A Computational Tour of Number Theory

Hao Chen R. Andrew Ohana Bharathwaj Palvannan
Travis Scholl Simon Spicer William Stein
Yannick Van Huele

March 9, 2014

Contents

Chapter 1

Introduction: tour of the main objects

In this book, we will explore several important central problems and objects of number theory, and for each to explain how—in practice (not just theory)—to *compute* with them. Books, papers, and web sites such as Wikipedia and Mathoverflow often give excellent descriptions of mathematical objects, algorithms, data, conjectures, and theorems. However, they rarely give concrete instructions so that you can manipulate them on a computer, with enough theoretical discussion so that you understand the limitations and capabilities of your tools. That is the mission of the book you are looking at.

Chapter 2

Class Numbers

Number fields are a natural generalization of the rational numbers $\mathbb{Q}$, to which ideas such as primes, and conjectures such as the Riemann Hypothesis, etc., all generalize. They have been intensely studied, initially motivated by work to prove Fermat's Last Theorem.

In this book we will assume you know abstract algebra; however, as a quick reminder, an *ideal* I in a commutative ring R is an additive subgroup of R such that for all $x \in R$ we have $xI \subset I$. A *number field* K is a finite algebraic extension of the rational numbers $\mathbb{Q}$; equivalently, it is a field obtained by adjoining a root α of some polynomial $f(x) \in \mathbb{Q}[x]$. The *ring of integers* $R = \mathcal{O}_K$ of a number field K is the set of all $\alpha \in K$ such that α is a root of some monic polynomial $f(x) \in \mathbb{Z}[x]$. It is interesting to prove that R (as defined) is closed under addition and multiplication (it is a ring). In the special case when $K = \mathbb{Q}$, we have $R = \mathbb{Z}$.

2.1 The ring of integers is a ring

As the name suggests, the ring of integers of a number field is in fact a ring. (Note, for example, that when $K = \mathbb{Q}$, the ring of integers is simply $\mathcal{O}_K = \mathbb{Z}$.) A sketch of the proof is given below. A more detailed exposition can be found in sections 2.2 and 2.3 of [Ste12].

To prove that $\mathcal{O}_K$ is a ring, let us first define a related object: Fix an algebraic closure $\bar{\mathbb{Q}}$ of the rational numbers. The set of *algebraic integers* $\bar{\mathbb{Z}}$ is the set of all $\alpha \in \bar{\mathbb{Q}}$ such that α is the root of some monic polynomial $f(x) \in \mathbb{Z}[x]$. Thus, if K is a number field, then – identifying K with a subfield of $\mathbb{Q}$ – we see that the ring of integers of K consists exactly of the algebraic integers lying in K: $\mathcal{O}_K = \bar{\mathbb{Z}} \cap K$.

Proposition 2.1.1. *The set $\bar{\mathbb{Z}}$ of algebraic integers is a ring.*

As $\bar{\mathbb{Z}}$ is a subset of $\bar{\mathbb{Q}}$, it suffices to show that $\bar{\mathbb{Z}}$ is closed under addition and multiplication. To do so, we use the following result:

Lemma 2.1.2. *Let $\alpha \in \bar{\mathbb{Q}}$. Then $\alpha \in \bar{\mathbb{Z}}$ if and only if $\mathbb{Z}[\alpha]$ is a finitely generated $\mathbb{Z}$-module.*

Proof. A proof of the lemma can be found in section 2.3 of [Ste12]. Now, suppose that $\alpha, \beta \in \bar{\mathbb{Z}}$ and note that

$$\mathbb{Z}[\alpha+\beta] \subseteq \mathbb{Z}[\alpha,\beta] \qquad \text{and} \qquad \mathbb{Z}[\alpha\beta] \subseteq \mathbb{Z}[\alpha,\beta]$$

By the lemma, both $\mathbb{Z}[\alpha]$ and $\mathbb{Z}[\beta]$ are finitely generated as $\mathbb{Z}$-modules. Let $\alpha_1, \ldots, \alpha_k$ and $\beta_1, \ldots, \beta_\ell$ be generators for $\mathbb{Z}[\alpha]$ and $\mathbb{Z}[\beta]$, respectively. Then, one can show that $\{\alpha_i\beta_j : 1 \leq i \leq k, 1 \leq j \leq \ell\}$ is a set of generators for the $\mathbb{Z}$-module $\mathbb{Z}[\alpha,\beta]$. Because $\mathbb{Z}$ is a noetherian ring and $\mathbb{Z}[\alpha,\beta]$ is a finitely generated $\mathbb{Z}$-module, $\mathbb{Z}[\alpha,\beta]$ is a noetherian $\mathbb{Z}$-module and, hence, every $\mathbb{Z}$-submodule is finitely generated. In particular, $\mathbb{Z}[\alpha+\beta]$ and $\mathbb{Z}[\alpha\beta]$ are both finitely generated $\mathbb{Z}$-modules. The lemma then tells us that $\alpha+\beta$ and $\alpha\beta$ are algebraic integers. □

Corollary 2.1.3. *Let K be a number field. Then $\mathcal{O}_K$, the ring of integers of K, is a ring.*

2.2 Arithmetic with ideals

We can multiply any two nonzero ideals $I, J \subset R$ by taking the ideal generated by all products of elements in I with elements in J. With this operation, the set of nonzero ideals is an infinite monoid (a group but without inverses). For example, if $R = \mathbb{Z}$, then the nonzero ideals are in bijection with positive integers, and the monoid is isomorphic to $\mathbb{Z}_{>0}$ under addition.

An ideal I is *principal* if there is some $\alpha \in R$ such that $I = \{\alpha b : b \in R\}$, in which case we write $I = (\alpha)$. Define an equivalence relation on nonzero ideals by $I \sim J$ if $I = J \cdot (\alpha)$ for some principal ideal (α). The *class group* $\mathrm{Cl}(R)$ of R is the quotient of the monoid of nonzero ideals modulo this equivalence relation, i.e., modulo the submonoid of principal ideals; it takes some work to show that the result is an abelian *group*, i.e., every ideal class has an inverse.

For example, if R is a principal ideal domain (PID), i.e., if every ideal is principal, then the class group is an abelian group of order 1. In the following example, we consider $\mathbb{Q}(\sqrt{-2013})$.

```
sage: K = QuadraticField(-2013); K
Number Field in a with defining polynomial x^2 + 2013
sage: C = K.class_group(); C
Class group of order 16 with structure C4 x C2 x C2 \
    of Number Field in i
with defining polynomial x^2 + 2013
sage: C.gens()
(Fractional ideal class (41, i + 23), Fractional \
    ideal class (47, i + 14),
 Fractional ideal class (2, i + 1))
sage: C.0 * C.1
Fractional ideal class (19, i + 1)
sage: (C.0)^4
Trivial principal fractional ideal class
```

2.3 Properties of the class group

One of the main theorems of algebraic number theory asserts that for any number field K, the class group $\mathrm{Cl}(R)$ is a *finite* abelian group. This immediately suggests some basic questions. For example, extending the computation above, if we let K vary over the imaginary quadratic fields $K = \mathbb{Q}(\sqrt{d})$ with $d \leq -1$ square free, we obtain a list of class numbers of these fields:

```
sage: h = lambda d : QuadraticField(d).class_number()
sage: v = [h(-d) for d in [1..500] if is_squarefree(-\
   d)]; v
[1, 1, 1, 2, 2, 1, 2, 1, 2, 4, 2, 4, 1, 4, 2, 3, 6, \
   6, 4, 3, 4, 4, 2, 2, 6,
4, 8, 4, 1, 4, 5, 2, 6, 4, 4, 2, 3, 6, 8, 8, 8, 1, 8,\
    4, 7, 4, 10, 8, 4, 5,
4, 3, 4, 10, 6, 12, 2, 4, 8, 8, 4, 14, 4, 5, 8, 6, 3,\
    6, 12, 8, 8, 8, 2, 6,
10, 10, 2, 5, 12, 4, 5, 4, 14, 8, 8, 3, 8, 4, 10, 8, \
   16, 14, 7, 8, 4, 6, 8,
10, 16, 1, 8, 10, 11, 12, 14, 12, 4, 8, 5, 10, 12, 8,\
    16, 12, 2, 4, 13, 4,
20, 4, 10, 9, 12, 6, 4, 8, 20, 20, 8, 3, 8, 6, 14, 8,\
    10, 4, 16, 12, 7, 8,
5, 10, 20, 12, 12, 2, 12, 8, 15, 12, 12, 6, 12, 7, 4,\
    16, 12, 16, 8, 4, 6,
13, 8, 20, 2, 22, 11, 8, 12, 6, 14, 20, 8, 3, 16, 12,\
    14, 20, 4, 18, 8, 6,
8, 8, 12, 10, 16, 3, 12, 8, 19, 8, 26, 10, 12, 10, \
   20, 8, 4, 22, 12, 24, 8,
3, 12, 18, 8, 6, 28, 8, 10, 5, 14, 16, 16, 4, 8, 6, \
   19, 18, 20, 12, 9, 12,
8, 10, 28, 16, 3, 20, 8, 17, 8, 20, 22, 16, 14, 12, \
   10, 8, 6, 20, 16, 20,
16, 2, 16, 16, 16, 16, 6, 20, 10, 12, 8, 9, 10, 10, \
   24, 2, 16, 12, 21, 12,
24, 4, 20, 8, 15, 8, 5, 8, 32, 14, 20, 6, 12, 14, 20,\
    8, 26, 30, 8, 7, 16,
8, 7, 16, 20, 16, 12, 20, 8, 25, 16, 20, 4, 20, 7, \
   20, 9, 12, 28, 24, 8, 3]
```

Looking at this list, it is natural to guess that 1 occurs only finitely many times – in fact, Gauss noticed that 1 only appears 9 times:

```
sage: v.count(1)
9
```

and *conjectured* that the corresponding 9 fields are the only quadratic imaginary fields with class number 1. Heegner [Hee52] (and independently, Stark [Sta69]) later proved this conjecture. Also, deep work of Goldfeld and Gross-Zagier involving elliptic curve L-functions, yielded an algorithm to find all quadratic imaginary fields with given class number, which Mark Watkins has made much more efficient in his thesis work. (The crucial input here was a proof that if E is the elliptic curve $y^2+y=x^3-7x+6$, then $\mathrm{ord}_{s=1} L(E,s)=3$, where $L(E,s)$ is the L-series of E.)

What about the other direction: positive d?

```
sage: h = lambda d : QuadraticField(d).class_number()
sage: v = [h(d) for d in [2..500] if is_squarefree(d)\
    ]; v
[1, 1, 1, 1, 1, 2, 1, 1, 1, 2, 1, 1, 1, 1, 1, 2, 1, \
    2, 1, 1, 2, 2, 1, 1,
2, 1, 2, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 2, 2, 1, 1,\
     2, 1, 1, 2, 1, 2,
3, 4, 1, 2, 1, 2, 1, 2, 1, 1, 2, 1, 1, 2, 1, 2, 2, 1,\
     1, 2, 2, 1, 2, 2,
1, 2, 2, 2, 1, 1, 4, 1, 1, 1, 1, 2, 1, 1, 3, 2, 4, 2,\
     1, 1, 2, 2, 1, 1,
2, 1, 1, 2, 1, 1, 4, 1, 2, 1, 2, 1, 1, 2, 2, 2, 2, 2,\
     2, 1, 1, 2, 4, 1,
1, 1, 2, 2, 2, 1, 1, 4, 1, 1, 1, 2, 1, 2, 4, 2, 2, 3,\
     8, 1, 3, 2, 4, 1,
6, 1, 2, 1, 1, 2, 2, 1, 1, 1, 3, 4, 3, 2, 2, 1, 1, 2,\
     2, 2, 1, 1, 2, 4,
1, 1, 1, 2, 1, 2, 2, 2, 4, 4, 1, 2, 2, 2, 1, 1, 2, 2,\
     1, 1, 2, 1, 1, 2,
1, 2, 2, 3, 4, 4, 3, 2, 1, 4, 1, 1, 2, 1, 2, 1, 2, 6,\
     1, 1, 1, 2, 2, 2,
1, 3, 2, 2, 2, 1, 4, 2, 1, 2, 2, 1, 1, 1, 1, 2, 2, 1,\
     4, 2, 1, 2, 2, 1,
1, 8, 5, 2, 2, 2, 2, 1, 4, 2, 1, 2, 1, 2, 1, 1, 1, 2,\
     6, 2, 2, 1, 1, 4,
4, 1, 4, 5, 8, 3, 4, 1, 2, 1, 2, 1, 1, 4, 1, 2, 1, 4,\
     1, 2, 2, 1, 3, 2,
2, 3, 2, 1, 1, 2, 2, 4, 2, 1, 1, 1, 2, 2, 1, 2, 5]
sage: v.count(1)
141
```

It's natural to guess that there there are infinitely many real quadratic fields with class number 1. This is an unsolved problem, though it is supported by the Cohen-Lenstra heuristics. In fact, even proving that there are infinitely many number fields (not just quadratic fields) with class number 1 is an unsolved problem.

HW (Volunteer): *What proportion of real quadratic fields are predicted to have class number* 1*?*

The Cohen-Lenstra heuristics [CL84] predicts that 75.446 % of real quadratic fields have class number 1, which agrees with computations by Riele and Williams [RW03].

2.4 Quadratic fields with nontrivial class group

One may also ask about quadratic number fields with class number greater than 1. Are there infinitely many real quadratic number fields with class number greater than 1? Are there real quadratic fields with arbitrarily large class number? The answer to both of these questions is yes and in fact, one can obtain a lower bound on the power of 2 dividing the class number of $\mathbb{Q}(\sqrt{d})$.

Proposition 2.4.1. *Let d be a square-free integer and let $K = \mathbb{Q}(\sqrt{d})$. Define*

$$d_K = \begin{cases} d & \text{if } d \equiv 1 (\text{mod } 4), \\ 4d & \text{if } d \equiv 2 \text{ or } 3 (\text{mod } 4). \end{cases}$$

Let t denote the number of distinct primes dividing d_K. If $t > 1$ then $2^{t-2} | h_K$, where h_K denotes the class number of K.

In our proof, we will assume some knowledge of class field theory, although one can also prove this result by carefully studying quadractic forms (see for example Section 3.8 of [BS66]). Let us first illustrate the idea of the proof by way of some examples. Let $d = 1105 = 5 \cdot 13 \cdot 17$ and let $K = \mathbb{Q}(\sqrt{1105})$. The following three quadratic (and thus abelian) extensions of K:

$$L_1 = K(\sqrt{5}) = \mathbb{Q}(\sqrt{5}, \sqrt{1105}), \qquad L_2 = K(\sqrt{13}), \quad \text{and} \quad L_3 = K(\sqrt{17}).$$

Note that L_1 is the compositum of K with $\mathbb{Q}(\sqrt{5})$, which is unramified outside of 5 ($5 \equiv 1 (\text{mod } 4)$). Hence, the extension L_1/K is unramified outside of 5. However, L_1 is also the compositum of K with $\mathbb{Q}(\sqrt{13 \cdot 17})$, which is unramified outside of 13 and 17 so the extension L_1/K is in fact unramified. Similarly, one can show that the extensions L_2/K and L_3/K are also unramified. Let L denote the compositum of these three extensions: $L = L_1 L_2 L_3$. Then the extension L/K is both abelian and unramified. Note however that

$$L = L_1 L_2 L_3 = \mathbb{Q}(\sqrt{5}, \sqrt{13}, \sqrt{17}, \sqrt{1105}) = \mathbb{Q}(\sqrt{5}, \sqrt{13}, \sqrt{1105}) = L_1 L_2$$

is actually a degree 4 extension of K. Let H denote the Hilbert class field of K. Then $4 = [L : K] | [H : K] = h_K$. So in fact a higher power of 2 than predicted in the proposition divides the class number of K. The difference is in part due to the fact that we are working with the ideal class group rather than the narrow ideal class group which arises somewhat more naturally in this context as will be seen in the proof of the proposition. In this particular example, the class number turns out to be exactly 4, as can be computed using Sage:

```
sage: QuadraticField(5*13*17).class_number()
4
```

Let us now prove the proposition.

Proof. Let d be a square free integer and let $p_1, p_2, \ldots, p_k$ be the odd primes dividing d (so that $d_K \in \{\pm 2^\ell p_1 \cdots p_k : \ell 0, 2, 3\}$ and $k = t$ or $k = t - 1$). For each i, let $p_i^* = (-1)^{(p-1)/2} p$. That is, p_i^* is plus or minus p_i with the sign chosen so that $p_i^* \equiv 1 (\bmod 4)$. Let $L_i = K(\sqrt{p_i^*})$. Then L_i/K is an abelian extension which is unramified at all finite primes. To see this, consider the following tower:

$$\begin{tikzcd}[arrows=dash]
 & L_i = \mathbb{Q}(\sqrt{d}, \sqrt{p_i^*}) \arrow[dl] \arrow[d] \arrow[dr] & \\
K = \mathbb{Q}(\sqrt{d}) \arrow[dr] & \mathbb{Q}(\sqrt{p_i^*}) \arrow[d] & \mathbb{Q}(\sqrt{d/p_i^*}) \arrow[dl] \\
 & \mathbb{Q} &
\end{tikzcd}$$

The field L_i is the compositum of K with either of the two other quadratic subfields of L_i, $\mathbb{Q}(\sqrt{p_i^*})$ and $\mathbb{Q}(\sqrt{d/p_i^*})$. Note that p_i is the only finite prime ramifying in $Q(\sqrt{p_i^*})/\mathbb{Q}$, so no other finite prime can be ramified in L_i/K. However, p_i does not ramify in $\mathbb{Q}(\sqrt{d/p_i^*})$ and, hence, does not ramify in L_i/K. Thus, for each $1 \leq i \leq k$, the extension L_i/K is unramified at all finite primes. Furthermore, L_i/K is a degree 2 extension so is abelian. Let L denote the compositum of all the L_i so that

$$L = \mathbb{Q}(\sqrt{d}, \sqrt{p_1^*}, \ldots, \sqrt{p_k^*})$$

Then L/K is an abelian extension of K which is unramified at all finite primes. If $d \equiv 2$ or $3 (\bmod 4)$, then we see that L/K is an extension of degree 2^k. On the other hand, if $d \equiv 1 (\bmod 4)$, we see that L/K is an extension of degree 2^{k-1}. In either case, L/K is an abelian extension of degree 2^{t-1} which is unramified at all finite primes of K (recall that t is the number of odd primes dividing d_K).

Let $\mathfrak{m}_\infty$ denote the modulus which is the product of all real primes of K so that the corresponding ray class group is $\mathrm{Cl}^+(K)$, the narrow (or strict) class group of K. It follows from class field theory that L is contained in the ray class field of K corresponding to $\mathfrak{m}_\infty$ and, thus, that $\mathrm{Gal}(L/K)$ is isomorphic to a quotient of $\mathrm{Cl}^+(K)$. So, in particular, 2^{t-1} divides the order of $\mathrm{Cl}^+(K)$. Let

$U_K = \mathcal{O}_K^\times$ denote the units of $\mathcal{O}_K$ and let U_K^+ denote the totally positive units of $\mathcal{O}_K$. Then there is an exact sequence:

$$1 \to U_K/U_K^+ \to \prod_{v \mid \mathfrak{m}_\infty} \{\pm 1\} \to \mathrm{Cl}^+(K) \to \mathrm{Cl}(K) \to 1$$

If $d < 0$, $K = \mathbb{Q}(\sqrt{d})$ has no real primes and we see that $\mathrm{Cl}^+(K)$ and $\mathrm{Cl}(K)$ are the same. If $d > 0$, K has two real primes. Dirichlet's Unit Theorem tells us that $U_K \approx \{\pm 1\} \times \mathbb{Z}$. Let ε be a fundamental unit of K: i.e., let $\varepsilon \in U_K$ such that $U_K = \langle \pm\varepsilon \rangle$. Then $U_K^+ = \langle \varepsilon \rangle$ or $U_K^+ = \langle \varepsilon^2 \rangle$. Hence, the quotient U_K/U_K^+ has order 2 or 4. It follows that $\mathrm{Cl}^+(K)$ has order either $2h_K$ or h_K and, hence, that 2^{t-2} always divides h_K. □

Chapter 3

Fermat, Faltings and ABC

3.1 Falting's theorem

Let $f(x,y) = 0$ be a smooth plane curve. If we consider the set of solutions to this equation over the complex numbers there is the natural notion of the *genus* of the curve: basically, the number of holes in the solution set, when the solution set is viewed as a Riemann surface. Genus one curves, for example, topologically look like doughnuts.

