$$

*4.* By a result of Paulin (Proposition 2.2, [63]), If $D<0$ is a integer congruent to 0 or 1 modulo 4, then

$$h(D)<\frac{1}{\pi}\sqrt{|D|}(2+\log|D|).$$

As $k\equiv 1(4)$ and $d\equiv 0$ or $1(4)$, $kd\equiv 0$ or $1(4)$ and $kd<0$, so

$$\begin{aligned}h(kd) &\leq 2\sqrt{k|d|}(2+\log 2k|d|)/\pi\\ &\leq 2.54\log|d|\sqrt{|d|}(2+\log(3.24(\log|d|)^2|d|))/\pi\\ &\leq 1.01\sqrt{|d|}(\log|d|)^2.\end{aligned}$$

*5.* For the term $Q=\prod_{p|k}\left(1-\frac{1}{p^2}\right)$, expanding this into a fraction gives

$$\prod_{p|k}\left(1-\frac{1}{p^2}\right)=\frac{(q_1^2-1)(q_2^2-1)}{q_1^2q_2^2},$$

and we have $H(Q)\leq 2.61(\log|d|)^4$ by our previous bounds on $q_i$. □

**Lemma 7.11.** *The quantity*

$$\sum_Q\frac{\chi(a)}{a}=:\frac{C}{d}$$

*is bounded in height by* $1.84|d|\log\log|d|$.

*Proof.* There are no forms $(a,b,a)$, so the sum $\chi(a)/a$ runs over the smaller half of the divisors of $d$, so

$$\sum_Q\frac{\chi(a)}{a}=\frac{C}{d},$$

where $C$ is an integer. By a result of Robin [76], for $n\geq 3$,

$$\frac{\sigma(n)}{n}\leq e^\gamma\log\log n+\frac{0.649}{\log\log n},$$

so that $|C|\leq\sigma(d)\leq 1.84|d|\log\log|d|$.

□

#### 7.3.2.2 Bound on the remainder

The minima $a$ are bounded by $|d|/P$, and combined with the previous bound on $k$, this gives

$$\begin{aligned}|A_{r,Q}| &\leq \frac{4\pi}{\sqrt{|d|}}|r|e^{-\pi|r|\sqrt{|d|}/(ka)} \\ &\leq \frac{4\pi}{\sqrt{|d|}}|r|e^{-\pi|r|P/(\sqrt{|d|}1.24(\log|d|)^2)}.\end{aligned}$$

So we have

$$\begin{aligned}\left|\sum_{Q}\sum_{\substack{r=-\infty \\ r\neq 0}}^{\infty} A_{r,Q}e^{\pi irb/(ka)}\right| &\leq \sum_{\substack{r=-\infty \\ r\neq 0}}^{\infty}\sum_{Q}|A_{r,Q}| \\ &\leq \frac{4\pi h(d)}{\sqrt{|d|}}\sum_{\substack{r=-\infty \\ r\neq 0}}^{\infty}|r|e^{-\pi|r|P/(\sqrt{|d|}1.69(\log|d|)^2)} \\ &= \frac{8\pi h(d)}{\sqrt{|d|}}e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)} \\ &\qquad \cdot (1-e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)})^{-2} \\ &\leq \frac{27.5h(d)}{\sqrt{|d|}}e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)},\end{aligned}$$

where the last line follows as $P \geq 1.69\sqrt{|d|}(\log|d|)^2$, which yields the bound

$$(1-e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)}) \geq (1-e^{-\pi}).$$