It turns out that the number of rational points on a plane curve – that is, the number of pairs of rational numbers (x,y) that satisfy $f(x,y) = 0$ – is highly dependent on the genus of the curve described by f. Genus zero curves are generally well understood: the solution set to $f(x,y) = 0$ is either empty, or it is infinite and described by a single parameter.

Genus one curves (modulo some extra conditions) are known as *elliptic curves*, and are the central objects of study in a large part of modern-day number theory. These will be discussed in later sections and chapters of the book, but the gist of the matter is that elliptic curves can have wildly varying numbers of rational points depending on the curve, from zero to infinity and everything in between; there is an entire industry dedicated to computing just how many points a particular elliptic curve has.

Curves of genus two and above, on the other hand, are a completely different affair.

Theorem 3.1.1. *Any smooth projective curve of genus two or greater has only finitely many rational points.*

This is known as *Falting's Theorem*, after the German mathemetician Gerd Falting who proved it in 1983 (for which he won a Fields Medal). The theorem is in fact more general than stated above: given any fixed number field K, the number of K-rational points on a smooth projective curve C of genus at least two is finite.

Unfortunately, Falting's Theorem is completely non-constructive: while it tells you that there will only ever be finitely many rational solutions to an equation describing a genus two or above curve, in no way does it outline how to go about finding those points. In fact, there is no general method known to find points on curves of higher genus; nor is it even known – or even suspected – if such a method exists or doesn't exist. The question, as they say, is wide open.

It thus shouldn't be a surprise that the question of finding points on higher genus curves is in itself an active area of research in number theory. As an example, in Diophantus' Arithmetica, the series of books on solving algebraic equations by the ancient Greek mathematician, there is only one instance of a higher genus curve being considered: problem 17 of book 6 boils down to finding rational points on the genus two curve described by

$$y^2 = x^6 + x^2 + 1$$

Because of Falting's Theorem we know that this equation only has finitely rational solutions. Two obvious solutions are $(x, y) = (0, \pm 1)$, while Diophantus himself gave the more interesting solution $(\frac{1}{2}, \pm\frac{9}{8})$. This single equation is the entire topic of the PhD thesis of Joe Wetherell (UC Berkeley 1998) [Wet97], which uses some pretty hefty algebraic geometry to show that no other solutions exist.

3.2 Fermat's last theorem

Fermat's Last Theorem – a problem from the 1600s! – asserts that whenever $n \geq 3$, then there are no positive integer solutions to the Diophantine equation

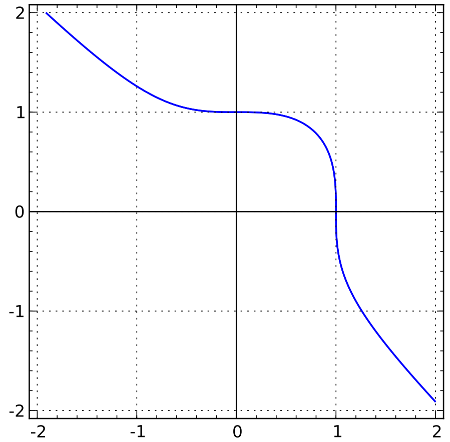
$$X^n + Y^n = Z^n.$$

This seemingly approachable problem has a long and colorful history. When $n = 3$ the equation is

$$X^3 + Y^3 = Z^3,$$

which is the (projective homogenous) equation of the plane cubic algebraic curve $x^3 + y^3 = 1$:

```
%var x y
implicit_plot(x^3 + y^3 == 1, (x,-2,2), (y,-2,2))
```



This is an example of an elliptic curve; you should see some rational points on it—namely $(1, 0)$ and $(0, 1)$—these correspond to torsion points on the elliptic curve, and to solutions to the Fermat equation with X or Y equal to 0.

For general $n \geq 5$, Fermat's assertion was finally resolved by much modern work in number theory, culminating with a major theorem of Andrew Wiles. The basic strategy, which is useful for attacking a wide range of Diophantine equations, is as follows. Given a specific counterexample $a^n + b^n = c^n$ to Fermat's claim (with say n prime), we associate an elliptic curve (the "Frey curve")

$$E: \quad y^2 = x(x - a^n)(x + b^n).$$

By counting solutions modulo p to the equation that defines this curve (or any nonsingular curve of the form $y^2 = x^3 + \alpha x + \beta$ for that matter), we define an L-series

$$L(E, s) = \prod_{p \text{ prime}} \frac{1}{1 - a_p p^{-s} + \varepsilon(p) \cdot p^{1-2s}} = \sum_{n \geq 1} \frac{a_n}{n^s},$$

where $\varepsilon(p) = 1$ if $p \nmid abc$ and $\varepsilon(p) = 0$ otherwise, and $a_p = p + 1 - \#E(\mathbb{F}_p)$. Old work of Hasse (and others) shows that this L-series defines a complex analytic function on some right half plane; people then conjectured (and subsequently proved!) that $L(E, s)$ extends to a holomorphic function on all $\mathbb{C}$. Birch and Swinnerton-Dyer even went so far as to conjecture that most everything about the arithmetic of E is determined by the leading coefficient of the Taylor expansion of $L(E, s)$ about $s = 1$ (see Chapter 7).

To the L-series $L(E, s)$, one can use Mellin transform to define another complex analytic function on the upper half plane

$$f(z) = \sum_{n \geq 1} a_n e^{2\pi i z} = \sum_{n \geq 1} a_n q^n,$$

where the coefficients a_n here are exactly the same as in the Dirichlet series representation of $L(E, s)$ above. Andrew Wiles (and Richard Taylor) proved

[Wil95]—using an ingenious argument involving an arithmetic analogue of Barry Mazur's deformation theory—that $f(z)$ has the property that for all 2×2 integer matrices γ with determinant 1 and lower left entry divisible by $N = \prod_{\ell | abc} \ell$, we have

$$f(z)dz = f(\gamma(z))d(\gamma(z)),$$

where γ acts on the upper half plane via linear fractional transformations. This function $f(z)$ is called a weight 2 cuspidal *modular form.*

Moreover, using arithmetic with *quaternion algebras* and geometry of modular curves, Ken Ribet had earlier proved that the coefficients a_n of the expansion of $f(z)$ must be congruent to the coefficients of a cuspform of "level 2", which is impossible, since there are no such nonzero forms. This contradiction proves Fermat's last theorem.

We can explicitly compute with many of the objects—elliptic curves, modular forms, etc.—appearing in the discussion above. For example, we do some computations with the elliptic curve corresponding to Fermat's equation for exponent $n = 3$ (this is *not* the corresponding Frey curve, but another model for $X^3 + Y^3 = 1$)

```
sage: E = EllipticCurve([0,0,1,0,-7]); E
Elliptic Curve defined by y^2 + y = x^3 - 7 over \
   Rational Field
sage: E.anlist(20)
[0, 1, 0, 0, -2, 0, 0, -1, 0, 0, 0, 0, 0, 5, 0, 0, 4,\
    0, 0, -7, 0]
sage: L = E.lseries(); L
Complex L-series of the Elliptic Curve defined by y^2\
    + y = x^3
over Rational Field
sage: L(1)
0.588879583428483
sage: L.taylor_series(1, 5)
0.59 + 0.45*z - 0.19*z^2 + 0.0042*z^3 + 0.033*z^4 - \
   0.018*z^5 + O(z^6)
sage: E.rank() # proves FLT for exponent 3
0
```

By computing the ranks of twists of E we can tell whether or not E has infinitely many rational points over the quadratic field $\mathbb{Q}(\sqrt{d})$, for various d. Running this code:

```
for d in [-5..5]:
    if d.is_squarefree() and d != 1:
       print d, E.quadratic_twist(d).rank()
```

yields:

```
-5 1
-3 0
```

```
-2 1
-1 0
2 1
3 0
5 1
```

This means, e.g., that in the field $\mathbb{Q}(\sqrt{5})$, generated by the golden ratio, there are infinitely many triples of coprime integers X, Y, Z such that $X^3 + Y^3 = Z^3$! The following code gives a few of these triples:

```
sage: f = EllipticCurve_from_cubic(x^3+y^3-z\
   ^3,[-1,1,0]).inverse();
sage: for i in range(1,6):
        r = f(i*f.domain().gens()[0]); r.\
           clear_denominators(); r
```

which outputs:

```
(-sqrt5 + 9 : sqrt5 + 9 : 12)
(2761*sqrt5 + 225 : -2761*sqrt5 + 225 : 3720)
(-3105301*sqrt5 + 6666948 : 3105301*sqrt5 + 6666948 :\
    13610574)
(7206070204127*sqrt5 + 4735673343225 :
    -7206070204127*sqrt5 + 4735673343225 :
    19652111284080)
(-45203693464083879145*sqrt5 + 25364447357048949231 :
    45203693464083879145*sqrt5 + 25364447357048949231\
        :
    116655528338821919868)
```

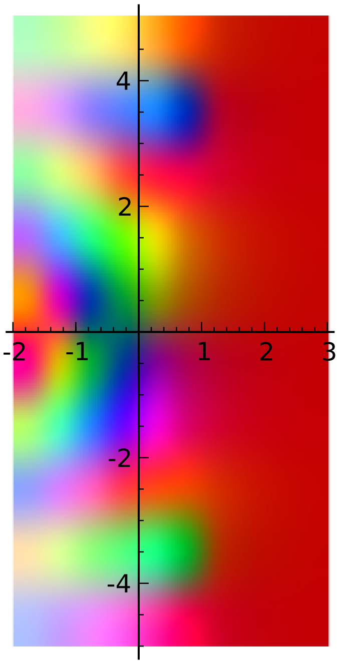
We have a short exact sequence

$$0 \to E(\mathbb{Q}) \to E(\mathbb{Q}(\sqrt{5})) \to E(\mathbb{Q}(\sqrt{5}))/E(\mathbb{Q}) \to 0. \tag{3.2.1}$$

Using Sage, we can compute the torsion and rank of each group. One obtains that $E(\mathbb{Q}) \cong \mathbb{Z}/3\mathbb{Z}$ and $E(\mathbb{Q}(\sqrt{5})) \cong \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}$. The group G acts on $E(\mathbb{Q}(\sqrt{5}))/E(\mathbb{Q})$ by -1. One can ask the question - is there an element x in $E(\mathbb{Q}(\sqrt{5}))$ on which G acts by -1 and such that $E(\mathbb{Q}(\sqrt{5})) \cong \mathbb{Z}\cdot < x > \oplus E(\mathbb{Q})$ (where the isomorphism is in the category of $\mathbb{Z}[G]$-modules). Let us write $E(\mathbb{Q}(\sqrt{5})) \cong \mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}$ (where the isomorphism is simply in the category of abeian groups). Let b be an element of $\mathbb{Z}/3\mathbb{Z}$ such that $\sigma(1,0) = (-1,b)$. Then one sees that $\sigma \cdot (1,b) = -(1,b)$. One can choose x to be $(1,b)$.

And we can explore the L-series itself:

```
sage: h = L.dokchitser(prec=200)
sage: complex_plot(h, (-2,3), (-5,5), plot_points=10)\
     # quite painful
```



As for the Riemann zeta function, there is a generalized Riemann Hypothesis, which postulates that the nontrivials zeros of $L(E, s)$ lie on a line, the line $\mathrm{Re}(s) = 1$. The first few imaginary parts of the nontrivial zeros of $L(E, s)$ are

```
sage: L.zeros(5)
[4.04304401, 6.04893540, 8.21765037, 9.42919921, \
    10.9087283]
```

The general structure of Wiles's approach also generalizes to many other Diophantine equations. Instead of obtaining a single congruence with forms in a space of dimension 0, we instead considers a potential large collection of spaces of modular forms, which we systematically compute (e.g., using modular symbols).

3.2.1 Extending Fermat's last theorem

There are many rings where Fermat's Last Theorem fails for many values of n. For example consider the ring $\mathbb{Z}[\zeta]$ where ζ is a primitive mth root of unity with $3|m$. Then $\xi := \zeta^{\frac{m}{3}}$ is a primitive 3rd root of unity so it satisfies $1 + \xi + \xi^2 = 0$. From this it is easy to see

$$(\xi)^{6k+5} + (\xi^2)^{6k+5} = (-1)^{6k+5}$$

for any non-negative integer k. Hence the triple $(\xi, \xi^2, -1)$ satisfies the Fermat equation $X^{6k+5} + Y^{6k+5} = Z^{6k+5}$ in the ring $\mathbb{Z}[\zeta]$.

In general, it is straightforward to construct counterexamples to FLT of degree d with degree d number fields:
Let $d \geq 3$ be a positive integer, and pick any two positive integers Y and Z with $Y < Z$. Let $K = \mathbb{Q}(\alpha)$, where α obeys $\alpha^d = Z^d - Y^d \in \mathbb{Z}$. Then clearly (α, Y, Z) is a Fermat triple lying in K.

However, it is still possible to extend Fermat's Last Theorem to number fields. Olivier Debarre and Matthew Classen [**?**] extend the results of Faltings to show, in particular:

Theorem 3.2.1 (Debarre-Classen)**.** *The equation $x^n + y^n = z^n$ has only finitely many solutions (x, y, z), where (x, y, z) lie in any field of degree $d \leq n - 2$. (So we're considering solutions in all of these fields simultaneously.)*

The theorem states that there are only a few counterexamples to Fermat's Last Theorem for each exponent n, except over number fields of degree $n - 1$ or greater. Frazer Jarvis and Paul Meekin [**?**] take a guess as to what these counterexamples are:

Conjecture 3.2.2 (Jarvis-Meekin)**.** *Let K be a number field of degree d and $n \geq d+2$. Then any solution (x, y, z) over K to $x^n + y^n = z^n$ satisfies $x + y = z$.*

3.3 The ABC conjecture

The abc conjecture concerns triples of coprime positive integers a, b, c such that

$$a + b = c.$$

The *radical* of an integer n is the product of the distinct prime divisors of n, and the radical of an abc triple a, b, c is $r = r(a, b, c) = r(abc)$. For example,

$$1 + 8 = 9$$

has radical $r = 2 \cdot 3 = 6$.

Conjecture 3.3.1 (Masser-Oesterlé)**.** *For every $\varepsilon > 0$ there are only finitely many triples a, b, c of coprime positive integers such that $a + b = c$ and $r^{1+\varepsilon} \leq c$, where $r = r(a, b, c)$.*

Given a, b it is trivial to find c with $a + b = c$, and usually the radical $r = r(a, b, c)$ is going to be bigger than c. A typical example is $a = 4$, $b = 15$. We have $c = 4 + 15 = 19$, and the radical is

$$2 \cdot 3 \cdot 5 \cdot 19 = 570 \gg 19.$$

For the radical to be small compared to c is quite special.

Now suppose for the moment that we have a counterexample to Fermat's Last Theorem (see Section 3.2 above), say

$$A^n + B^n = C^n,$$

with A, B, C coprime positive integers and $n \geq 5$ (say). Letting $a = A^n, b = B^n, c = C^n$, we have

$$r(a,b,c) = r(A^n, B^n, C^n) = r(A,B,C) \leq ABC < C^3 < C^n = c.$$

According to Conjecture 3.3.1 (with any choice of ε), there can be only finitely many such triples (A^n, B^n, C^n). In particular, Conjecture 3.3.1 implies Fermat's Last Theorem for all sufficiently large n. [1]

There's much computational work that goes into understanding and refining the ABC conjecture. For example, Lenstra defined a notion of the *quality* of a triple $a + b = c$ to be

$$q(a,b,c) = \frac{\log(c)}{\log(r)}.$$

Here is a Sage interact to compute the quality:

```
@interact
def f(a=3, b=4):
    c = a + b
    print "%s + %s = %s"%(a,b,c)
    v = prod(set(prime_divisors(a) + prime_divisors(b\
        ) + prime_divisors(c)))
    q = log(c)/log(v)
    print "quality = ", float(q)
```

Notice that if we chose ε to get equality in Conjecture 3.3.1, then $q(a,b,c) = 1 + \varepsilon$. As mentioned above, triples a, b, c usually have very low quality, since r is typically much bigger than c. However, there are some known triples a, b, c of high quality, but the ABC conjecture asserts there aren't too many. In fact, here is an equivalent version of the ABC conjecture:

Conjecture 3.3.2. *For every $h > 1$ there are only finitely many triples a, b, c with quality bigger than h.*

The highest quality triple ever found[2] is $2 + 3^{10} \cdot 109 = 23^5$, where

$$q(a,b,c) = \frac{\log(23^5)}{\log(2 \cdot 3 \cdot 109 \cdot 23)} = 1.62991168412\ldots$$

Even if you don't believe the ABC conjecture, you might believe this:

Conjecture 3.3.3 (Weak ABC)**.** *Among all a, b, c triples, there is an absolute upper bound on the quality.*

[1] An accepted proof of FLT is known (due to Wiles). There is no accepted proof of ABC, but there is a claimed one, which may or may not be right...

> "The problem with wrong proofs to correct statements is that it is hard to give a counterexample." – Hendrik Lenstra, http://www.ucs.louisiana.edu/~avm1260/lenstra.html

[2] See http://www.math.leidenuniv.nl/~desmit/abc/.

Exercise 3.3.4. Does the above weaker conjecture imply Fermat's Last Theorem for all sufficiently large exponents n?

Exercise 3.3.5. If possible, formulate ABC over number fields in terms of a notion of "quality". See `http://www.math.columbia.edu/~goldfeld/ABC-Conjecture.pdf` for a precise statement of ABC over number fields. What does this imply about FLT over number fields?

In the paper mentioned in Exercise 1.5.5, Goldfeld formed a slightly different version of ABC:

(ABC')Let A, B, C be nonzero, coprime integers such that $A + B + C = 0$. Let $N = \prod_{p|ABC} p$. Then for every $\epsilon > 0$, there exists $\kappa(\epsilon) > 0$ such that

$$\max(|A|, |B|, |C|) < \kappa(\epsilon)N^{1+\epsilon}$$

The first question is if this conjecture is equivalent to ABC. One sees that ABC implies ABC': indeed, fix any $\epsilon > 0$, then κ can be chosen to be any number greater than $\max\{\frac{\max\{|A|,|B|,|C|\}}{N^{1+\epsilon}}\}$. This maximum taken over a set is finite since by ABC all but finitely many element in this set is smaller than 1.

Proof that ABC' implies ABC: Fix $\epsilon > 1$, choose δ s.t. $\epsilon > \delta > 1$, then by ABC', there exists some positive constant $\kappa(\delta)$ such that $\max(|A|, |B|, |C|) < \kappa(\delta)N^{1+\delta}$ holds for all triples. Now for any triple s.t. $N^{1+\epsilon} < \max(|A|, |B|, |C|)$ we have $N^{1+\epsilon} < N^{1+\delta}$, which can only hold for finitely many N. Hence N is bounded above, and we see from ABC' that $\max(|A|, |B|, |C|)$ is bounded above. So there's finitely many such triples.

To generalize this to number field, one needs to extends the notion of absolute value and radical by the following: Let K be a number field and V_K denotes the set of primes of K. Define

$$H_K(a, b, c) = \prod_{v \in V_K} \max\{|a|_v, |b|_v, |c|_v\}$$

and

$$rad_K(a, b, c) = \prod_{P \in IK(a,b,c)} N_{K/\mathbb{Q}}(P)$$

where $I_K(a, b, c)$ is the set of all prime ideals P of OK for which $|a|_v, |b|_v, |c|_v$ are not equal. The abc conjecture for K states that for any $\epsilon > 0$, there exists $\kappa(\epsilon, K) > 0$ such that

$$H_K(a, b, c) < \kappa(\epsilon, K)rad_K(a, b, c)^{1+\epsilon}$$

3.3.1 Infinitely many triples with quality greater than 1

For p an odd prime and $n \geq 1$ an integer, consider the a, b, c triple

$$a = 1, \qquad b = 2^{p(p-1)n} - 1, \qquad c = 2^{p(p-1)n}.$$

Since $\varphi(p^2) = p(p-1)$ is the order of $(\mathbb{Z}/p^2\mathbb{Z})^*$, we have $2^{p(p-1)n} \equiv 1 \pmod{p^2}$, so $p^2 \mid b$. Thus

$$r = r(a,b,c) = 1 \cdot p \cdot (\text{other stuff from } b) \cdot 2 \leq \frac{2b}{p} < b < c,$$

so

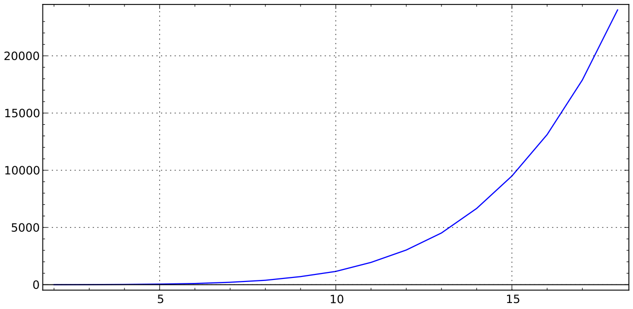
$$q(a,b,c) = \frac{\log(c)}{\log(r)} > 1.$$

Here is a Sage interact to compute a, b, c as above, along with their quality. Note that if you put at all largve values in for either n or p, then the resulting numbers are huge, and the computation of the quality (as implemented) will take forever.

```
@interact
def f(p=3, n=1):
    if not is_prime(p) or p<=2:
        print "p must be an odd prime!"
    if n < 1:
        print "n must be >= 1"
    a = 1
    b = 2^(p*(p-1)*n) - 1
    c = 2^(p*(p-1)*n)
    print "%s + %s = %s"%(a,b,c)
    print "(now computing quality...)"
    sys.stdout.flush()
    v = prod(set(prime_divisors(a) + prime_divisors(b\
       ) + prime_divisors(c)))
    q = log(c)/log(v)
    print "quality = ", float(q)
```

3.3.2 Triples with quality at least 1.4

The following is a plot of the number of abc triples with $c \leq 10^x$ and quality at least 1.4, drawn on a log scale, where $x \leq 10^{18}$. This data was computed by the ABC@HOME project.



The (strong) ABC conjecture asserts that the blue line is bounded above! As is typical is typical with most questions in "asymptotic number theory", the numerical data suggests—to the naive observer—the exact opposite. See [BMSW07] for an article about this sort of tension in another context:

> "We have a network of heuristics and conjectures regarding rational points, and we have massive data accumulated to exhibit instances of the phenomena. Generally, we would expect that our data supports our conjectures; and if not, we lose faith in our conjectures. But here there is a somewhat more surprising interrelation between data and conjecture: they are not exactly in open conflict one with the other. We discuss various aspects of this story, including recent heuristics and data that attempt to resolve this mystery."

Student Project Idea: (Andrew Ohana is mainly doing this) *Completely forget about the ABC conjecture, pretend you are an applied mathematician, and model the function that counts the number of ABC triples of quality q up to 10^x.*

1. Beg, barrow, or steel the complete dataset from the ABC@Home project. This could be pretty challenging, since currently their site is down with "`Warning: mysql_pconnect(): Too many connections in /data/project/abc/html/inc/db.inc on line 39...`". That said, I know the people who run this site, and we can contact them directly.

2. For a few hundred values of q with $1 \leq q \leq 1.67$, use the data to "compute or plot" the function

   $$f_q(x) = \#\{\text{number of ABC-triples with } c < 10^x \text{ of quality} \geq q\}.$$

   The plot will look like the blue curve above.

3. Use various models of your choosing to find a smooth function that best fits that curve? Your smooth function will be parametrized by q, i.e., you get a different function for each value of q. I don't have any idea what sort of function to expect. A polynomial? Exponential? Something else? Note that according to the ABC conjecture, the function $f_q(x)$ is supposed to be *bounded above* for each $q > 1$! (However, just pretend you don't know about that and you're a physicist or something....)

4. What happens if you let the parameter q in the above model go to ∞?

5. Can you find a different model for $f_q(x)$ that *is* bounded above (hence consistent with the ABC conjecture), and also fits the data?

The above seems to me to be the most obvious first thing to do if one had computing power and wanted to make the ABC conjecture in the first place. However, the data to do the above computation wasn't available until very recently, so maybe nobody did it (which I find sort of shocking if true).

3.4 Some consequences of the ABC conjecture

1. In [Ich98], assuming the generalized ABC conjecture for number fields, H. Ichimura shows that for a given real quadratic field k, there exists infinitely many prime numbers p such that the Iwasawa λ-invariant is 0. This is closely related to Greenberg's conjecture.

2. In [Elk91], N. Elkies shows that the generalized ABC conjecture for number fields implies the Mordell conjecture.

Chapter 4

Public-key Cryptography

> "Nowadays, when a Number Theorist applies for a grant, he says that Number Theory is used in cryptography, and so doing Number Theory is good for National Security. Back then, since it was before the discovery of America, they said Number Theory is used in music. But I won't comment on the progress of civilization since then."
> – Hendrik Lenstra,
> `http://www.ucs.louisiana.edu/~avm1260/lenstra.html`

Public-key cryptography involves many of the same ideas and tools as the rest of computational number theory, and so in this book we will consider it on an equal footing. In particular, we will consider the RSA cryptosystem, whose security relies on the difficulty of factoring integers, and also elliptic curve-based cryptosystems such as ElGamal and elliptic curve Diffie-Hellman.

There are two intimately related sides to cryptography: creating methods to send encrypted messages, and cracking proposed encryption methods. It's extremely difficult to be good at creating methods for encrypting messages, but bad at attacking them, since one must be very aware of attack techniques in order to guard against them. It's probably less important to be good at making encryption systems if you want to be good at cracking them. It is natural that the typical cryptography group at a big company would put more effort into creating and implementing cryptosystems than into attacking them, since they have little commercial motivation for investing billions into attacking cryptosystems, and publicity about them doing so probably wouldn't look good (and indeed, such attack technology might have legal implications).

> "Former NSA technology boss Prescott Winter has a word for the kind of security he sees even at large, technologically sophisticated companies: Appalling"
> – `http://tinyurl.com/pgb45j8`, Slashdot, Oct 3, 2013.

For the cryptography aspect of this course, we'll consider the cryptosystems above, and some of the attacks on them, and how all of this work has a deep interplay with many other objects we consider.

4.1 The Standard protocols

To get started, let's quickly go over a couple of bog standard public-key cryptography protocols: Diffie-Hellman, RSA, and elliptic curve ElGamal.

4.1.1 Diffie-Hellman

The Diffie-Hellman key exchange protocol enables two programs A and B to agree on a common secret in full view of an adversary. For example, Diffie-Hellman (but on an elliptic curve) is used when you connect your web browser to the `https://cloud.sagemath.com` website in order to agree on a secret key that is used to encrypt all further data.

Here's how it works.

1. Programs A and B agree on a prime number p and a primitive root g modulo p, i.e., a number that generates $(\mathbb{Z}/p\mathbb{Z})^*$.

2. Program A compute a random number a and program B computes a random number b.

3. Program A sends $g^a \pmod{p}$ and program B sends $g^b \pmod{p}$.

4. Program A and program B compute the secret
$$s = (g^a)^b \pmod{p} = (g^b)^a \pmod{p}.$$

One subtlety is that there is a fast algorithm to compute $g^n \pmod{p}$ – to do this, write n in binary, and compute $g \cdots g$ (n times) by a couple of repeated squarings, reducing modulo p at each step... which is *easy*.

The main theoretical assumption one makes is that given $g^n \pmod{p}$ and $g \pmod{p}$ it is "difficult" to compute n, i.e., to find $\log_g(g^n \pmod{p})$. This is the *discrete log problem.* There are some primes p for which this assumption is obviously false. For example, if $p-1$ factors as a product of many small numbers, then solving the discrete log problem in $(\mathbb{Z}/p\mathbb{Z})^*$ is no more difficult than solving it in groups with these much smaller orders, hence easy. So, if when using Diffie-Hellman, it is critical to check that $p-1$ doesn't factor too much; the best you can hope for is that $p-1 = 2q$, with q prime, and this is what people often use.

See Section 4.5 below for a discussion of the complexity of finding a primitive root.

Once you commit to seriously using Diffie-Hellman modulo a prime, you have to worry about how big to make p. If you make p enormous, then everything takes more memory and is slower. For example, for large p establishing a connection to `https://cloud.sagemath.com` might take so much time (seconds!) that you'll just loose patience. To decide how big a p to take, you basically have to learn about algorithms for attacking the discrete log problem, look at what challenge problems people have solved in practice, then bet the farm on a particular choice size of p. In practice, people usually trust organizations such as NIST (which is advised by NSA, which knows a thing or two) to make the decisions for them.

Exercise 4.1.1. How big of primes p to people actually use?

There is an algorithm called "baby-step giant-step" which can solve the discrete logarithm mod p using time and space $O(\sqrt{p})$; we'll consider this algorithm in more detail later, since it is extremely important in computational number theory.

The Diffie-Hellman key exchange is amenable to the *man-in-the-middle attack*, where an adversary intercepts all transmissions, and agrees on two different secrets, one with program A and a *different secret* with program B. Then the adversary decrypts/encrypts all traffic between A and B. This is not theoretical – exactly this attack is used successfully in practice on the Internet (see recent news articles).

4.1.2 RSA

The RSA cryptosystem allows a program A to publish a public key for all to see. Any program B can then send A a message that is encrypted using this public key. Also, A can sign documents using their corresponding private key, and B can use A's public key to verify that indeed A signed the document. This RSA solves two problems: (1) being able to encrypt a message to A without any specific activity on A's part (unlike with Diffie-Hellman), and (2) being able to digitally sign a document. The main issue is that B has to trust that A's public key is... really A's public key, and in practice this is a major pain.

Here's how it works.

1. Program A chooses two prime numbers $p \neq q$ and computes $n = pq$, computes $\varphi(n) = (p-1)(q-1)$, and chooses an integer $e > 1$ that is coprime to $\varphi(n)$. Program A then publishes (n, e). Also, program A computes d such that $ed \equiv 1 \pmod{\varphi(n)}$.

2. Program B encrypts a message $m \pmod{n}$ to A by computing $m^e \pmod{n}$. (We assume the message m is coprime to n.)

3. Program A decryptes the message $m^e \pmod{n}$ by computing $m = (m^e)^d \pmod{n}$, where we use that $\varphi(n)$ is the order of $(\mathbb{Z}/n\mathbb{Z})^*$.

4. Also, if program A wishes to sign a message m, then program A publishes m and $m^d \pmod{n}$.

5. Program B can verify that program A signed m by computing $(m^d)^e \pmod{n}$ and checking that this equals m.

The security of RSA relies on the secret d, which is easy to compute if you know $\varphi(n)$. However, knowing $\varphi(n)$ requires factoring n, which seems—as far as we know—to be difficult in general. All the same remarks as for Diffie-Hellman apply above; in order to know how big to choose n, you need to be intimately familiar with approaches to factoring large integers n, or you look at what people have done and trust expert recommendations. As always, there is a tradeoff between speed and security.

Exercise 4.1.2. Install the gpg program ("gnu privacy guard"), generate an RSA key, and post the public part of it on your webpage, so that other people can send you secret messages. What bitsize did you choose?

4.2 Elliptic curve cryptography

For any prime power q and any elliptic curve

$$E: \quad y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}_q$, you can build various cryptosystems using the abelian group $E(\mathbb{F}_q)$. In 1985, Neal Koblitz and Victor Miller were the first to really push this idea in cryptography.

Elliptic curves are used *a lot* in cryptography these days. The primary motivation for using elliptic curves instead of $(\mathbb{Z}/n\mathbb{Z})^*$ or $\mathbb{F}_q^*$ is that... professionals tell us to. More precisely, based on guesswork and the ability of attackers, there's a belief that to get "sufficient security" using elliptic curves involves much smaller key sizes than getting the same security using $(\mathbb{Z}/n\mathbb{Z})^*$ or $\mathbb{F}_q^*$. These smaller key sizes result in faster protocols – e.g., connecting to a secure webpage is faster, or the license key you type in when installing software is much shorter.

4.2.1 Diffie-Hellman on an elliptic curve

Instead of choosing a primitive root g in $(\mathbb{Z}/p\mathbb{Z})^*$, choose a point $G \in E(\mathbb{F}_q)$. It's not critical that G generate the possibly non-cyclic group $E(\mathbb{F}_q)$. Instead, it's important that G have large prime order. The fact that it is possible to efficiently compute the order of G – even when q has hundreds of digits ! – is very deep and surprising, and is called the Schoof-Elkies-Atkin algorithm (or SEA for short). That this is possible is absolutely crucial to really making elliptic curve cryptography viable.

With G agreed upon, progams A and B publish aG and bG, for random secrets a and b, which can both be computed efficiently by writing a and b as sums of powers of 2 (i.e., in binary). Then the shared secret is

$$S = b(aG) = a(bG).$$

4.2.2 RSA on an elliptic curve?

Well, let's see... WARNING: I've never read anything about this: I just thought it through right now – William.

RSA relies on being able to construct and publish an algebraic group (in this case $\mathbb{G}_m(\mathbb{Z}/n\mathbb{Z})$) such that you know the order of the group, but it is difficult for other people to compute the order of that group. If the algebraic group is an elliptic curve over the field $\mathbb{F}_q$, then this is not possible, since the SEA algorithm ensures that anybody can efficiently compute the order of $E(\mathbb{F}_q)$. A direct analogue of RSA on $E(\mathbb{F}_q)$ would have public key $(E/\mathbb{F}_q, e)$, where e is some random number modulo the exponent of the group $E(\mathbb{F}_q)$. One would encode a message as a point $M \in E(\mathbb{F}_q)$ then encrypt it as $eM \in E(\mathbb{F}_q)$.

To encode a message as a point on the curve, you of course first break the message up into smaller pieces, add salt (randomize it somehow), then get a possible x coordinate. By solving a quadratic equation in $\mathbb{F}_q$, which can be done efficiently in practice, one constructs a point (x, y), or if there are no solutions, simply slightly modify x to get a solution (since half of the x's work). Solving a quadratic equation involves extracting a square root, which can be done efficiently in practice (see [[reference to something]]).

However, there's a major problem with this protocol, which involves getting confused about what the discrete log problem is. If you know eM and e, you can compute M by multiplying eM by the multiplicative inverse of e modulo the order of M, and this order is easy to compute since we know $\#E(\mathbb{F}_q)$. The discrete log problem in $E(\mathbb{F}_q)$ is a different problem:

Discrete Log Problem: *Given M and eM, what is e?*

Another analogue of RSA would be to generalize the notion of elliptic curve to obtain an object E defined over $\mathbb{Z}/n\mathbb{Z}$, where $n = pq$ is a product of distinct primes. Using the theory of *group schemes* one can put such a more general definition on a rigorous footing. The isomorphism $\mathbb{Z}/n\mathbb{Z}$
$isom\mathbb{Z}/p\mathbb{Z} \oplus \mathbb{Z}/q\mathbb{Z}$ induces a canonical isomorphism

$$E(\mathbb{Z}/n\mathbb{Z}) \cong E(\mathbb{Z}/p\mathbb{Z}) \oplus E(\mathbb{Z}/q\mathbb{Z}), \tag{4.2.1}$$

but (presumably it is easy to prove that) one must factor n in order to compute this isomorphism.

Exercise 4.2.1. Prove (if easy/true!?) that if you know the isomorphism (4.2.1), then you can factor n.

One then encrypts a message $M \in E(\mathbb{Z}/n\mathbb{Z})$ as eM, which one can compute. But wait, given a message M, how do you encode it as a point on the curve? To do so, you need to be able to extract square roots in the ring $\mathbb{Z}/n\mathbb{Z}$. I can't think of any way to do this without factoring n. Of course, if you choose E to have an obvious point on it, then you can do lots of arithmetic to find many other points on E. However, they are all basically random—you can only encrypt nonsense.

Incidentally, a critically important approach to factoring an integer n is to consider the group $E(\mathbb{Z}/n\mathbb{Z})$ for a certain choice of curve with an easy-to-find point on it, and to do arithmetic there, as if n were prime. If something goes wrong, one discovers a factorization of n. This is H. Lenstra's Elliptic Curve factorization Method (ECM). It turns out it is very good at splitting off medium sized factors of an integer. Given a large integer n, the ECM algorithm finds primes $p \mid n$ that are bigger than trial division finds, but usually much smaller than $\sqrt{n}$. I've heard that this algorithm is crucial as an intermediate step in "number field sieve" attacks on RSA itself.

4.2.3 ElGamal on an elliptic curve

The ElGamal cryptosystem makes sense in any cyclic abelian group, and provides a way to solve some of the problems that RSA addresses.

Here's how it works.

1. Program A chooses an elliptic curve E over $\mathbb{F}_q$, a point $P \in E(\mathbb{F}_q)$ of large order, and computes a secret random number a. Program A then publishes the public key $(E/\mathbb{F}_q, P, aP)$. The private key is a.
2. Program B encrypts a message $M \in E(\mathbb{F}_q)$ as follows. Program B computes a random secret integer r and encrypts M as the pair $(rP, M+raP)$.
3. Program A decrypts the message by computing $M = M + raP - arP$, which program A can compute because it knows the secret number a.

Conceptually, ElGamal is just an obvious extension of Diffie-Hellman. Program A does its part of Diffie-Hellman and "puts it out there" as a public key. Then, program B does its part of Diffie-Hellman to construct a shared secret – namely raP – and just uses that shared secret as a one-time pad to encrypt the message.

With RSA there is perfect symmetry between the public and private keys, but with ElGamal the public key is two points on an elliptic curve and the private key is a number. Thus ElGamal does not naturally also produce a way of constructing digitial signatures.

4.3 The Elliptic curve digital signature algorithm

The Elliptic Curve Digital Signature Algorithm (ECDSA) is a widely used algorithm that solves—using elliptic curves—another problem that RSA solves, namely digitally signing messages. It is, for example, used as part of every bitcoin transaction. As we saw above, even with ElGamal, it is not at all obvious how to solve this problem! ECSDA was invented by Scott Vanstone in 1992. (Speculation: The internet suggests that Vanstone maybe (?) invented ECDSA while working at the NSA, then started a company called Certicom, patented some elliptic curves then licensed the patents back to the NSA for $25 million dollars? And there are aliens 'n stuff. Anyways, I'm sure Koblitz can clear this up.)

4.3.1 The ECDSA protocol

Whenever we choose a random element "of $\mathbb{F}_p$" below, we choose it to be neither 0 or 1.