### 7.3.3 The linear form in logarithms

Putting this together, we have

$$\frac{2h(k)\log\epsilon}{\sqrt{k}}\cdot\frac{h(kd)\pi}{\sqrt{k|d|}} = \frac{\pi^2 CQ}{3d} + \sum_{Q}\sum_{\substack{r=-\infty \\ r\neq 0}}^{\infty} A_{r,Q}e^{\pi irb/(ka)},$$

which yields a linear form in logarithms:

$$\begin{aligned}\left|\frac{\sqrt{|d|}6h(k)h(kd)}{kCQ}\log\epsilon+\pi\right| &= \left|\frac{\sqrt{|d|}6h(k)h(kd)}{kCQ}\log\epsilon - 2i\log i\right| \\ &= \left|\frac{3\sqrt{d}h(k)h(kd)}{kCQ}\log\epsilon+\log i\right| \\ &\leq \frac{27.5\cdot 3\sqrt{|d|}h(d)}{\pi|C|Q}e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)} \\ &\leq 52.6|d|e^{-\pi P/(\sqrt{|d|}1.69(\log|d|)^2)},\end{aligned}$$

where the last line follows from $h(d) < \sqrt{|d|}$, $|C| \geq 1$, $|Q| \geq 1/2$. We now recollect the bounds on the heights of the appearing terms, letting $Q_1$ and $Q_2$ be the denominator and numerator of $Q$ respectively:

$$\begin{aligned} H(k) &\leq 1.69(\log|d|)^2, \\ H(h(k)) &\leq 0.64\log|d|, \\ H(\sqrt{d}) &= \sqrt{|d|}, \\ H(h(kd)) &\leq 1.01\sqrt{|d|}(\log|d|)^2, \\ H(C) &\leq 1.84|d|\log\log|d|, \\ H(Q_1), H(Q_2) &\leq 2.61(\log|d|)^4, \\ H(\epsilon) &\leq \exp(1.69\log|d|\log\log|d|) =: A_1, \\ H(i) &= 1 =: A_2. \end{aligned}$$

The height of the coefficient $\beta = \frac{\sqrt{d}3h(k)h(kd)Q_1}{kCQ_2}$ is bounded by

$$\begin{aligned} H(\beta) &\leq H(\sqrt{d})H\left(\frac{3h(k)h(kd)Q_1}{kCQ_2}\right) \\ &\leq \sqrt{|d|}1.69(\log|d|)^2 1.84|d|\log\log|d|2.61(\log|d|)^4 \\ &\leq 8.12|d|^{3/2}(\log|d|)^6\log\log|d| =: B. \end{aligned}$$

We will apply the following lower bound of Waldschmidt and Mignotte to this linear form in two logarithms.

**Theorem 7.4** (Theorem, [61]). *Let $\beta, \alpha_1, \alpha_2$ be three non-zero algebraic numbers of exact degrees $D_0, D_1, D_2$. Let $D$ be the degree of the field $\mathbb{Q}(\beta, \alpha_1, \alpha_2)$ over $\mathbb{Q}$. For $j = 1, 2$, let $\log\alpha_j$ be any determination of the logarithm of $\alpha_j$, and let $A_j$ be an upper bound for the height of $\alpha_j$ and $\exp|\log\alpha_j|$; further define*

$$S_0 = D_0 + \log B, \qquad\qquad S_j = D_j + \log A_j,$$

*and assume*

$$\Lambda = \beta\log\alpha_1 - \log\alpha_2 \neq 0.$$

*Let*

$$T = 4 + \frac{S_0}{D_0} + \log\left(D^2 \cdot \frac{S_1}{D_1} \cdot \frac{S_2}{D_2}\right),$$

*then*

$$|\Lambda| > \exp\left(-5 \cdot 10^8 D^4 \cdot \frac{S_1}{D_1} \cdot \frac{S_2}{D_2} \cdot T^2\right).$$

*Proof of Theorem 7.3.* By the Gelfond–Schneider theorem, $\log \epsilon$ and $\log i$ are linearly independent over $\overline{\mathbb{Q}}$, so $\Lambda \neq 0$. Applying the theorem to our linear form in logarithms, we have

$$D_0 = D_1 = D_2 = 2, \quad D = [\mathbb{Q}(\beta, \epsilon, i) : \mathbb{Q}] = 8, \quad S_0 = 2 + B,$$
$$S_1 = 2 + \log A_1, \quad S_2 = 2, \quad T = 5 + \frac{\log B}{2} + \log(32 S_1).$$