Protocol: *Program A digitally signs a message m using ECDSA as follows:*

1. [Setup Protocol] Choose a prime power q, an elliptic curve $E/\mathbb{F}_q$, and a point $G \in E(\mathbb{F}_q)$ of prime order p (which has nothing to do with q), and define a set-theoretic map (in any way) $\phi : \mathbb{F}_q \to \mathbb{F}_p^*$. Choose a random secret number $d \in \mathbb{F}_p^*$, and let $Q = dG$. The public key is (E, G, Q, p) and the private key is d.

2. [Hash] Hash the message m to an element $z \in \mathbb{F}_p^*$.

3. [Random Point] Choose a random $k \in \mathbb{F}_p^*$, and compute $kG \in E(\mathbb{F}_q)$.

4. [Compute Signature] Compute

$$r = \phi(x(kG)) \in \mathbb{F}_p^* \text{ and } s = (z + rd)/k \in \mathbb{F}_p$$

In the unlikely case $s = 0$, choose a new random point in step 3; otherwise, the signature of the message m is the pair (r, s) of elements of $\mathbb{F}_p^*$.

Protocol: *Program B verifies a digital signature* (r, s) *of a message* m *using ECDSA as follows:*

1. [Hash] Hash m to the same $z \in \mathbb{F}_p^*$ as above.

2. [Verify] The signature is valid if $r = \phi(x(C))$, where

$$C = \frac{z}{s}G + \frac{r}{s}Q \in E(\mathbb{F}_q).$$

Here z/s and r/s are the quotients in the field $\mathbb{F}_p$, and we view $\langle G \rangle \subset E(\mathbb{F}_q)$ as a one-dimensional F_p-vector space.

Proposition 4.3.1. *If* (r, s) *is a valid signature, then Program B will conclude that it is a valid signature.*

Proof. Suppose that (r, s) is a valid signature, so

$$s = \frac{z + rd}{k} \in \mathbb{F}_p^*.$$

Then $k = \frac{z+rd}{s}$, so

$$kG = \frac{z}{s}G + \frac{rd}{s}G = \frac{z}{s}G + \frac{r}{s}Q = C,$$

hence $\phi(x(C)) = \phi(x(kG)) = s$.

□

4.3.2 Playstation 3 hacked!

In 2010, Sony implemented ECDSA for the Playstation 3, in order to control what PS3 owners could run on their game consoles. Evidently, they crossed some line, and specifically made it so people could no longer run Linux (and pirate games) on their PS3. So some people decided to attempt to crack their implementatio of ECDSA. Studying many examples, the noticed a pattern. It turned out that instead of choosing the parameter k at random in Step 3 of the signature protocol, they always used the same value of k! Oops.

Why is this a problem? Suppose we have in hand two valid signatures (r, s) and (r', s') with corresponding hashed messages z and z'. Recall that the private key is some $d \in \mathbb{F}_p^*$, so our goal is to find d. As in the proof of Proposition 4.3.1, we have

$$\frac{z + rd}{s} = k = k' = \frac{z' + r'd}{s'} \tag{4.3.1}$$

Oops, that's one linear equation in one unknown, namely the private key d. So with two signatures, one can solve for d, get the private key, and it's game over.

More generally, when implementing ECDSA, it is extremely important to choose k sufficiently randomly. If there is something sufficiently predictable about k – even a few bits – then one can get information toward (4.3.1), and with sufficiently many signatures, recover d.

4.4 The Bitcoin elliptic curve

Bitcoin appears to be the first successful digital currency in history. It's revolutionary, in that it is not supported or controlled by governments, banks, or anything else real or political.[1]. There are many aspects of how the system works (and how bitcoins are mined), but there is one critical part that involves computational number theory, and that's what we'll discuss. In order to spend bitcoin, a certain message is digitally signed, which involves computationa of a hash followed by application of the ECDSDA algorithm. Thus for Bitcoin, trust in governments, banks, armies, credit ratings, etc., is replaced trust in this elliptic curve

$$y^2 = x^3 + 7 \quad \text{over} \quad \mathbb{F}_p, \text{ where } p = 2^{256} - 2^{32} - 2^9 - 2^8 - 2^7 - 2^6 - 2^4 - 1.$$

This curve has cardinality the 256-bit prime

$$\begin{aligned}\#E(\mathbb{F}_p) = &115792089237316195423570985008687907852837564279074\\ &904382605163141518161494337\end{aligned}$$

[1] Many mainstream institutions are threatened by bitcoin: `http://www.coindesk.com/capital-one-closes-bank-account-bitcoin/`. Bitcoin also makes it very difficult for powerful governments to steal all the money of people who do things they don't like: `http://www.extremetech.com/computing/168139-fbi-unable-to-seize-600000-bitcoins-from-silk-road-operator`

The fixed base point G on the curve is

$$\begin{aligned} G =(&55066263022277343669578718895168534326250603453777594175500187360389116729240,\\ &32670510020758816978083085130507043184471273380659243275938904335757337482424) \end{aligned}$$

which obviously generates $E(\mathbb{F}_p)$, since it generates a subgroup of a group of prime order.

The curve E and point G above have the official name **secp256k1**, where the 256 refers to the number of bits of p, and the k is the first letter of Neal Koblitz's last name. This curve is somehow special in that there is a way to more cleverly compute kG, which results in very real speedups in practice, as compared to computing kG on a "generic" 256-bit elliptic curve. There's an extensive discussion online about actual implementations, where they claim a 30% speedup in practice, while also expressing serious reservations about using a special curve as the foundation for such an important cryptosystem – finding a fast way to solve the discrete log problem on this one particular curve would have a major impact on a billion dollar economic system.

Koblitz remarks: "[This speedup is described in] Jerry Solinas' Crypto '97 talk, and is very concrete and understandable. It boils down to writing number in the base rho, where $\rho = (-1+\sqrt{-7})/2$. That talk was the first Crypto talk by an NSA person."

4.4.1 Bitcoin hacked!

We saw above when discussing the PS3 crack that it is critical that every single time program A signs a message using ECDSA, that a *different* random value k is used. If any k is ever re-used, the private key can then be easily found. In August 2013, exactly this problem surfaced in many of the bitcoin wallet implementations for Android, and it was actually exploited:

> "It looks as though, at least on occasion, the Java-based PRNG on Android will repeat its pseudorandom sequences, thanks to a flaw in Android's so-called SecureRandom Java class. The Bitcoin Forum has already reported the theft of close to BTC56 (worth about US $6000) from a number of people. A list of known-vulnerable Android Bitcoin wallets has been published by the Bitcoin Project, with instructions on what to do when the various wallet apps are fixed to use better-quality random numbers."

`http://nakedsecurity.sophos.com/2013/08/12/android-random-number-flaw-implicated-in-bitcoin-`

The fix for users was to create a new bitcoin public/private key pair using a more secure approach, and transfer all of their money to themselves.

4.5 Finding a primitive root mod p

In this section, a primitive root mod p will refer to a generator of the cyclic group $(\mathbb{Z}/p\mathbb{Z})^\times$. First, let us consider an easy example. Let us assume that the prime p satisfies the relation $p-1=2\cdot q$, where q is also a prime number. Now, the order of any element in $(\mathbb{Z}/p\mathbb{Z})^\times$ is either 1, 2, q or $2\cdot q$. Now, -1 is not a square in $(\mathbb{Z}/p\mathbb{Z})^\times$, since 4 does not divide $p-1$. Hence, either 2 or -2 (and exactly one of them) will be a generator for the cyclic group $(\mathbb{Z}/p\mathbb{Z})^\times$. So, it suffices to computee 2^q and check if it equals 1 or -1. A naive algorithm for finding a primitive root would be systematically computing $m^{(p-1)/2}$, for every element m in $(\mathbb{Z}/p\mathbb{Z})^\times$. If an element m satisfies the equality $m^{(p-1)/2} = -1($ mod $p)$, then the element m must be a primitive root.

There are polynomial time algorithms which assume the validity of the generalized Riemann Hypothesis. For instance, in [Sho92], the algorithm to find a primitive root is of order $O(\log^6(p))$. There are some polynomial time algorithms for computing primitve roots. For instance, see [DD06].

There are $\phi(p-1)$ generators in $(\mathbb{Z}/pZ)^\times$, where $\phi(x)$ is the Euler-toitent function. If one chooses an element from $(\mathbb{Z}/p\mathbb{Z})^\times$ at random, the probability that this element is a generator is $\phi(p-1)/(p-1)$. Artin has conjetured that that a given integer n which is not equal to 1, -1 and which is not a perfect square is a primitive root for infinitely many primes p. The conjecture is still open, though some progress has been made in [GM84] and [HB86]. Artin's conjecture gives us some hope that sampling at random will lead to a primitive root. By the prime number theorem, $\phi(p-1)/(p-1) \sim 1/\log(p)$. In practice, by selecting an element at random, one expects to find a primitive root after $(p-1)/\phi(p-1)$ tries, which is of order $O(\log(p))$ (which is significantly smaller than $O(\log^6(p))$).

Chapter 5

Enumerating Elliptic Curves

The main goal in this chapter is to understand the following deep recent theorem:

Theorem 5.0.1. *There is an algorithm that takes as input a positive integer N and outputs all elliptic curves defined over $\mathbb{Q}$ of conductor N.*

Along the way, we'll learn many interesting theorems related to elliptic curves and about how to compute modular forms using modular symbols.

Definition 5.0.2 (Elliptic Curve)**.** An elliptic curve over a field K is a nonsingular (geometrically irreducible) genus one curve E defined over K with a distinguished rational point $\mathcal{O} \in E(K)$.

5.1 The Discriminant

Let E be an elliptic curve defined over $\mathbb{Q}$. Using the Riemann-Roch theorem and that E is a genus one curve with a rational point, we can prove that there is an (affine) Weierstrass equation of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6. \tag{5.1.1}$$

The "x-coordinate of a point" is a rational function $x : E \to \mathbb{P}^1$ that has degree 2, and the y-coordinate is a rational function of degree 3. The subscripts in the above equation are chosen so that the degree of each term is 6. The *discriminant* of (5.1.1) is a polynomial function the coefficients of the curve:

$$\Delta = \frac{c_4^3 - c_6^2}{1728} = -8\,b_4^3 + 9\,b_2b_4b_6 - b_2^2b_8 - 27\,b_6^2, \tag{5.1.2}$$

where

$$c_4 = b_2^2 - 24\,b_4, \quad c_6 = -b_2^3 + 36\,b_2b_4 - 216\,b_6,$$

and

$$\begin{aligned}&b_2 = a_1^2 + 4\,a_2, \quad b_4 = a_1 a_3 + 2\,a_4, \quad b_6 = a_3^2 + 4\,a_6,\\ &b_8 = a_2 a_3^2 - a_1 a_3 a_4 + a_1^2 a_6 - a_4^2 + 4\,a_2 a_6.\end{aligned}$$

NOTE: I literally just copied these formulas from the source code of Sage... and 7 years ago, I implemented them in Sage based on Silverman's book (which got c_6 wrong!).

```
sage: E = EllipticCurve([1,2])
sage: E.discriminant??
sage: E.b_invariants??
```

We have $\Delta = 0$ if and only if the curve is nonsingular, which is always the case for an elliptic curve, by definition.

Definition 5.1.1 (Minimal Discriminant)**.** The minimal discriminant of E is the smallest (in absolute value) discriminant of any Weierstrass equation (5.1.1) for E.

Theorem 5.1.2. *Fix a discriminant $0 \neq \Delta \in \mathbb{Z}$. Then there are only finitely many isomorphism classes of elliptic curves with discriminant Δ.*

As we will see, Theorem 5.1.2 is implied by a finiteness theorem about integral points. It's also implied by modularity of elliptic curves. And, it's a special case of Faltings' finiteness theorem on abelian varieties. Also, making it explicit has connections with the ABC conjecture.

5.1.1 Discriminant 1

Suppose E is an elliptic curve over $\mathbb{Q}$ with discriminant 1. Then there is a Weierstrass equation as in (5.1.1) with integer coeffecients a_1, a_2, a_3, a_4, a_6 such that $\Delta = 1$. There are also the corresponding b-invariants b_2, b_4, b_6, b_8; and c-invariants c_4 and c_6. Now if $\Delta = 1$ then c_4 and c_6 satsify

$$c_4^3 - c_6^2 = 1728$$

which looks an awful lot like the equation for an elliptic curve. Set $x = c_4$, $y = c_6$, and rearrange to get

$$y^2 = x^3 - 1728.$$

The integral points on this curve can be found by the Sage code

```
sage: E = EllipticCurve([0,0,0,0,-1728])
sage: E.integral_points()
```

which outputs

```
[(12 : 0 : 1)]
```

so the only possibilities are $c_4 = 12$ and $c_6 = 0$. Using these values, we can solve for b_4 and then b_6 in terms of b_2 using the equations for c_4 and c_6. This comes out to $24b_4 = b_2^2 - 12$ and $432b_6 = b_2^3 - 36b_2$. Now we can use these equations to solve for b_8 in terms of b_2

$$6912b_8 = b_2^4 - 72b_2^2 - 432.$$

This equation modulo 6912 shows b_2 is a solution of $x^4 - 72x^2 - 432 \equiv 0 \mod 6912$. However this equation has no solutions, as can be verified by Sage:

```
sage:     for x in range(6912):
sage:         if mod(x^4 - 72*x - 432,6912) == 0:
sage:             print(x)
```

which does not output anything. Therefore no such b_2 exists so the curve could not have had discriminant 1. However, in general there are lots of Weirstrass equations with discriminant 1 as long as we allow coefficients in $\mathbb{Q}$. For example the curve $y^2 - 4x^3 - x^2 + \frac{11}{12}x + \frac{35}{432}$ has discriminant 1.

5.2 Finding curves with bounded discriminant

Motivating Problem: Given a positive integer M, find a representative for each isomorphism class of elliptic curve over $\mathbb{Q}$ with $|\Delta| \leq M$.

According to (5.1.2), we have

$$1728\Delta = c_4^3 - c_6^2, \tag{5.2.1}$$

where Δ, c_4, c_6 are all integers.

Theorem 5.2.1. *The ABC conjecture implies that there are only finitely many pairs* (c_4, c_6) *that satisfy* (5.2.1) *with* $|\Delta| \leq M$.

Proof. First assume that Δ is positive (in which case c_4 must also be positive). We have an equation

$$A + B = C, \tag{5.2.2}$$

where $A = 1728\Delta$, $B = c_6^2$, $C = c_4^3$ are all positive integers.

To each triple with $\gcd(A, B, C) = 1$ with $A < 1728M$, there are only finitely many other triples (dA, dB, dC) with $dA < 1728M$, so we may restrict to the case that $\gcd(A,B,C)=1$. The quality of the coprime triple (5.2.2) is

$$\begin{aligned}
q(A,B,C) &= \frac{\log(C)}{\log(r(ABC))} \\
&= \frac{\log(c_4^3)}{\log(r(1728\Delta c_6^2 c_4^3))} \\
&= \frac{3\log(c_4)}{\log(r(6\Delta|c_6|c_4))} \\
&= \frac{3\log(c_4)}{\log(r(6\Delta))+\log(r(|c_6|))+\log(r(c_4))} \\
&\geq \frac{3\log(c_4)}{\log(6\Delta)+\log(|c_6|)+\log(c_4)} \\
&> \frac{3\log(c_4)}{\log(6M)+\log(c_4^{3/2})+\log(c_4)} \\
&= \frac{3\log(c_4)}{\log(6M)+\frac{5}{2}\log(c_4)} \to 1.2 \text{ as } c_4 \to \infty
\end{aligned}$$

where in the inequalities we make the denominator as big as possible, first assuming each summand is square free, then using the $c_6^2 < c_4^3$. The (strong) ABC conjecture implies that there are finitely many coprime triples (A,B,C) with quality bigger than 1.2, so we conclude that there are only finitely many c_4 (and hence c_6) as above.

Now suppose $\Delta < 0$ and $c_4 > 0$. Then, we have

$$c_4^3 + (-1728\Delta) = c_6^2$$

Set $A = c_4^3$, $B = -1728\Delta$, and $C = c_6^2$. As before, we may assume that $\gcd(A,B,C)=1$. The quality of this coprime triple is

$$\begin{aligned}
q(A,B,C) &= \frac{\log(C)}{\log(r(ABC))} \\
&= \frac{\log(c_6^2)}{\log(r(-1728\Delta c_6^2 c_4^3))} \\
&= \frac{2\log(|c_6|)}{\log(r(6|\Delta||c_6|c_4))} \\
&= \frac{2\log(|c_6|)}{\log(r(6|\Delta|))+\log(r(|c_6|))+\log(r(c_4))} \\
&\geq \frac{2\log(|c_6|)}{\log(6|\Delta|)+\log(|c_6|)+\log(c_4)} \\
&> \frac{2\log(|c_6|)}{\log(6M)+\log(|c_6|)+\log(|c_6|^{2/3})} \\
&= \frac{2\log(|c_6|)}{\log(6M)+\frac{5}{3}\log(|c_6|)} \to 1.2 \text{ as } |c_6| \to \infty
\end{aligned}$$

and we see once again that the (strong) ABC conjecture implies only finitely many such triples exist.

For the final case, suppose that $\Delta < 0$ and $c_4 \leq 0$. Then we have

$$-c_4^3 + c_6^2 = -1728\Delta$$

One quickly verifies that an analagous argument to those above will not work in this case. However, this is not a problem as an even simpler argument – independent of the ABC conjecture – is available in this case: Using our bound on the discriminant gives

$$-c_4^3 + c_6^2 \leq 1728M$$

and there are only finitely many pairs of nonnegative integers whose sum is less than or equal to $1728M$. □

By iterating through curves with bounded discrimant we can actually see the bounds proved in 5.2.1. Figure 5.2.1 shows the quality of an elliptic curve as a function of c_6. It was made by iterating over curves with $|\Delta| < 10^{10}$, $\Delta < 0$, $c_4 > 0$, and $\gcd(c_4^3, -1728\Delta) = 1$. This corresponds to the second case in the proof above. Each dot represents an elliptic curve satisfying the requirements with height given by the quality as defined in the proof, $q(c_4^3, -1728\Delta, c_6^2)$. The line represents the llbound $\frac{2\log(|c_6|}{\log(6M)+\frac{5}{3}\log(|c_6|)}$.

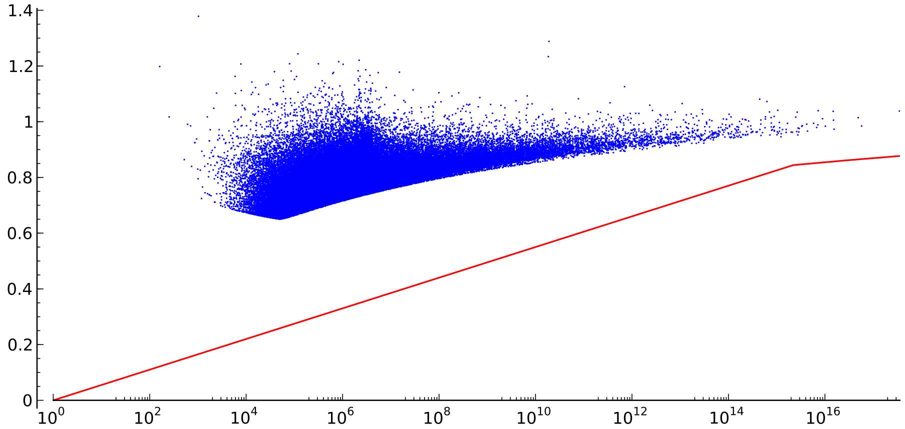


Figure 5.2.1: Quality as a function of c_6

Of course, due to Andrew Ohana's observations, I have failed to convince you that the strong ABC conjecture is actually true (see Section 3.3.2), so I will understand if you do not consider this argument to be convincing!