Now Waldschmidt's result concludes

$$|\Lambda| \geq \exp\left(-5 \cdot 10^8 \cdot 8^4 (1 + \log A_1)\left(5 + \frac{\log B}{2} + \log(64 + 32 \log A_1)\right)^2\right).$$

Simplifying and using our lower bound $|d| \geq 10^{21}$, we obtain the lower bound

$$|\Lambda| \geq \exp\left(-1.2 \cdot 10^{16} (\log|d|)^3 \log\log|d|\right).$$

We have the upper bound

$$\exp\left(-\pi C (\log|d|)^3 / 1.24 + \log|d| + \log 50.4\right),$$

and combining it with our lower bound gives

$$C(\log|d|)^3 \log\log|d| \leq 4.8 \cdot 10^{15} (\log|d|)^3 \log\log|d| + 1.24 \log 51.6 + 1.24 \log|d|,$$

and taking $C = 5 \cdot 10^{15}$ violates this inequality, yielding the theorem. □

# References


[1] NIST digital library of mathematical functions. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[2] J. D. Achter. Detecting complex multiplication. In *Computational aspects of algebraic curves*, volume 13 of *Lecture Notes Ser. Comput.*, pages 38–50. World Sci. Publ., Hackensack, NJ, 2005.

[3] E. Alwaise. An algorithm for numerically inverting the modular $j$-function. *Res. Number Theory*, 4(2):Paper No. 20, 6, 2018.

[4] Y. André. Finitude des couples d'invariants modulaires singuliers sur une courbe algébrique plane non modulaire. *J. Reine Angew. Math.*, 505:203–208, 1998.

[5] A. Baker. Linear forms in the logarithms of algebraic numbers. I, II, III. *Mathematika*, 13:204–216; ibid. 14 (1967), 102–107; ibid. 14 (1967), 220–228, 1966.

[6] A. Baker. On the class number of imaginary quadratic fields. *Bull. Amer. Math. Soc.*, 77:678–684, 1971.

[7] D. Berkane, O. Bordellès, and O. Ramaré. Explicit upper bounds for the remainder term in the divisor problem. *Math. Comp.*, 81(278):1025–1051, 2012.

[8] D. J. Bernstein. Doubly focused enumeration of locally square polynomial values. In *High primes and misdemeanours: lectures in honour of the 60th birthday of Hugh Cowie Williams*, volume 41 of *Fields Inst. Commun.*, pages 69–76. Amer. Math. Soc., Providence, RI, 2004.

[9] A. C. Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.

[10] E. Besson. Points rationnels de la fonction gamma d'Euler. *Arch. Math. (Basel)*, 103(1):61–73, 2014.

[11] F. Beukers and C. J. Smyth. Cyclotomic points on curves. In *Number theory for the millennium, I (Urbana, IL, 2000)*, pages 67–85. A K Peters, Natick, MA, 2002.

[12] R. Bianconi. Some model theory of hypergeometric and Pfaffian functions. *South Amer. J. Log.*, 2(2):297–318, 2016.

[13] Y. Bilu, D. Masser, and U. Zannier. An effective "theorem of André" for $CM$-points on a plane curve. *Math. Proc. Cambridge Philos. Soc.*, 154(1):145–152, 2013.

[14] G. Binyamini. Density of algebraic points on Noetherian varieties. *Geom. Funct. Anal.*, 29(1):72–118, 2019.

[15] J. W. Bober and G. A. Hiary. New computations of the riemann zeta function on the critical line. *Experimental Mathematics*, 27(2):125–137, 2018.

[16] E. Bombieri and W. Gubler. *Heights in Diophantine geometry*, volume 4 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2006.

[17] J. M. Borwein and P. B. Borwein. *Pi and the AGM*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons, Inc., New York, 1987. A study in analytic number theory and computational complexity, A Wiley-Interscience Publication.

[18] G. J. Boxall and G. O. Jones. Algebraic values of certain analytic functions. *Int. Math. Res. Not. IMRN*, (4):1141–1158, 2015.

[19] N. Brisebarre and G. Philibert. Effective lower and upper bounds for the Fourier coefficients of powers of the modular invariant $j$. *J. Ramanujan Math. Soc.*, 20(4):255–282, 2005.