One goal of proving Theorem 5.2.1 was to hopefully give some explicit upper bound on the number of curves with discriminant $|\Delta| < M$, assuming some effective version of ABC, which might come out of the computations we've been doing. Of course, this is not what we get from our data...

See [BMSW07, §3.4] for the following conjecture:

Conjecture 5.2.2. *There are $cX^{5/6}$ curves with $|\Delta| \leq X$.*

`http://mathoverflow.net/questions/96285/average-rank-of-elliptic-curves-over-mathbb`

The paper [SW02] is about systematic enumeration of all pairs of integers (c_4, c_6) such that we have simultaneously

$$|\Delta| \leq 10^{12} \quad \text{and} \quad |c_4| \leq 1.44 \cdot 10^{12}.$$

We do not know that this gives us all curves with $|\Delta| \leq 10^{12}$, but it does give us hundreds of millions of elliptic curves. Here's a quick demo of using the Stein-Watkins tables from Sage (see `http://wstein.org/ecdb/format.txt` for the meaning of the data).

```
sage: v = SteinWatkinsPrimeData(0)
sage: c = v.next(); c
Stein-Watkins isogeny class of conductor 11
sage: list(c)
[[[0, -1, 1, 0, 0], ’(1)’, ’1’, ’5’],
 [[0, -1, 1, -10, -20], ’(5)’, ’1’, ’5’],
 [[0, -1, 1, -7820, -263580], ’(1)’, ’1’, ’1’]]
sage: v = SteinWatkinsPrimeData(1)
sage: c = v.next(); c
Stein-Watkins isogeny class of conductor 100000937
sage: list(c)
[[[1, 0, 1, -472, -3951], ’[1]’, ’1’, ’1’]]
sage: v = SteinWatkinsAllData(0)
sage: c = v.next(); c
Stein-Watkins isogeny class of conductor 11
sage: list(c)
[[[0, -1, 1, 0, 0], ’(1)’, ’1’, ’5’],
 [[0, -1, 1, -10, -20], ’(5)’, ’1’, ’5’],
 [[0, -1, 1, -7820, -263580], ’(1)’, ’1’, ’1’]]
sage: c = v.next(); c
Stein-Watkins isogeny class of conductor 14
sage: list(c)
[[[1, 0, 1, -1, 0], ’(2,1)’, ’1’, ’6’],
 [[1, 0, 1, -11, 12], ’[1,2]’, ’1’, ’6’],
 [[1, 0, 1, 4, -6], ’(6,3)’, ’1’, ’6’],
 [[1, 0, 1, -36, -70], ’[3,6]’, ’1’, ’6’],
 [[1, 0, 1, -171, -874], ’(18,1)’, ’1’, ’2’],
 [[1, 0, 1, -2731, -55146], ’[9,2]’, ’1’, ’2’]]
sage: v = SteinWatkinsAllData(10)
sage: c = v.next(); c
```

```
Stein-Watkins isogeny class of conductor 1000002
sage: list(c)
[[[1, 1, 0, -63, -1539], '(3,6,1)', 'X', '1']]
```

5.2.1 Number fields of bounded discriminant

One can also define the discriminant of a number field and one obtains a similar finiteness theorem:

Theorem 5.2.3 (Hermite's Theorem)**.** *For each positive integer N, there exist only finitely many number fields with discriminant $|\Delta| \leq N$.*

Proof. See section III.2 of (Neukirch) □

John Jones has a searchable database of number fields of small degree at `http://hobbes.la.asu.edu/NFDB/`

Let us use this database and Hermite's Theorem to study a certain class of fields. Let K be a number field and let p be an odd prime. A field K_∞ is said to be a $\mathbb{Z}_p$-extension of K if K_∞/K is Galois and $\mathrm{Gal}(K_\infty/K) \approx \mathbb{Z}_p$. We will need a few facts about $\mathbb{Z}_p$-extensions. The results that follow can be found in chapter 13 of [Was97]. For each $n \geq 0$, there is a unique intemediate field $K \subseteq K_n \subseteq K_\infty$ such that $\mathrm{Gal}(K_n/K) \approx \mathbb{Z}/p^n\mathbb{Z}$. The only primes dividing the discriminant of K_n are those primes dividing the discriminant of K and possibly (always, if n is large enough) the prime p. Suppose now that $K = \mathbb{Q}(\sqrt{-d})$ is an imaginary quadratic field. Using class field theory, one can show that there are exactly two $\mathbb{Z}_p$ extensions K_∞^+ and K_∞^- which are Galois over $\mathbb{Q}$, called the *cyclotomic* extenstion and *anticyclotomic* extension, respectively. The Galois group of K_n^+ over $\mathbb{Q}$ is the direct product:

$$\mathrm{Gal}(K_n^+/\mathbb{Q}) \cong \mathrm{Gal}(K_n^+/K) \times \mathrm{Gal}(K/\mathbb{Q})$$

while the Galois group of K_n^- over $\mathbb{Q}$ is the semidirect product

$$\mathrm{Gal}(K_n^-/\mathbb{Q}) \cong \mathrm{Gal}(K_n^-/K) \rtimes \mathrm{Gal}(K/\mathbb{Q})$$

Exercise 5.2.4. Constructing $\mathbb{Z}_p$ extensions.

1. Using John Jones' database, construct K_1^+ for various choices of $K = \mathbb{Q}(\sqrt{-d})$ and (small) p.

2. Let μ_m denote the group of m-th roots of unity, and let μ_{p^∞} denote the group of all p-th power roots of unity. It can be shown that K_∞^+ is a subfield of $K(\mu_{p^\infty})$. For the fields K_1^+ that you found, verify that $K_1^+ \subseteq K(\mu_{p^2})$. When is $K_1^+ \subseteq K(\mu_p)$?

3. Now construct K_1^- for various choices of d and p.

4.

5.3 The Conductor

The conductor of an elliptic curve (defined over the rational numbers or a number field) is another invariant, like the minimal discrimant, which encodes information about the local structure of the elliptic curve at each bad prime.

Theorem 5.3.1. *For each positive integer N, there are finitely many elliptic curves over $\mathbb{Q}$ of conductor N, and there is an algorithm to enumerate all of them that is polynomial time in N.*

5.3.1 Reduction

Suppose E is an elliptic curve over $\mathbb{Q}$, and let

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6 \qquad (5.3.1)$$

be a global minimal model, so the a_i are in $\mathbb{Z}$ and the discriminant Δ of the above equation is minimal in absolute value, among all such Weierstrass equations for E. For each prime number p, it makes sense to consider the algebraic curve $E_{\mathbb{F}_p}$ over the finite field $\mathbb{F}_p$ obtained by reducing (5.3.1) modulo p. The curve $E_{\mathbb{F}_p}$ is nonsingular if and only $p \nmid \Delta$. When $p \mid \Delta$, there are three possible ways in which $E_{\mathbb{F}_p}$ can be singular:

1. split nodal, e.g., $y^2 = x^2(x+9) \pmod{11}$: there are two distinct tangent lines at the singular point, with slopes in $\mathbb{F}_p$.
2. nonsplit nodal, e.g., $y^2 = x^2(x+23) \pmod{37}$: there are two distinct tangent lines at the singular point, but the slopes are in $\mathbb{F}_{p^2}$ instead of $\mathbb{F}_p$.
3. cuspidal, e.g., $y^2 = x^3 \mod 17$: there is a cuspidal singularity.

We call the first two cases *multiplicative reduction*, since in those cases the group of nonsingular points in $E_{\mathbb{F}_p}$ is $\mathbb{G}_m$ in the split case, or it's the nontrivial twist of $\mathbb{G}_m$ in the nonsplit case. We call the nodal case *additive reduction* since the group of nonsingular points is $\mathbb{G}_a$. (To see these isomorphisms, draw a line through the singular point and consider the other point of intersection; this sets up a bijection between lines through the origin and nonsingular points.)

5.3.2 "Definition" via an incomplete formula that you will remember

The conductor N of E is

$$N = \prod_{p\mid\Delta} p^{v_p},$$

where

$$v_p = \begin{cases} 1 & \text{if } E \text{ has multiplicative reduction at } p \\ 2 & \text{if } E \text{ has additive reduction at } p \geq 5 \\ \leq 5 & \text{if } E \text{ has additive reduction at } p = 3 \\ \leq 8 & \text{if } E \text{ has additive reduction at } p = 2\ . \end{cases}.$$

In each additive case, $v_p \geq 2$. The above is what everybody easily remembers about the conductor of an elliptic curve. I don't know of any simple recipe for v_p when $p = 2, 3$ is a prime of additive reduction is complicated.

Remark 5.3.2. See Section 5.3.7 below for a generalization (with examples) of the above formula to number fields, where you will find out whether $v_{\mathfrak{p}} \leq 2$ for $\mathrm{char}(\mathfrak{p}) \geq 5$ and if $v_{\mathfrak{p}}$ can be arbitrarily large.

The (partial) definition above is pretty arbitrary, lacking any conceptual insight. How does it generalize to elliptic curves over numbers fields? What does it have to do with other things called "conductors" and "levels" in number theory?

> **Summary:** The conductor of E is a product of the primes of bad reduction for E. It is exactly divisible by the primes of multiplicative reduction, and divisible by p^2 for primes of additive reduction.

5.3.3 Definition using the functional equation

Consider the L-series

$$L(E,s) = \prod_{\text{all } p} \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{1-2s}},$$

where $\varepsilon(p) = 0$ if $p \mid \Delta$ and $\varepsilon(p) = 1$ otherwise, and

$$a_p = p + 1 - \#E_{\mathbb{F}_p}(\mathbb{F}_p).$$

(NOTE: These a_p have nothing to do with the coefficients a_i in the Weierstrass equation (5.3.1) above.) The deep theorem that every elliptic curve E over $\mathbb{Q}$ is modular implies that $L(E,s)$ extends (uniquely) to a holomorphic function on the entire complex plane. It's easier to understand the completed L-series

$$\Lambda(E,s) = 2\pi^{-s}\Gamma(s)L(E,s).$$

The modularity theorem also implies that there is a unique integer N and sign $\epsilon \in \{1, -1\}$ such that $\Lambda(E,s)$ satisfies the following functional equation:

$$\Lambda(E,s) = \epsilon \cdot N^{1-s}\Lambda(E, 2-s). \tag{5.3.2}$$

The number N is the *conductor of* E.

> **Summary:** The conductor of E is a constant that appears naturally in the functional equation for the L-series of E.

5.3.4 (*) Definition using modular forms

This section gets a star because it is how I think about the conductor.

Expand the Euler product for $L(E,s)$ from Section 5.3.3 to obtain a Dirichlet Series

$$L(E,s) = \prod_p L_p(E,s) = \sum_{n\geq 1} \frac{a_n}{n^s},$$

where the a_n are integers. The analytic function $L(E,s)$ on the complex plane is connected via Mellin transform to the holomorphic function

$$f_E(z) = \sum_{n\geq 1} a_n e^{2\pi i n z}$$

on the upper half plane. Alternatively, writing $q = q(z) = e^{2\pi i z}$, we write

$$f_E(q) = \sum_{n\geq 1} a_n q^n,$$

which we view either as function of z on the upper half plane, or as a function of q on the open unit disk.

The *conductor of* E is the smallest positive integer N such that for all matrices

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) = \{\text{matrices in } \mathrm{SL}_2(\mathbb{Z}) \text{ that are upper triangular mod } N\},$$

we have

$$f(\gamma(z))d(\gamma(z)) = f(z)dz,$$

where $\gamma(z) = \frac{az+b}{cz+d}$ is a linear fractional transformation. Thus f is a cuspidal modular form of level the conductor N of E.

> **Summary:** The conductor of E is the level of the modular form attached to E.

5.3.5 Definition using Néron models and Ogg's formula

The *Néron model* $\mathcal{E}$ of E is the unique smooth commutative group scheme over $\mathbb{Z}$ with generic fiber $\mathcal{E}_{\mathbb{Q}} = E$ such that for every smooth scheme $X/\mathbb{Z}$ the natural map

$$\mathcal{E}(X) \to E(X_{\mathbb{Q}})$$

is an isomorphism. The statement that the above map is an isomorphism means that any morphism $X_{\mathbb{Q}} \to E$ on generic fibers can be extended uniquely to a morphism $x \to \mathcal{E}$ of schemes over $\mathbb{Z}$. It is a theorem of Néron that his model exists.

The *component group* of E at p is the quotient

$$\Phi_{E,p} = \mathcal{E}_{\mathbb{F}_p}/\mathcal{E}^0_{\mathbb{F}_p}$$

of the reduction of $\mathcal{E}$ modulo p by its identity component. We have an exact sequence

$$0 \to \mathcal{E}^0_{\mathbb{F}_p} \to \mathcal{E}_{\mathbb{F}_p} \to \Phi_{E,p} \to 0,$$

and $\Phi_{E,p}$ is a finite flat group scheme over $\mathbb{F}_p$.

The exponent $v_p = \mathrm{ord}_p(N)$ in the conductor $N = \prod p^{v_p}$ is

$$v_p = \mathrm{ord}_p(\Delta) - (\#\Phi_{E,p}(\overline{\mathbb{F}}_p) - 1) \qquad \text{(Ogg's formula)},$$

where Δ is the minimal discriminant of E.

Software such as Sage uses Tate's algorithm to compute $\#\Phi_{E,p}(\overline{\mathbb{F}}_p)$. This is implemented in Sage over arbitrary number fields.

Ogg's formula implies that given an elliptic curve E, knowing any two of the following invariants determines the third:

1. The minimal discriminant of E.

2. The conductor of E.

3. The orders of the components groups of E.

For example, Aly Deines uses this in her thesis work (extending work of Shuzo Takahashi and Ken Ribet), where she uses Shimura curves to determine invariants 2 and of a certain curve, which then determines 2. She then studies hypothesis under which 1 and 2 determine the curve itself.

The Tamagawa numbers c_p of E are

$$c_p = \#\Phi_{E,p}(\mathbb{F}_p).$$

Since $\Phi_{E,p}(\mathbb{F}_p)$ is a subgroup of $\Phi_{E,p}(\overline{\mathbb{F}}_p)$, the numbers c_p divide the numbers $\overline{c}_p = \#\Phi_{E,p}(\overline{\mathbb{F}}_p)$ in Ogg's formula. The Tamagawa numbers appear in the *conjectural* formula of Birch and Swinnerton-Dyer:

$$\frac{L^r(E,1)}{r!} = \frac{\Omega_E \cdot \prod_{p\mid N} c_p \cdot \mathrm{Reg}_E \cdot \#\text{Ш}(E)}{\#E(\mathbb{Q})^2_{\mathrm{tor}}},$$

where r is conjectured to be both $\mathrm{ord}_{s=1} L(E,s)$ and the rank of $E(\mathbb{Q})$.

> **Summary:** The conductor of an elliptic curve is obtained from the minimal discriminant by reducing the power of each prime by the order of the corresponding component group (minus 1).

5.3.6 Definition using Galois representations

Let E be any elliptic curve over $\mathbb{Q}$. For each prime ℓ, consider the two-dimensional Galois representation

$$\overline{\rho}_{E,\ell} : G_{\mathbb{Q}} \to \mathrm{Aut}(E[\ell])$$

obtained by considering the action of the absolute Galois group $G_{\mathbb{Q}} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the group

$$E[\ell] = \{P \in E(\overline{\mathbb{Q}}) : \ell P = 0\}$$

of ℓ-torsion points on E. We associate an integer $N(\overline{\rho}_{E,\ell})$, called the *Serre level*, to this representation. Note that Serre level, as we will define it, makes sense for *any* Galois representation $\overline{\rho} : G_{\mathbb{Q}} \to V$ over $\mathbb{F}_\ell$, not just for representations attached to elliptic curves. Write

$$N(\overline{\rho}) = \prod_{\text{primes } p \neq \ell} p^{n(p)},$$

where it remains to define the numbers $n(p)$, which will measure subtle properities of the ramification of $\overline{\rho}$.

Since $\overline{\rho}$ is continuous and $\mathrm{Aut}(V)$ is a finite set, there is a finite Galois exension $K/\mathbb{Q}$ such that $\overline{\rho}$ factors through $\mathrm{Gal}(K/\mathbb{Q})$. We have a surjective map $G_{\mathbb{Q}} \twoheadrightarrow \mathrm{Gal}(K/\mathbb{Q})$ followed by an inclusion $\mathrm{Gal}(K/\mathbb{Q}) \subset \mathrm{Aut}(V)$, with K a Galois extension. More concreately, the field K is the extension of $\mathbb{Q}$ generated by the x and y coordinates of all ℓ-torsion points on E, which we often write

$$K = \mathbb{Q}(E[\ell]).$$

Fix a prime number p, let $\mathcal{O}_K$ be the ring of integers of K,, then factor p as a product of prime ideals

$$p\mathcal{O}_K = \prod_{i=1}^{g} \mathfrak{p}_i.$$

Let $\mathfrak{p}$ be a choice of any one of the $\mathfrak{p}_i$; which we choose does not impact the definition at all, as it turns out, since the primes $\mathfrak{p}_i$ are all conjugate under the action of $\mathrm{Gal}(K/\mathbb{Q})$. With these choices, the decomposition group is

$$D = \{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) : \sigma(\mathfrak{p}) = \mathfrak{p}\}.$$

Consider the finite field $\mathbb{F}_{\mathfrak{p}} = \mathcal{O}_K/\mathfrak{p}$. One proves that there is an exact sequence

$$1 \to I \to D \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p) \to 1,$$

where the *inertia group* I is by definition the kernel, and surjectivity of the map $D \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$ is not supposed to be obvious. and surjectivity of the map $D \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$ is not supposed to be obvious. Thus

$$I = \{\sigma \in D : \sigma(x) \equiv x \pmod{\mathfrak{p}} \text{ all } x \in \mathcal{O}_K\},$$

and it is natural to define further smaller subgroups of D by letting

$$\begin{aligned} G_i &= \{\sigma \in I : \sigma(x) \equiv x \pmod{\mathfrak{p}^{i+1}} \text{ all } x \in \mathcal{O}_K\} \\ &= \ker(I \to \mathrm{Aut}((\mathcal{O}_K/\mathfrak{p}^{i+1})/(\mathbb{Z}/p^{i+1}\mathbb{Z}))) \end{aligned}$$

for all $i \geq 0$. These finite groups

$$G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots$$

are the *higher ramification groups* associated to our fixed choice of prime $\mathfrak{p}$ over p.

Finally, we define

$$n(p) = \sum_{i=0}^{\infty} \frac{1}{[G_0 : G_i]} \dim V/V^{G_i},$$

where

$$V^{G_i} = \{x \in V : \sigma(x) = x \text{ all } \sigma \in G_i\}.$$

The groups G_i are trivial for all sufficiently large i, so the sum has only finitely many nonzero terms. The representation sentation V is unramified at p if and only if G_0 acts trivially, so we have $n(p) = 0$ if and only if V is unramified at p (for primes $p \neq \ell$).

Returning to our elliptic curve E, we have now defined, for each prime ℓ, an integer $N(\overline{\rho}_{E,\ell})$, which is by definition coprime to ℓ.

Definition 5.3.3 (Conductor)**.** The *conductor* N_E of E is the least common multiple of the integers $N(\overline{\rho}_{E,\ell})$, over all primes ℓ.

Corollary 5.3.4. *A prime p is ramified in $\mathbb{Q}(E[\ell])$ for some $\ell \neq p$ if and only if $p \mid N_E$.*

The *criterion of Néron-Ogg-Shafarevich* is an improvement of the above corollary—it's a criterion purely in terms of ramification—for whether or not a prime p is a prime of bad reduction for an elliptic curve.