[20] R. Bröker and A. V. Sutherland. An explicit height bound for the classical modular polynomial. *Ramanujan J.*, 22(3):293–313, 2010.

[21] K. Chandrasekharan. *Elliptic functions*, volume 281 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1985.

[22] D. Charles. Complex multiplication tests for elliptic curves, 2004.

[23] S. Chowla. An extension of Heilbronn's class–number theorem. *The Quarterly Journal of Mathematics*, os-5(1):304–307, 01 1934.

[24] H. Cohen. *A course in computational algebraic number theory*, volume 138 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1993.

[25] S. Cooper. Inversion formulas for elliptic functions. *Proc. Lond. Math. Soc. (3)*, 99(2):461–483, 2009.

[26] D. A. Cox. *Primes of the form $x^2 + ny^2$*. Pure and Applied Mathematics (Hoboken). John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2013. Fermat, class field theory, and complex multiplication.

[27] J. E. Cremona and T. Thongjunthug. The complex AGM, periods of elliptic curves over $\mathbb{C}$ and complex elliptic logarithms. *J. Number Theory*, 133(8):2813–2841, 2013.

[28] D. Doud. A procedure to calculate torsion of elliptic curves over $\mathbf{Q}$. *Manuscripta Math.*, 95(4):463–469, 1998.

[29] L. van den Dries. *Tame topology and o-minimal structures*, volume 248 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1998.

[30] L. van den Dries and C. Miller. On the real exponential field with restricted analytic functions. *Israel J. Math.*, 85(1-3):19–56, 1994.

[31] R. Dupont. Fast evaluation of modular functions using Newton iterations and the AGM. *Mathematics of Computation*, 80(275):1823–1847, 2011.

[32] C.-G. Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, A28:1–19, 1942.

[33] A. M. Gabrièlov. Projections of semianalytic sets. *Funkcional. Anal. i Priložen.*, 2(4):18–30, 1968.

[34] I. García-Selfa, M. A. Olalla, and J. M. Tornero. Computing the rational torsion of an elliptic curve using Tate normal form. *J. Number Theory*, 96(1):76–88, 2002.

[35] C. F. Gauss. Circa seriem infinitatam $1 + \frac{\alpha\beta}{1.\gamma}x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1.2.\gamma(\gamma+1)}xx + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1.2.3.\gamma(\gamma+1)(\gamma+2)x^3} +$ etc. *Commentationes societatis regiae scientiarum Gottingensis recentiores*, II, 1813.

[36] C. F. Gauss. *Disquisitiones arithmeticae*. Springer-Verlag, New York, 1986. Translated and with a preface by Arthur A. Clarke, Revised by William C. Waterhouse, Cornelius Greither and A. W. Grootendorst and with a preface by Waterhouse.

[37] F. Grube. Ueber die Anziehungscomponente eines gerade elliptischen Cylinders in der Richtung der Axe, wenn die Elementaranziehung irgend einer Potenz der Entfernung umgekehrt proportional ist. *J. Reine Angew. Math.*, 65:62–73, 1866.

[38] D. Harvey and J. Van Der Hoeven. Integer multiplication in time O(n log n). working paper or preprint, Mar 2019.

[39] H. Hasse. Neue Begründung der komplexen Multiplikation. Erster Teil: Einordnung in die allgemeine Klassenkörpertheorie. *J. Reine Angew. Math.*, 157:115–140, 1927.

[40] K. Heegner. Diophantische Analysis und Modulfunktionen. *Math. Z.*, 56:227–253, 1952.

[41] H. Heilbronn. On the class–number in imaginary quadratic fields. *The Quarterly Journal of Mathematics*, os-5(1):150–160, 01 1934.

[42] L. K. Hua. *Introduction to number theory*. Springer-Verlag, Berlin-New York, 1982. Translated from the Chinese by Peter Shiu.

[43] F. Johansson. Arb: efficient arbitrary-precision midpoint-radius interval arithmetic. *IEEE Transactions on Computers*, 66:1281–1292, 2017.

[44] G. Jones and H. Schmidt. Pfaffian definitions of Weierstrass elliptic functions. *Mathematische Annalen*, 2020.

[45] M. Kaneko and D. Zagier. Supersingular $j$-invariants, hypergeometric series, and Atkin's orthogonal polynomials. In *Computational perspectives on number theory (Chicago, IL, 1995)*, volume 7 of *AMS/IP Stud. Adv. Math.*, pages 97–126. Amer. Math. Soc., Providence, RI, 1998.

[46] E. Kani. Idoneal numbers and some generalizations. *Ann. Sci. Math. Québec*, 35(2):197–227, 2011.

[47] L. V. Kantorovich and G. P. Akilov. *Functional analysis in normed spaces*. Translated from the Russian by D. E. Brown. Edited by A. P. Robertson. International Series of Monographs in Pure and Applied Mathematics, Vol. 46. The Macmillan Co., New York, 1964.

[48] A. G. Khovanskiĭ. *Fewnomials*, volume 88 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1991. Translated from the Russian by Smilka Zdravkovska.

[49] L. Kühne. An effective result of André-Oort type. *Ann. of Math. (2)*, 176(1):651–671, 2012.

[50] H. Labrande. Computing Jacobi's theta in quasi-linear time. *Math. Comp.*, 87(311):1479–1508, 2018.

[51] S. Lang. *Introduction to transcendental numbers*. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1966.

[52] S. Lang. *Complex multiplication*, volume 255 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1983.

[53] S. Lang. *Elliptic functions*, volume 112 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1987. With an appendix by J. Tate.

[54] M. H. Le. Upper bounds for class numbers of real quadratic fields. *Acta Arith.*, 68(2):141–144, 1994.

[55] D. H. Lehmer. The Mechanical Combination of Linear Forms. *Amer. Math. Monthly*, 35(3):114–121, 1928.

[56] The LMFDB Collaboration. The L-functions and modular forms database. `http://www.lmfdb.org`, 2021. [Online; accessed 25 May 2021].

[57] D. Marker. Khovanskii's theorem. In B. T. Hart, A. H. Lachlan, and M. A. Valeriote, editors, *Algebraic Model Theory*, pages 181–193. Springer Netherlands, Dordrecht, 1997.

[58] D. Masser. Rational values of the Riemann zeta function. *J. Number Theory*, 131(11):2037–2046, 2011.

[59] D. Masser. *Auxiliary Polynomials in Number Theory*. Cambridge Tracts in Mathematics. Cambridge University Press, 2016.

[60] B. Mazur. Modular curves and the Eisenstein ideal. *Inst. Hautes Études Sci. Publ. Math.*, (47):33–186 (1978), 1977. With an appendix by Mazur and M. Rapoport.

[61] M. Mignotte and M. Waldschmidt. Linear forms in two logarithms and Schneider's method. *Math. Ann.*, 231(3):241–267, 1977/78.

[62] H. L. Montgomery and P. J. Weinberger. Notes on small class numbers. *Acta Arith.*, 24:529–542, 1973/74.

[63] R. Paulin. An explicit André-Oort type result for $\mathbb{P}^1(\mathbb{C}) \times \mathbb{G}_m(\mathbb{C})$ based on logarithmic forms. *Publ. Math. Debrecen*, 88(1-2):21–33, 2016.

[64] J. Pila. Integer points on the dilation of a subanalytic surface. *Q. J. Math.*, 55(2):207–223, 2004.

[65] J. Pila. Mild parameterization and the rational points of a Pfaff curve. *Comment. Math. Univ. St. Pauli*, 55(1):1–8, 2006.

[66] J. Pila. On the algebraic points of a definable set. *Selecta Math. (N.S.)*, 15(1):151–170, 2009.

[67] J. Pila. Counting rational points on a certain exponential-algebraic surface. *Ann. Inst. Fourier (Grenoble)*, 60(2):489–514, 2010.