Theorem 5.3.5 (Criteron of Néron-Ogg-Shafarevich)**.** *A prime p is ramified in $\mathbb{Q}(E[\ell])$ for all but finitely many $\ell \neq p$ if and only if $p \mid N_E$.*

(In fact, their theorem is stronger, asserting that $p \mid N_E$ if and only if p is ramified in $\mathbb{Q}(E[\ell^\infty])$ for all $p \neq \ell$.)

An Explicit example

Let E be the Fermat Cubic, given in short Weierstrass form by the equation $y^2 = x^3 - 432$. Let's try to compute the conductor using Galois representations. First fix $\ell = 2$ and start by computing $N(\overline{\rho}_{E,\ell})$. We will need to find $K = \mathbb{Q}(E[\ell])$. It is not hard to see the 2-torsion points have to have $y = 0$ so they are exactly the cube roots of 432. Therefore K is the Galois closure of $\mathbb{Q}(432^{1/3})$.

Next we want to find the space V of ℓ-torsion points that $\bar{\rho}_{E,\ell}$ acts on. In sage this can be calculated with

```
E = EllipticCurve(’27a1’).short_weierstrass_model(); \
   E
K.<a> = QQ[432^(1/3)].galois_closure()
V = E.change_ring(K)(0).division_points(2); V
```

which outputs

```
Elliptic Curve defined by y^2 = x^3 - 432 over \
    Rational Field
[(0 : 1 : 0),
     (1/18144*a^4 + 13/42*a : 0 : 1),
     (-1/54432*a^4 - 55/126*a : 0 : 1),
     (-1/27216*a^4 + 8/63*a : 0 : 1)]
```

Note that we know V has 4 elements because it is a 2 dimensional $\mathbb{F}_\ell$-vector space.

Recall $N(\bar{\rho}_{E,\ell}) = \prod_{\text{primes } p \neq l} p^{n(p)}$. So the next step is to find the $n(p)$. Fix p and fix P a prime ideal in $\mathcal{O}_K$ lying over p. To find $n(p)$ we need to find the ramification groups G_i. It can be shown that p is ramified in $\mathcal{O}_K$ if and only if G_0 (the inertia subgroup) is nontrivial. Hence $N(\bar{\rho}_{E,\ell})$ may be defined as the product over primes that ramify in K (as unramified primes p will have $n(p) = 0$). A prime p ramifies in K if and only if p divides the discriminant of K. Using Sage by typing `K.discriminant()`, we find that the discriminant of K is $-34992 = -2^4 \cdot 3^7$, so the only p to check is 3 (we skip 2 since $\ell = 2$). Next we can calculate the ramification groups explicitly:

```
p = 3
G = K.galois_group()
P = K.factor(p)[0][0]
G_0 = G.ramification_group(P,0); G_0.cardinality()
G_1 = G.ramification_group(P,1); G_1.cardinality()
G_2 = G.ramification_group(P,2); G_2.cardinality()
```

which outputs

```
6
3
1
```

So G_0 is the whole Galois group, G_1 is the unique group of index 2, and G_2 is trivial. Next we find the corresponding fixed subspace of V by each group, though we may throw out G_2 as it is trivial.

```
V_0 = [v for v in V if all(E.change_ring(K).point([h(\
    v[0]),h(v[1]),h(v[2])])
     == v for h in G_0)]; V_0
V_1 = [v for v in V if all(E.change_ring(K).point([h(\
    v[0]),h(v[1]),h(v[2])])
     == v for h in G_1)]; V_1
```

which outputs

```
[(0 : 1 : 0)]
[(0 : 1 : 0)]
```

hence $V_0 = V^{G_0}$ and $V_1 = V^{G_1}$ are both trivial. So then

$$n(3) = \sum_{i=0}^{1} \frac{1}{[G_0 \colon G_i]} \dim V/V^{G_i} = \frac{1}{1}(2) + \frac{1}{2}(2) = 3$$

By above $p = 3$ was the only prime that ramifies in K so we have

$$N(\bar{\rho}_{E,2}) = \prod_{\text{primes } p \neq \ell} p^{n(p)} = 3^3 = 27.$$

Since we already know the conductor of E is 27 we are done. Note that this definition is not great computationally since there is no obvious way to know how many ℓ's to use in calculating $\operatorname{lcm} N(\bar{\rho}_{E,\ell})$. It happened with this example that only one ℓ was needed.

Serre's conjecture

Since we're so close, it's worth mentioning Serre's conjecture, though we have not yet discussed how to associate Galois representations $\overline{\rho}_{f,\lambda}$ to modular forms f.

Theorem 5.3.6 (Serre's Conjecture). *Suppose $\overline{\rho} : G_{\mathbb{Q}} \to V$ is a 2-dimensional absolutely irreducible mod ℓ Galois representation such that*

$$\det(\overline{\rho}(\text{complex conjugation})) = -1.$$

Then there is a modular eigenform of level $N(\overline{\rho})$ such that $\overline{\rho} \approx \overline{\rho}_{f,\lambda}$, for some prime $\lambda \mid \ell$.

Serre also gives a formula for a weight $k(\overline{\rho}) \in \mathbb{Z}$, and conjectures that $f \in S_{k(\overline{\rho})}(\Gamma_1(N(\overline{\rho})))$. I *think* all cases of Serre's conjectures are now known (unless there's some weird corner case when $\ell = 2$), due to work of Khare, Wintenberger, Dieulefait, Ribet, Diamond, and many others, which grew out of Taylor-Wiles's work on Fermat's Last Theorem. There are also generalizations of this result to totally real fields. Computational, the above theorem is often extremely helpful, since it provides a specific finite dimensional space in which the relevant modular form must live.

> **Summary:** The conductor of an elliptic curve is the least common multiple of the Serre levels of the mod ℓ Galois representations associated to the elliptic curve.

5.3.7 The conductor of E over a number field

Then

$$N = \prod_{\mathfrak{p}} \mathfrak{p}^{f(E/K_{\mathfrak{p}})}$$

where the product runs over all nonzero prime ideals $\mathfrak{p}$ of $\mathcal{O}_K$ and

$$f(E/K_{\mathfrak{p}}) = \begin{cases} 0 & \text{if } E/K_{\mathfrak{p}} \text{ has good reduction at } \mathfrak{p} \\ 1 & \text{if } E/K_{\mathfrak{p}} \text{ has multiplicative reduction at } \mathfrak{p} \\ 2 & \text{if } \mathfrak{p} | p \geq 5 \text{ and } E/K_{\mathfrak{p}} \text{ has additive reduction at } \mathfrak{p} \end{cases}$$

and

$$2 \leq f(E/K_{\mathfrak{p}}) \leq 2 + 3\operatorname{ord}_{\mathfrak{p}}(3) + 6\operatorname{ord}_{\mathfrak{p}}(2)$$

if $\mathfrak{p}|p$ for $p \in \{2, 3\}$ and $E/K_{\mathfrak{p}}$ has additive reduction at $\mathfrak{p}$ (note that only one of $\operatorname{ord}_{\mathfrak{p}}(3)$ and $\operatorname{ord}_{\mathfrak{p}}(2)$ is nonzero).

```
sage: E = EllipticCurve('256a')
sage: print E
sage: for n in range(1,8):
sage:     K.<alpha> = NumberField(x^n - 2)
sage:     EK = E.base_extend(K);
sage:     print n, EK.conductor().factor()
Elliptic Curve defined by y^2 = x^3 + x^2 - 3*x + 1 \
   over Rational Field
(Fractional ideal (2))^8
(Fractional ideal (alpha))^10
(Fractional ideal (alpha))^20
(Fractional ideal (alpha))^8
(Fractional ideal (-alpha))^32
(Fractional ideal (2, alpha))^26
(Fractional ideal (2, alpha))^44
```

```
sage: E = EllipticCurve('243a');
sage: print E
sage: for n in range(1,8):
sage:     K.<alpha> = NumberField(x^n - 3)
sage:     EK = E.base_extend(K);
sage:     print EK.conductor().factor()
Elliptic Curve defined by y^2 + y = x^3 - 1 over \
   Rational Field
(Fractional ideal (3))^5
(Fractional ideal (alpha))^8
(Fractional ideal (alpha))^3
(Fractional ideal (alpha))^14
(Fractional ideal (-alpha))^17
(Fractional ideal (3, alpha))^4
(Fractional ideal (3, alpha))^23
```

5.4 Modularity of elliptic curves over $\mathbb{Q}$

Barry Mazur once wrote a great article entitled *Number Theory as Gadfly* which is about what it means for an elliptic curve over $\mathbb{Q}$ to be *modular.*

Theorem 5.4.1 (Breuil, Conrad, Diamond, Taylor, Wiles)**.** *Every elliptic curve over $\mathbb{Q}$ is modular.*

This theorem is incredibly important to numerous algorithms for computing with elliptic curves. As with the definition of conductor, there are many ways of viewing modularity, with deep connections between them.

5.4.1 Definition using modular forms

Let E be an elliptic curve. In Section 5.3.4 we defined a function

$$f = f_E(q) = \sum_{n=1}^{\infty} a_n q^n,$$

and it is relatively easy to prove that f_E defines a holomorphic function on the upper half plane

$$\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\} \subset \mathbb{C}.$$

We then defined the conductor N as the smallest positive integer so that

$$f(z)dz = \frac{f(q)}{q}dq$$

is invariant under the action of the congruence subgroup $\Gamma_0(N)$. Theorem 5.4.1 is the assertion that there is *some* N such that f is invariant under $\Gamma_0(N)$. The function f is then a normalized cuspidal modular eigenforms level N and weight 2, and we write

$$f \in S_2(\Gamma_0(N))_{\mathrm{new}}$$

> **Summary:** An elliptic curve E over $\mathbb{Q}$ is modular if the function $f(q) = \sum a_n q^n$, with $a_p = p + 1 - \#E(\mathbb{F}_p)$ is a modular form.

5.4.2 (*) Definition using systems of Hecke eigenvalues

This section gets a star because it is how I think about modularity.

The *Hecke algebra* is a commutative ring $\mathbb{T}$ that acts on the space $S_2(\Gamma_0(N))$ of cusp forms. The Hecke algebra is generated by Hecke operators T_n, one for each integer n

$$\mathbb{T} = \mathbb{Z}[T_1, T_2, T_3, \ldots].$$

That $f = \sum a_n q^n$ is a normalized eigenform means that for each integer n, we have

$$T_n(f) = a_n \cdot f,$$

i.e., the coefficients of f are the eigenvalues of the linear transformations T_n. There are also recurrences satisfied by the T_n, so that knowing them for prime n determines all of them. Thus alternatively, E is modular if there is some eigenvector $v \in S_2(\Gamma_0(N))$ such that for all primes p we have

$$T_p(v) = (p + 1 - \#E(\mathbb{F}_p)) \cdot v.$$

> **Summary:** An elliptic curve E over $\mathbb{Q}$ is modular if the point counts $\#E(\mathbb{F}_p)$ are encoded in the system of Hecke eigenvalues of some eigenvector in $S_2(\Gamma_0(N))$.

Moreover, and this is key, one can replace $S_2(\Gamma_0(N))$ by any $\mathbb{T}$-module that is isomorphic to it.

5.4.3 Definition in terms of modular curves

Let E be an elliptic curve over $\mathbb{Q}$, let N be the conductor of E, and consider the left action of $\Gamma_0(N)$ on the extended upper half plane

$$\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{Q} \cup \{i\infty\}$$

by linear fractional transformations:

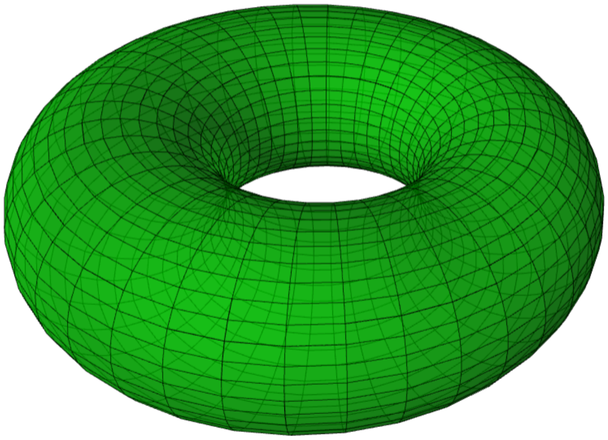
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az+b}{cz+d}.$$

The modular curve of level $\Gamma_0(N)$ is the compact Riemann *surface*

$$X_0(N)(\mathbb{C}) = \Gamma_0(N)\backslash\mathfrak{h}^*.$$

For example, when $N = 11$, the real surface $X_0(11)(\mathbb{C})$ has genus 1:

```
sage: dimension_cusp_forms(11)
1
sage: u = var('u')
sage: revolution_plot3d((cos(u),sin(u)),(u,0,2*pi),\
    axis=(2,0),color='green',
       opacity=.7, mesh=True, frame=False)
```



Developing the theory of modular functions, one proves that $X_0(N)(\mathbb{C})$ is the set of complex points of an algebraic curve $X_0(N)$ defined over $\mathbb{Q}$ (or even over $\mathbb{Z}$ – see Katz-Mazur!). Theorem 5.4.1 then asserts that there is a surjective morphism of algebraic curves $X_0(N) \to E$.

Remark 5.4.2. Fun and useless fact: There are modular curves $X_0(N)$ of genus g for each $g \leq 149$, but there is no modular curve of genus 150. See Csirik, et al.

5.4.4 Definition in terms of L-series

In Section 5.3.3 we defined the conductor of E as a constant that arises in the functional equation for $L(E, s)$. The assertion that E is modular is *a priori* a *stronger* statement than the assertion that $L(E, s)$ is entire and satisfies the functional equation (5.3.2). If E is modular, than $L(E, s) = L(f, s)$ is indeed entire and satisfies the functional equation; however, as f is a modular form, we also have for every Dirichlet character χ, that $L(f^{\chi}, s)$ is also entire and satisfies a functional equation. There is a *converse theorem*, due to Weil, which asserts that if $L(E, s)$ is entire and satisfies a functional equation, *and* so do "all of its twists", then $L(E, s) = L(f, s)$ for some modular form f. *Warning: I don't know the precise formulation of this statement – exercise for a student in class to fill this in.*

The above converse theorem was considered for a long time to be the best evidence for the modularity conjecture before it was proved. This might be why the modularity statement has had (at least) the following names:

1. The conjecture of Weil
2. The Taniyama-Shimura-Weil conjecture
3. The Taniyama-Weil conjecture:
   page 341 of *A Course in Number Theory*, By H. E. Rose

4. The Shimura-Taniyama conjecture:
`http://www.ams.org/notices/199511/forum.pdf`

5. The Modularity theorem:
`http://en.wikipedia.org/wiki/Modularity_theorem`

5.5 Converse theorems

We have seen that given a modular form f of weight k and conductor N, one can associate an L-series $L(f,s)$ to it. The L-series $L(f,s)$ can be analytically continued to the whole complex plane and it satisfies a certain functional equation that we describe below.

Define a holomorphic function $\Lambda_N(s,f) = (2\pi/\sqrt{N})^{-s}\Gamma(s)L(s,f)$, where the function $\Gamma(s)$ is the classical Gamma function that extends the factorial function. Also, define a holomorphic function $g(z)$ as follows $g(z) = (-i\sqrt{N}z)^{-k}f(-1/Nz)$. One sees that g is also a modular form of weight k and level N. Now, we have the following functional equation

$$\Lambda_N(s,f) = \Lambda_N(k-s,g). \tag{5.5.1}$$

The converse theorems deal with the following question (as is indicative of the name) - given an L-series that satisfies a certain functional equation, does the Mellin transform of the L-series correspond to a modular form ?

The first converse theorem was proved by Hecke ([Hec36], [Hec59]), for modular forms of level 1.

Theorem 5.5.1 (Hecke)**.** *Let $f(z) = \sum_{n=0}^{\infty} a_n e^{2\pi i z}$ be a holomorphic function on the upper half plane $\mathbb{H}$. Let $k \geq 2$ be an even number. Assume that there exists a $\nu > 0$, such that for all $n \in \mathbb{N}$, $a_n = O(n^\nu)$. The following statements are equivalent*

1. *$f(z)$ is a modular form of level 1 and weight k.*

2. *The function $\Lambda(s,f)$ can be analytically continued to the whole s-plane, satisfying the functional equation*

$$\Lambda_1(s,f) = (-1)^{k/2}\Lambda(k-s,f) \tag{5.5.2}$$

$$\Lambda_1(s,f) = (-1)^{k/2}\Lambda(k-s,f) \tag{5.5.3}$$

and

$$\Lambda(s,f) + a_0/s + (-1)^{k/2}a_0/(k-s) \tag{5.5.4}$$

is holomorphic on the whole s-plane and bounded on any vertical strip.

This was further extended by Weil ([Wei67]) for modular forms, without any restrictions on the level. However, all the twisted L-series should also satisfy certain functional equations.

Theorem 5.5.2 (Weil)**.** *Fix positive integeres N and k and suppose that $L(s) = \sum_{n=1}^{\infty} a_n/n^s$ satisfies the following condition*

1. *$L(s)$ is absolutely convergent for $Re(s)$ sufficiently large.*

2. *for each primitive Dirichlet character χ of modulus r and $(r, N) = 1$,*

$$\Lambda(s,\chi) = (2\pi)^{-2}\Gamma(s)\sum_{n=1}^{\infty} a_n\chi(n)n^{-s} \tag{5.5.5}$$

   continues to an entire function of s and bounded on vertical strips.

3. *Each such $\Lambda(s,\chi)$ satisfies the functional equation*

$$\Lambda(s,\chi) = \omega_\chi r^{-1}(r^2N)^{k/2-s}\Lambda(k-s,\overline{\chi}), \tag{5.5.6}$$

   where

$$\omega_\chi = i^k\chi(N)g(\chi)^2 \tag{5.5.7}$$

   and the Gauss sum

$$g(\chi) = \sum_{n \mod r} \chi(n)e^{2\pi in/r} \tag{5.5.8}$$

Then the function $f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi iz}$ belongs to $S_2(\Gamma_0(N))$.

One can ask if Weil's converse theorems can be used to prove modularity. However, it somehow requires us to know that the Hasse-Weil L-function is entire. This seems difficult. Indeed, proving this is a major component of the BSD conjecture.

Check this letter by Jean-Pierre Serre to David Goss - `http://smf4.emath.fr/Publications/Gazette/2002/91/smf_gazette_91_55-57.pdf`. It seems that Serre was initially skeptical about the truth of the modularity conjecture, and that Weil's converse theorems provided him evidence to believe in the modularity conjecture. It seems that this convinced Serre that the name of Weil should be included in the Taniyama-Shimura-Weil conjecture. Also, see a mathoverflow answer here - `http://mathoverflow.net/questions/75335/what-is-the-reason-for-modularity-results?answertab=votes#tab-top`.

5.6 Modularity over real quadratic fields

A number field K is *totally real* if every embedding $K \to \mathbb{C}$ has image in $\mathbb{R}$. If we order totally real fields by the absolute value of their discriminant, than the list is

$$\mathbb{Q}, \mathbb{Q}(\sqrt{5}), \mathbb{Q}(\sqrt{2}), \ldots$$

Exercise 5.6.1. The implicit exercise in the above statement is to show that if K has degree at least 3, then the absolute value of the discriminant of K is greater than $8 = \mathrm{disc}(\mathbb{Q}(\sqrt{2}))$. We can show this using the Minkowksi bound. Also, searching `http://hobbes.la.asu.edu/NFDB/` shows that the smallest discriminant totally real cubic field is defined by $x^3 - x^2 - 2x + 1$ and has discriminant 49.

The following major new result appeared on the arxiv recently (see `http://arxiv.org/abs/1310.7088`).