[68] J. Pila. O-minimality and the André-Oort conjecture for $\mathbb{C}^\times$. *Ann. of Math. (2)*, 173(3):1779–1840, 2011.

[69] J. Pila and J. Tsimerman. Ax-Lindemann for $\mathcal{A}_3$. *Ann. of Math. (2)*, 179(2):659–681, 2014.

[70] J. Pila and A. J. Wilkie. The rational points of a definable set. *Duke Math. J.*, 133(3):591–616, 2006.

[71] G. Pólya and G. Szegő. *Problems and theorems in analysis. Vol. II.* Springer-Verlag, New York-Heidelberg, german edition, 1976. Theory of functions, zeros, polynomials, determinants, number theory, geometry, Die Grundlehren der Mathematischen Wissenschaften, Band 216.

[72] K. Ramachandra. Contributions to the theory of transcendental numbers. I, II. *Acta Arith.*, 14:65–72; ibid. 14 (1967/1968), 73–88, 1967/68.

[73] M. Raynaud. Courbes sur une variété abélienne et points de torsion. *Invent. Math.*, 71(1):207–233, 1983.

[74] M. Raynaud. Sous-variétés d'une variété abélienne et points de torsion. In *Arithmetic and geometry, Vol. I*, volume 35 of *Progr. Math.*, pages 327–352. Birkhäuser Boston, Boston, MA, 1983.

[75] G. Robin. Estimation de la fonction de Tchebychef $\theta$ sur le $k$-ième nombre premier et grandes valeurs de la fonction $\omega(n)$ nombre de diviseurs premiers de $n$. *Acta Arith.*, 42(4):367–389, 1983.

[76] G. Robin. Grandes valeurs de la fonction somme des diviseurs et hypothèse de Riemann. *J. Math. Pures Appl. (9)*, 63(2):187–213, 1984.

[77] J. B. Rosser and L. Schoenfeld. Approximate formulas for some functions of prime numbers. *Illinois J. Math.*, 6(1):64–94, 03 1962.

[78] T. Schneider. Arithmetische Untersuchungen elliptischer Integrale. *Math. Ann.*, 113(1):1–13, 1937.

[79] I. G. Shevtsova. Refinement of estimates for the rate of convergence in Lyapunov's theorem. *Dokl. Akad. Nauk*, 435(1):26–28, 2010.

[80] C. Siegel. Über die classenzahl quadratischer zahlkörper. *Acta Arithmetica*, 1(1):83–86, 1935.

[81] J. P. Sorenson. Sieving for pseudosquares and pseudocubes in parallel using doubly-focused enumeration and wheel datastructures. In *Algorithmic number theory*, volume 6197 of *Lecture Notes in Comput. Sci.*, pages 331–339. Springer, Berlin, 2010.

[82] J. P. Sorenson and J. Webster. Two algorithms to find primes in patterns. *Math. Comp.*, 89(324):1953–1968, 2020.

[83] H. M. Stark. A complete determination of the complex quadratic fields of class-number one. *Michigan Math. J.*, 14:1–27, 1967.

[84] I. Stewart and D. Tall. *Complex analysis*. Cambridge University Press, Cambridge, 2 edition, 2018.

[85] J. Tsimerman. The André-Oort conjecture for $\mathcal{A}_g$. *Ann. of Math. (2)*, 187(2):379–390, 2018.

[86] M. Watkins. Class numbers of imaginary quadratic fields. *Math. Comp.*, 73(246):907–938, 2004.

[87] M. Watkins. A spectral proof of class number one. *Mathematische Zeitschrift*, 293(1):383–406, Oct 2019.

[88] P. J. Weinberger. Exponents of the class groups of complex quadratic fields. *Acta Arith.*, 22:117–124, 1973.

[89] A. J. Wilkie. Model completeness results for expansions of the ordered field of real numbers by restricted Pfaffian functions and the exponential function. *J. Amer. Math. Soc.*, 9(4):1051–1094, 1996.

[90] A. J. Wilkie. A theorem of the complement and some new o-minimal structures. *Selecta Math. (N.S.)*, 5(4):397–421, 1999.