Theorem 5.6.2 (Freitas, Le Hung, Siksek)**.** *Every elliptic curve over a real quadratic field is modular.*

Part of their motivation was to generalize Fermat (see Section 3.2) to real quadratic fields, which is the subject of another paper they recently posted at `http://arxiv.org/abs/1307.3162`.

5.6.1 Definition in terms of Hilbert modular forms

We sketch one interpretation of what it means for an elliptic curve E over a real quadratic field $K = \mathbb{Q}(\sqrt{d})$ to be *modular*. For this, we will introduce Hilbert modular forms, following Dembele-Voight's `http://arxiv.org/abs/1010.5727`.

Let $\mathfrak{n}$ be an ideal in the ring $\mathcal{O}_K$ of integers of K, and fix the two distinct embeddings

$$\iota_1 : K \hookrightarrow \mathbb{R} \quad \text{and} \quad \iota_2 : K \hookrightarrow \mathbb{R}.$$

Let $\mathrm{GL}_2^+(\mathcal{O}_K)$ be the group of 2×2 matrices γ with entries in $\mathcal{O}_K$ such that both $\iota_1(\det(\gamma)) > 0$ and $\iota_2(\det(\gamma)) > 0$. The analogue over K of the congruence subgroup $\Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$ is

$$\Gamma_0(\mathfrak{n}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathcal{O}_K) : c \in \mathfrak{n} \right\}.$$

For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(K)$, we let $a_i, b_i, c_i, d_i \in \mathbb{R}$ denote the entries of $\iota_i(\gamma)$, for $i = 1, 2$.

A *holomorphic function* on an open subset of $\mathbb{C} \times \mathbb{C}$ is a complex-valued function that is continuous and holomorphic in each of its variables. Suppose $f = f(z_1, z_2) : \mathfrak{h} \times \mathfrak{h} \to \mathbb{C}$ be a holomorphic function on the product of two copies of the upper half plane. For $\gamma \in M_2(K)$, let

$$(f|\gamma)(z) = \frac{\det \gamma_1}{(c_1 z_1 + d_1)^2} \cdot \frac{\det \gamma_2}{(c_2 z_2 + d_2)^2} f(\gamma(z))$$

Definition 5.6.3 (Hilbert Modular Form)**.** A *Hilbert modular form* over K of parallel weight $(2, 2)$ and level $\mathfrak{n}$ is a holomorphic function $f : \mathfrak{h} \times \mathfrak{h} \to \mathbb{C}$ such that for all $\gamma \in \Gamma_0(\mathfrak{n})$, we have $f|\gamma = f$.

Let $M_{(2,2)}(\Gamma_0(\mathfrak{n}))$ denote the space of Hilbert modular forms f, as above.

As in Section 5.4.2, there are commuting Hecke operators $T_{\mathfrak{m}}$ acting on $M_{(2,2)}(\Gamma_0(\mathfrak{n}))$, for each nonzero ideal $\mathfrak{m}$ of $\mathcal{O}_K$. Together these generate (over $\mathbb{Z}$) a commutative Hecke algebra $\mathbb{T}$. For example, suppose K has narrow class number 1 (so every ideal is generated by a totally positive element), and that π is a totally positive generator for a prime ideal $\mathfrak{p} \nmid \mathfrak{n}$; then, we have

$$T_{\mathfrak{p}}(f) = f|\left(\begin{smallmatrix} \pi & 0 \\ 0 & 1 \end{smallmatrix}\right) + \sum_{\alpha \in \mathbb{F}_{\mathfrak{p}}} f|\left(\begin{smallmatrix} 1 & \alpha \\ 0 & \pi \end{smallmatrix}\right).$$

The elliptic curve E over K of conductor $\mathfrak{n}$ is *modular* if there is an eigenvector $f \in M_{(2,2)}(\Gamma_0(\mathfrak{n}))$ such that for all $\mathfrak{p} \nmid \mathfrak{n}$ we have

$$T_{\mathfrak{p}}(f) = (\mathrm{Norm}(\mathfrak{p}) + 1 - \#E(\mathbb{F}_{\mathfrak{p}})) \cdot f.$$

5.6.2 The First example

The paper `http://wstein.org/papers/sqrt5/` is about enumerating *all* elliptic curves over $K = \mathbb{Q}(\phi)$, with $\phi = \frac{1+\sqrt{5}}{2}$ the golden ratio, up to the first curve of rank 2. The smallest level so that there is an elliptic curve is $\mathfrak{n} = (5\phi - 3)$, which is an ideal of norm 31. A curve of conductor $\mathfrak{n}$ is

$$y^2 + xy + \phi y = x^3 + (-\phi - 1)\, x^2.$$

In the paper, we enumerate the curves of norm conductor up to 1831 by first finding all Hilbert modular forms of this this level, then for each form, we find a corresponding elliptic curve *by hook or crook*. Unlike the situation with elliptic curves over $\mathbb{Q}$, there's no straightforward way (in general) to just write down the curve corresponding to a Hilbert modular form, so we use combination of many ad hoc search procedures. Computing the Hilbert modular forms turns out be fast, using an algorithm of Dembele (see his *Explicit Computations of Hilbert Modular Forms on* $\mathbb{Q}(\sqrt{5})$), though implementing that algorithm in a way that is fast is challenging.

5.7 Computing all elliptic curves over the rational numbers using modular symbols

Cremona's book *Algorithms for Modular Elliptic Curves* is about the following:

> **Problem:** Given an integer N, find all elliptic curves over $\mathbb{Q}$ of conductor N.

Recall from Section 5.4.2 that for an elliptic curve E to be modular means that the integers

$$a_p = a_p(E) = p + 1 - \#E(\mathbb{F}_p) \in \mathbb{Z},$$

for p prime, are also the eigenvalues of the Hecke operators T_p acting on some eigenvector in the space of cusp forms $S_2(\Gamma_0(N))$.

To solve the above problem, we have the following strategy:

1. Compute a $\mathbb{T}$-module V such that for every elliptic curve of conductor N, the system of eigenvalues $\{a_p(E)\}$ is guaranteed to occur in V. For example, one could take $V = S_2(\Gamma_0(N))$.
2. Given finitely many of the integers $\{a_p(E)\}$, find an equation for a curve E with those a_p.
3. Find the finitely many curves isogenous to E.

5.7.1 Computing systems of eigenvalues using modular symbols

Modular symbols is one powerful systematic approach to finding all systems of rational Hecke eigenvalues $a_p(E)$ corresponding to elliptic curves of conductor N. The basic idea is to explicitly compute the homology group

$$H_1(X_0(N), \mathbb{Z}),$$

which is equipped with an action of Hecke operators T_p. Then we do linear algebra with these matrices to compute the systems of Hecke eigenvalues.

Let $\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{Q} \cup \{\infty\}$ Recall that $X_0(N) = \Gamma_0(N)\backslash\mathfrak{h}^*$, where the *cusps* are the elements of

$$\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}.$$

To compute $H_1(X_0(N), \mathbb{Z})$, we consider the slightly bigger relative homology group

$$H_1(X_0(N), \mathbb{Z}) \subset H_1(X_0(N), \mathbb{Z}; \{\text{cusps}\}).$$

Manin proveed that $H_1(X_0(N), \mathbb{Z}; \{\text{cusps}\})$ is generated by all geodesic paths $\{\alpha, \beta\}$ between elements $\alpha, \beta \in \mathbb{P}^1(\mathbb{Q})$. More precisely, Manin proved that $H_1(X_0(N), \mathbb{Z}; \{\text{cusps}\})$ is the quotient of the free abelian group generated by all symbols $\{\alpha, \beta\}$, modulo the following relations, and modulo any torsion:

- $\{\alpha, \beta\} = -\{\beta, \alpha\}$,

- $\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$.

Remark 5.7.1. The notation $\{\alpha, \beta\}$ means the path *from* α to β, i.e., it is *ordered*, despite being written like a set – this is a historical accident, I guess, but doesn't cause confusion.

Hecke Operators

For a matrix γ, let $\gamma\{\alpha, \beta\} = \{\gamma(\alpha), \gamma(\beta)\}$. The Hecke operators T_p, for p prime to N, act on $H_1(X_0(N), \mathbb{Z}, \{\text{cusps}\})$ as follows:

$$T_p(\{\alpha, \beta\}) = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \{\alpha, \beta\} + \sum_{r=0}^{p-1} \begin{pmatrix} 1 & r \\ 0 & p \end{pmatrix} \{\alpha, \beta\}.$$

When $p \mid N$ the action is the same, except we omit $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

The $*$-Involution

Complex conjugation on $X_0(N)$ induces an involution on $H_1(X_0(N), \mathbb{Z}, \{\text{cusps}\})$ given by

$$*\{\alpha, \beta\} = \{-\alpha, -\beta\}.$$

The star involution $*$ commutes with the Hecke operators (and everything else), and for applications we frequently compute in the quotient vector space

$$H_1(X_0(N), \mathbb{Q}, \{\text{cusps}\})/(* - 1),$$

since the (new) eigenspaces for the Hecke operators appear with multiplicity 1 instead of 2 here, which makes many problems an order of magnitude easier.

Manin's trick

Manin observed using continued fractions that there is an explicit *finite* subset of paths $\{\alpha, \beta\}$ that generate $H_1(X_0(N), \mathbb{Z}; \{\text{cusps}\})$.

Proposition 5.7.2. *Write*

$$\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{Z}) = \Gamma_0(N)\gamma_1 \ \cup \cdots \cup \Gamma_0(N)\gamma_d,$$

where $\gamma_1, \ldots, \gamma_d$ are a choice of right coset representatives for the (not normal unless $N = 1$) subgroup $\Gamma_0(N)$ of $\mathrm{SL}_2(\mathbb{Z})$. The homology group $H_1(X_0(N), \mathbb{Z}; \{cusps\})$ is generated by the symbols $\{\gamma_i(0), \gamma_i(\infty)\}$ for $i = 1, \ldots, d$.

Proof. The proof is a straightforward argument that $\{0, \alpha\}$ can be written in terms of such symbols using the partial convergents of the continued fraction of αe [Ste07, Prop. 3.11] for two proofs, and also Chapter 2 of [Cre97]. □

Manin's presentation

There is a bijection between the right cosets

$$\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{Z})$$

and the elements of

$$\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z}) = \{(c:d) : 1 \leq c, d \leq N, \gcd(c,d,N) = 1\}/(\mathbb{Z}/N\mathbb{Z})^*.$$

A right coset $\Gamma_0(N)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ corresponds to the element $(c:d) \in \mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$.

The *Manin symbol* (c,d) is

$$(c,d) = \left\{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(0), \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)(\infty)\right\} = \left\{\frac{b}{d}, \frac{a}{c}\right\} \in H_1(X_0(N), \mathbb{Z}, \{\text{cusps}\}).$$

where $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ is any choice of matrix such that $c \equiv u \pmod N$ and $d \equiv v \pmod N$.

Manin's trick ensures that the finitely many symbols (c,d) generate $H_1(X_0(N), \mathbb{Z}, \{\text{cusps}\})$, as $(c:d)$ runs through representatives for $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. Manin also figured out exactly what the relations are between the Manin symbols.

Each element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ acts on the right on pairs (u,v) by matrix multiplication (of a row vector times a matrix). The group $\mathrm{SL}_2(\mathbb{Z})$ is generated by the elements $\sigma = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ of order 4 and $\tau = 1{-}110$ of order 6.

Theorem 5.7.3 (Manin)**.** *The relative homology group $H_1(X_0(N), \mathbb{Z}, \{cusps\})$ is isomorphic to the quotient of the free abelian group on symbols $x = (c,d)$, for $(c:d) \in \mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$, modulo the following relations and modulo any torsion:*

- $x + x\sigma = 0,$
- $x + x\tau + x\tau^2 = 0.$

For a partial proof and references, see [Ste07, Thm. 3.13].

We now have an algorithm to compute the space of modular symbols of level N:

1. Explicitly enumerate the elements of $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$.
2. Quotient out by the 2-term and 3-term relations listed above.

We do the quotient process over $\mathbb{Q}$ as follows. First, we quotient out by the 2-term relations, which basically identifies half of the symbols the negatives of the other half. We then create a sparse matrix with rows the 3-term relations, but with the 2-term relations identified, and put it in reduced row echelon form. From the echelon form, we read off free generators, i.e., a subset of Manin symbols, and a way of explicitly writing all other Manin symbols as $\mathbb{Q}$-linear combinations of this subset.

We usually do not need a presentation over $\mathbb{Z}$, but if we do, one way to compute it is to use Hermite normal form to find a basis for the $\mathbb{Z}$-module generated by the symbols (u,v) inside the $\mathbb{Q}$-vector space that we just computed.

5.7.2 Computing all curves with given conductor

Our motivating problem is to give an algorithm to compute each elliptic curve over $\mathbb{Q}$ of conductor N. In this section, we *sketch* an algorithm. *For every single step below, we have omitted numerous tricks and optimizations that speed up the computation, which are absolutely critical in practice*; they are mostly described in detail on [Cre97].

1. **Presentation for** $H = H_1(X_0(N), \mathbb{Q}, \{\text{cusps}\})/(1 - *)$: Compute a list of linearly independent Manin symbols $x_i = (u, v)$ such that every Manin symbol is an explicit $\mathbb{Q}$-linear combination of the x_i.

2. **Find the new rational eigenspaces:** Let p be the smallest prime not dividing N, and compute the matrix with respect to the basis of x_i for the Hecke operator T_p acting on H. Compute the characteristic polynomial f of T_p, factor f as a product $f = \prod g_i^{e_i}$, and compute each subspace $V_i = \ker(g_i(T_p))$ for which g_i is a linear polynomial. If $\dim(V_i) > \deg(g_i^{e_i})$, then discard V_i, since it is old, and has nothing to do with elliptic curves of conductor N. Repeat the above procedure increasing p and computing T_p on the V_i instead of H, until each remaining V_i has dimension 1. (Theorems guarantee this process terminates.)

3. **Modular forms:** To each eigenspace V_i found in Step 2, compute the eigenvalues $a_2, a_3, a_5, \ldots$, of the Hecke operators T_p, up to some bound B. Use the following recurrence to compute a_n for all integers $n \leq B$: for $\gcd(n, m) = 1$, we have $a_{nm} = a_n a_m$, and for prime powers p^r we have $a_{p^r} = a_p a_{p^{r-1}} - \varepsilon(p) p a_{p^{r-2}}$, where $\varepsilon(p) = 1$ if $p \nmid N$ and $\varepsilon(p) = 0$ for $p \mid N$. Let $f_i = \sum_{n=1}^{B} a_n q^n$, which is an approximation to the modular form attached to V_i. To each f_i there is a corresponding elliptic curve $E_i/\mathbb{Q}$, and this gives representives for all elliptic curves over $\mathbb{Q}$ of conductor N, up to isogeny.

4. **Complex lattices:** To each modular form f_i in Step 3, there is a period lattice
$$\Lambda_i = \left\{ \int_\gamma 2\pi i f_i(z) dz : \gamma \in H_1(X_0(N); \mathbb{Z}) \right\} \subset \mathbb{C}.$$
Over $\mathbb{C}$, we have $E_i(\mathbb{C}) \cong \mathbb{C}/\Lambda_i$. Unfortunately, I have not yet explained how to compute either $H_1(X_0(N); \mathbb{Z})$ or $\int_\gamma 2\pi i f_i(z) dz$, and both are somewhat involved. Moreover, it's critical to algebraically find 2 elements of $H_1(X_0(N); \mathbb{Z})$ that map to a basis of Λ_i, since we only obtain approximations for the elements of Λ_i, and if we have more than 2 we don't know for sure how to obtain a basis from them.

5. **Compute Weierstrass equations:** Once we have computed approximations to the lattices $\Lambda_i \subset \mathbb{C}$, we compute the corresponding Weirstrass equation by using the Weierstrass $\wp$ function. More precisely, write $\Lambda_i = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, such that $\tau = \omega_1/\omega_2 \in \mathfrak{h}$. Apply elements of $\mathrm{SL}_2(\mathbb{Z})$ to

transform τ into an element in the standard fundamental domain, so $\mathrm{Re}(\tau) \leq 1/2$ and $|\tau| \geq 1$. The c_4, c_6 invariants of the corresponding elliptic curve over $\mathbb{Q}$ given by explicit power series in terms of τ and ω_2 (see [Cre97, §2.14]). A deep theorem of Edixhoven (that "the Manin constant is an integer") implies that c_4 and c_6 are integers. We compute them numerically, to sufficient precision to determine them as integers (possibly increasing B above). Once we are done, we write down the Weierstrass equation over $\mathbb{Q}$, recompute the a_n from that, for n sufficiently large to determine f_i, and verify that indeed the curve we found is E_i. Because of the modularity theorem, this final step proves that we really found E_i, despite any numerical issues along the way.

The only complete optimized implementation of this algorithm is

`https://github.com/JohnCremona/eclib`

which is included with Sage, though there is no easy way to use it from the interpreter yet. Cremona has run this algorithm on every integer $N \leq 300{,}000$, and the data is available at

`http://homepages.warwick.ac.uk/~masgaj/ftp/data/`

5.7.3 How to compute period integrals

Suppose we are given an approximation

$$\sum_{n=1}^{B} a_n q^n \in S_2(\Gamma_0(N)),$$

to a newform

$$f = \sum_{n=1}^{\infty} a_n q^n$$

with $a_n \in \mathbb{Z}$. Our goal is to find the corresponding elliptic curve E, which is defined over $\mathbb{Q}$.

As mentioned above, one systematic approach to doing this is to approximate the lattice

$$\Lambda = \left\{ \int_\gamma 2\pi i f(z) dz : \gamma \in H_1(X_0(N), \mathbb{Z}) \right\}.$$

The tricky step in computing Λ is computing integrals of the form

$$\int_\alpha^\beta 2\pi i f(z) dz.$$

A first naive thing to do is to write

$$\begin{aligned}\int_\alpha^\beta 2\pi i f(z)dz &= \int_\alpha^\beta 2\pi i \sum_{n=1}^\infty a_n e^{2\pi i n z} dz\\ \text{``}=\text{''} &\sum_{n=1}^\infty \int_\alpha^\beta 2\pi i a_n e^{2\pi i n z} dz\\ &= \sum_{n=1}^\infty \left(a_n e^{2\pi i n\beta} - a_n e^{2\pi i n \alpha}\right)\end{aligned}$$

If $\beta = i\infty$ the sum is

$$\sum_{n=1}^\infty \frac{a_n}{e^{-\infty}} = 0.$$

However, what happens when $\beta \in \mathbb{Q}$ is a rational number? Then we get

$$\sum_{n=1}^\infty \frac{a_n}{n} e^{2\pi i n \beta},$$

and we can't hope for meaningful convergence!

So we need another idea, which starts with the following observation. For any $\alpha, \beta \in \mathfrak{h}^*$, extend the notation $\{\alpha, \beta\}$ to mean the class of a geodesic path from α to β in $\mathfrak{h}^*$, modulo the action of $\Gamma_0(N)$ and the homology relations. For any $\alpha \in \mathfrak{h}^*$ and $\gamma \in \Gamma_0(N)$, we have

$$\begin{aligned}\{0, \gamma(0)\} &= \{0, \alpha\} + \{\alpha, \gamma(\alpha)\} + \{\gamma(\alpha), \gamma(0)\}\\ &= \{0, \alpha\} + \{\alpha, \gamma(\alpha)\} + \{\alpha, 0\} = \{\alpha, \gamma(\alpha)\}.\end{aligned}$$

Thus

$$\int_0^{\gamma(0)} 2\pi i f(z)dz = \int_\alpha^{\gamma(\alpha)} 2\pi i f(z) dz \tag{5.7.1}$$

$$= \sum_{n=1}^\infty \frac{a_n}{n} e^{2\pi i n \gamma(\alpha)} - \sum_{n=1}^\infty \frac{a_n}{n} e^{2\pi i n \alpha}, \tag{5.7.2}$$

and this converges as long $\alpha \in \mathfrak{h}$. There is a tension that emerges: we can choose α and γ to make the convergence fast, but they might not be useful. In particular, the period map

$$\Phi_f : H_1(X_0(N), \mathbb{Q}, \{\text{ cusps }\}) \to \mathbb{C}$$

is a linear map (over $\mathbb{Q}$), so if we can compute the images of two random elements of $H_1(X_0(N), \mathbb{Q}, \{\text{ cusps }\})$, then we have a shot at computing Φ_f on *anything.* In the next section we illustrate computing periods with an exaple.

5.7.4 Example: Conductor 43

In this section, we sketch how to find all isogeny classes of elliptic curves of conductor $N = 43$ using the modular symbols method. First, we compute the space modular symbols of conductor 43, by computing the quotient of the vector space on generators the Manin symbols

$$x_0 = (0,1), x_1 = (1,0), (1,1), (1,2), \ldots, x_{43} = (1,42)$$

by the relations

$$\begin{aligned} 0 &= (u,v) + (u,v)\sigma \\ 0 &= (u,v) + (u,v)\tau + (u,v)\tau^2. \end{aligned}$$

More explicitly, these relations are

$$\begin{aligned} 0 &= (u,v) + (v,-u) \\ 0 &= (u,v) + (u+v,-u) + (v,-u-v). \end{aligned}$$

Ignoring optimizations, we can do this by creating a matrix with 44 columns, corresponding to the symbols $x_0, \ldots, x_{43}$, and 86 rows, corresponding to the above relations, and computing its reduced row echelon form. Sage does this and gives a subset of 7 Manin symbols, such that all other Manin symbols are uniquely $\mathbb{Q}$-rational linear combinations of these 7, modulo the above relations:

```
sage: M = ModularSymbols(43); M.basis()
((1,0), (1,31), (1,32), (1,38), (1,39), (1,40), \
    (1,41))
```

Anything can be written in terms of these symbols:

```
M( (1,0) )
```

(1,0)

```
M( (2,3) )
```

(1,40) - (1,41)

Remark 5.7.4. There is nothing canonical about this list of 7 symbols; we could have found a different subset of 7 symbols that form a basis. In fact, making different choices can have a major impact on the sparseness of the Hecke operators T_p that we compute later, which can significantly impact the runtime of algorithms.

Next, we compute the Hecke operator T_2 and use it to find an elliptic curve.

```
T2 = M.hecke_operator(2); T2.matrix()
```

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 & 2 & -1 & 1 \\ 0 & -1 & 0 & 1 & 1 & -1 & 1 \\ 0 & 0 & -2 & 0 & 2 & -2 & 1 \\ 0 & 0 & -1 & 0 & 1 & 0 & -1 \end{pmatrix}$$

The characteristic polynomial of this matrix is

```
T2.charpoly().factor()
```

$$(x-3) \cdot (x+2)^2 \cdot (x^2-2)^2$$

The factor of $(x-3)$ corresponds to the Eisenstein series E_2

```
EisensteinForms(43,2).basis()[0].q_expansion(prec=10)\
    *(7/4)
```

$$\frac{7}{4} + q + 3q^2 + 4q^3 + 7q^4 + 6q^5 + 12q^6 + 8q^7 + 15q^8 + 13q^9 + \cdots$$

The other two irreducible factors, $(x+2)$ and x^2-2, correspond to Galois orbits of cuspidal eigenforms.

```
S = CuspForms(43,2); D = S.newforms(names='a')
for f in D:
    print f
```

$$f_1 = q - 2q^2 - 2q^3 + 2q^4 - 4q^5 + \cdots$$
$$f_2 = q + a_1 q^2 - a_1 q^3 + (-a_1 + 2)\, q^5 + \cdots$$

Noe that f_1 has coefficients in $\mathbb{Q}$ and f_2 has coefficients in the quadratic field $\mathbb{Q}(\sqrt{2})$:

```
D[1].base_ring()
```

```
Number Field in a1 with defining polynomial x^2 - 2
```

As a double check, let's find the coefficients of f_1 directly using modular symbols:

```
V = (T2+2).kernel()
[V.hecke_matrix(n)[0,0] for n in [1..6]]
```

[1, -2, -2, 2, -4, 4]

Next, let's find periods! We choose some arbitrary $\alpha \in \mathfrak{h}$ and $\gamma \in \Gamma_0(43)$, and evaluate the period integral $\int_\omega 2\pi i f_1(z) dz$ along the path

$$\{\alpha, \gamma(\alpha)\} = \{0, \gamma(0)\}.$$

Let's just take $\alpha = i = \sqrt{-1}$. To make an interesting element of $\Gamma_0(43)$ use the Euclidean algorithm:

```
xgcd(43,2)
```

(1, 1, -21)

```
gamma = matrix(2,[-21,-1,43,2]); gamma
```

$$\begin{pmatrix} -21 & -1 \\ 43 & 2 \end{pmatrix}$$

We are now in a position to consider (5.7.1)

$$\int_0^{\gamma(0)} 2\pi i f(z)dz = \sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n\gamma(\alpha)} - \sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n\alpha}$$
$$= ? - ?$$

We have

```
alpha = I
gamma_alpha = (-21*I - 1)/(43*I + 2); gamma_alpha
```

$\frac{1}{1853}i - \frac{905}{1853}$

We have

$$\sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n\alpha} = \sum_{n=1}^{\infty} \frac{a_n}{n} e^{-2\pi i n} \sim 0.0018639510538063274,$$

which we compute as

```
sum(float(an[n]/n * exp(-2*pi*n)) for n in [1..100])
```

Since $e^{-2\pi 100} \sim 10^{-273}$ the above approximation is probably quite accurate. Likewise, we compute with $\gamma(\alpha)$:

$$\sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi i n\gamma(\alpha)} = \sum_{n=1}^{\infty} \frac{a_n}{n} e^{2\pi n\cdot\left(-\frac{905}{1853}i - \frac{1}{1853}\right)} \sim \text{useless},$$

which is useless because this is so big:

```
N(exp(2*pi*I*gamma_alpha*100))
```

$0.380674761646642 + 0.602191787477105i$

Trying something bigger is useful:

```
N(exp(2*pi*I*gamma_alpha*2000))
```

$0.000312297661686191 + 0.00109058443661966i$

So we brute force through and get this approximation for the other integral:

$$0.00213843650496767 - 0.000422376254514594i$$

Thus

$$\int_0^{\gamma(0)} 2\pi i f(z)dz \sim (0.00213843650496767 - 0.000422376254514594i) - 0.0018639510538063274 = 0.00027$$

The actual elliptic curve period lattice is

```
EllipticCurve('43a').period_lattice().basis()
```

$$(5.46868952996758, 2.73434476498379 + 1.36318241817043i)$$

The integral we compute must be a $\mathbb{Z}$ linear combination of this basis, so we just computed 0. Indeed, this is what the algebra shows:

```
phi = M.cuspidal_submodule().decomposition()[0].\
   rational_period_mapping()
phi(M([0,-1/2]))
```

```
(0, 0)
```

5.7.5 Remarks on finding all curves isogenous to E

Basically, see that ANTS paper.

Chapter 6

Prime Numbers

Main unsolved problem: the Riemann Hypothesis

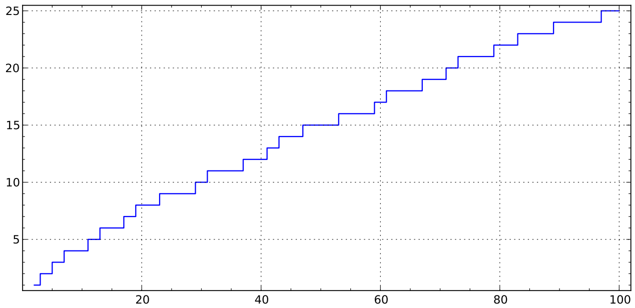
6.1 The Riemann hypothesis

The prime numbers $2, 3, 5, 7, 11, \ldots,$ have fascinated mathematicians for thousands of years. Euclid proved there are infinitely many: if $p_1, \ldots, p_n$ are primes, then $p_1 \cdots p_n + 1$ is an integer divisible by some prime p that isn't equal to any p_i.

Let's compute the primes up to 100:

```
sage: prime_range(100)
[2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, \
   47, 53, 59, 61, 67, 71,
 73, 79, 83, 89, 97]
```

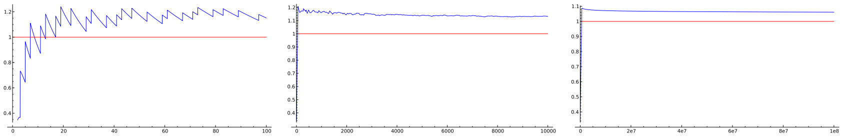
And draw a plot of the function $\pi(x)$ that counts the number of primes up to x for $x < 100$.



The *Prime Number Theorem*, which was proved over a century ago, asserts that $\pi(x) \sim x/\log(x)$:

```
@interact
```

```
def f(B=[10^n for n in [2..9]]):
    show(plot(lambda x: prime_pi(x)/(x/log(x)), (x,2,\
       B))
       + line([(0,1),(B,1)],color='red'))
```



The *Riemann Hypothesis*, which remains completely unsolved today, asserts that for all $x \geq 2.01$,

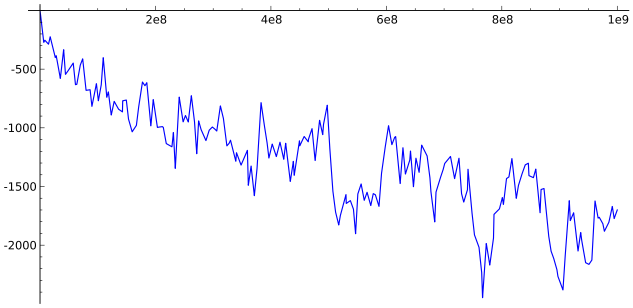
$$|\pi(x) - \mathrm{Li}(x)| \leq \sqrt{x} \cdot \log(x),$$

where

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log(t)}$$

In the range of the following plots, $\pi(x) - \mathrm{Li}(x)$ is much smaller than $\sqrt{x}\log(x)$.

```
@interact
def f(B=[10^n for n in [2..9]]):
    print "sqrt(B)*log(B) = ", round(sqrt(B)*log(B))
    show(plot(lambda x: prime_pi(x) - Li(x), (x,2,B))\
       )
```



Up to 10^9 we have $|\pi(x) - \mathrm{Li}(x)|$ is much smaller than $\sqrt{10^9} \cdot \log(10^9) \sim 655{,}327$.

HW (Andrew and Bharath): *Can we prove anything toward* $\sqrt{x} \cdot \log(x)$*?*

One can follow a naive approach and use basic calculus. One sees easily that $\pi(x) \leq x$. Similarly, if $x \geq 2$, then $\log(x) \geq \log(2)$. So, we have that $\mathrm{Li}(x) \leq \frac{x}{log(2)}$. Using triangular inequality, one can deduce that $|\pi(x) - \mathrm{Li}(x)| \leq x \cdot (1 + \frac{1}{\log(2)}) \leq 3x$.

Chapter 7

Birch and Swinnerton-Dyer

This chapter is about the Birch and Swinnerton-Dyer conjecture, which ties together the invariants of an elliptic curve.

7.1 Statement of the Conjecture

Let E be an elliptic curve over $\mathbb{Q}$ with L-series

$$L(E,s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

which we know by Theorem 5.4.1 extends to an entire function on $\mathbb{C}$. Let $L^*(E,1)$ denote the leading coefficient of the Taylor series of $L(E,s)$ about $s=1$. Also, let

$$r_{\text{an}} = r_{\text{an}}(E) = \operatorname{ord}_{s=1} L(E,s) \quad \text{and} \quad r_{\text{alg}} = r_{\text{alg}}(E) = \operatorname{rank}(E(\mathbb{Q})).$$

That $E(\mathbb{Q})$ is *finitely generated* is a nontrivial theorem of Mordell. The L-series of E is related to the invariants of E by the following conjecture, whose formulation by Bryan Birch and Sir Peter Swinnerton-Dyer in the 1960s took an enormous amount of computer computation and theoretical imagination.

Conjecture 7.1.1 (Birch and Swinnerton-Dyer)**.** *We have* $r_{\text{alg}} = r_{\text{an}}$ *and*

$$L^*(E,1) = \frac{\Omega_E \cdot \prod_{p\mid N} c_p \cdot \operatorname{Reg}_E \cdot \#\text{Ш}(E)}{\#E(\mathbb{Q})_{\text{tor}}^2}.$$

In the conjecture, Ω_E is the least real period, or twice the least real period if the period lattice (which we encountered in Section 5.7.2) is rectangular. The Tamagawa numbers

$$c_p = \#\Phi_{E,p}(\mathbb{F}_p)$$

are the order of the group of $\mathbb{F}_p$-rational points in the component of group of E at p, which we encountered in Section 5.3.5 when defining the conductor. The torsion subgroup $E(\mathbb{Q})_{\mathrm{tor}}$ in the denominator is simply the subgroup of elements of finite order in the abelian group $E(\mathbb{Q})$; a theorem of Mazur [Maz77] classifies the 16 possibilities for $E(\mathbb{Q})$.

The real number Reg_E is the absolute value of the discriminant of the Néron-Tate canonical height pairing matrix on a basis for $E(\mathbb{Q})/\operatorname{tor}$. Specifically, if $P \in E(\mathbb{Q})$, we define the (Néron-Tate) canonical height $h(P)$ via a naïve logarithmic height function on points in $E(\mathbb{Q})$:

$$h(P) = \lim_{n\to\infty} \frac{\widehat{h}(nP)}{n^2}$$

There are many choices of naïve logarithmic height functions that result in the same canonical height function (hence its name); one of the simplest that works is taking $\widehat{h}(Q)$ to be the natural logarithm of the denominator of the x-coordinate of Q (see Advanced Silverman, chapter 6 for a more complete discussion of height functions). The canonical height function defines a quadratic form on points in $E(\mathbb{Q})$:

$$\langle P, Q\rangle := \frac{h(P+Q) - h(P) - h(Q)}{2}$$

The regulator Reg_E is the absolute value of the matrix $(\langle P, Q\rangle)$, where P and Q range over a basis of generators of the free part of $E(\mathbb{Q})$.

Finally, the Shafarevich-Tate group $Ш(E)$, pronounced "Sha" for the first letter of Shafarevich's name, is the kernel

$$\begin{aligned} Ш(E) &= \ker\left(H^1(\mathbb{Q}, E) \to \prod_{p\leq\infty} H^1(\mathbb{Q}_p, E)\right), \\ &= \Big\{ \ (X/\mathbb{Q}, \iota) \text{ with } E\times X \xrightarrow{\iota} X \text{ and } X(\mathbb{Q}_p)\neq\emptyset \text{ all } p\leq\infty\Big\} / \sim, \end{aligned}$$

where ι is a simply transitive action and $\mathbb{Q}_\infty = \mathbb{R}$. Thus $Ш(E)$ is an abelian group, and $\#Ш(E)$ is its order.

Example 7.1.2. Consider the elliptic curve E defined by the projective equation $x^3 + y^3 + 60z^3$. Then the curve $3x^3 + 4y^3 + 5z^3 = 0$ defines an element of $Ш(E)$, as explained in Mazur's *On the passage from local to global in number theory.*

7.2 Some of what is and isn't known

In this section we give a flavor of some of what is and is not known about Conjecture 7.1.1. Any real progress at all on any of the challenges listed below as "Open Problems" would be extremely exciting, because it seems that the mathematically community is completely stumped by all of them right now, and the best advice from the old masters is:

"A new idea is needed." – Nick Katz, when I asked him about BSD.

7.2.1 The Rank

When $r_{\rm an} \leq 1$, work of Gross-Zagier on Heegner points (and some analytic work on L-functions by Murty-Murty or Waldspurger or Bump-Friedberg-Hoffstein, and of course Wiles et al.) combined with Kolyvagin's theorems about Euler systems yields extensive results. In particular, we have

Theorem 7.2.1 (mainly Kolyvagin, Gross-Zagier)**.** *Suppose that E is an elliptic curve over $\mathbb{Q}$ and $r_{\rm an} \leq 1$. Then $r_{\rm an}(E) = r_{\rm alg}(E)$.*

We can also computationally construct many specific examples of curves E of rank 2 or 3 such that $r_{\rm an}(E) = r_{\rm alg}(E)$. However, beyond rank 3 we are completely ignorant, in the following sense:

Open Problem 7.2.2. Prove that there is at least one elliptic curve over $\mathbb{Q}$ such that $r_{\rm an}(E) \geq 4$.

The smallest conductor of a rank 4 curve is 234446, and the curve is

$$y^2 + xy = x^3 - x^2 - 79x + 289.$$

7.2.2 The Shafarevich-Tate group Ш

The group $Ш(E/\mathbb{Q})$ is a subgroup of $H^1(\mathbb{Q}, E)$, which is an enormous torsion group; in fact, one can prove that it has infinitely many elements of all order (see [LT58, Ste02])).

An implicit assertion of Conjecture 7.1.1 is that the group $Ш(E/\mathbb{Q})$ is finite. In fact, due to the work of Kolyvagin and Gross-Zagier mentioned above, we know this in some generality:

Theorem 7.2.3 (Kolyvagin, Gross-Zagier)**.** *If E is an elliptic curve over $\mathbb{Q}$ with $r_{\rm an} \leq 1$ then $Ш(E/\mathbb{Q})$ is a finite group.*

Not only is $Ш(E/\mathbb{Q})$ finite when $r_{\rm an} \leq 1$, but there is an algorithm to compute $Ш(E/\mathbb{Q})$ when $r_{\rm an} \leq 1$, which has been made *pratical* in many cases (see Section 7.2.4). However, our knowledge stops the moment we discard the hypothesis that $r_{\rm an} \leq 1$.

Open Problem 7.2.4. Prove that there is an elliptic curve E over $\mathbb{Q}$ with $r_{\rm an} \geq 2$ such that $Ш(E/\mathbb{Q})$ is finite.

In an attempt to computationally convince myself that $Ш(E/\mathbb{Q})$ is finite for the curve E with label 389a of rank 2, I verified that $Ш(E/\mathbb{Q})[p] = 0$ for all but 19 primes $p \leq 48859$ (see [SW12]), the 19 primes are excluded because they are either bad, supersingular, or a p-adic L-function vanishes to higher order making certain computations more difficult.

7.2.3 The Analytic order of Ш

For an elliptic curve E, let $Ш_{\mathrm{an}}(E)$ denote order of $Ш(E)$ that is predicted by Conjecture 7.1.1, so

$$Ш_{\mathrm{an}}(E) = \frac{L^*(E,1)\cdot \#E(\mathbb{Q})_{\mathrm{tor}}^2}{\Omega_E \cdot \prod_{p\mid N} c_p \cdot \mathrm{Reg}_E} \in \mathbb{R}.$$

For curves of analytic rank at most 1, we know a lot about the real number $Ш_{\mathrm{an}}(E)$ due to work of Birch and Gross-Zagier, which is reflected in the following theorem:

Theorem 7.2.5 (Birch, Gross-Zagier)**.** *If E is an elliptic curve with $r_{\mathrm{an}} \leq 1$, then $Ш_{\mathrm{an}}(E) \in \mathbb{Q}$.*

Birch proved the above theorem when $r_{\mathrm{an}} = 0$ using modular symbols, and when $r_{\mathrm{an}} = 1$ it is a consequence of the Gross-Zagier formula.

Open Problem 7.2.6. Prove that there is some elliptic curve E of rank at least 2 such that $Ш_{\mathrm{an}}(E)$ is not transcendental.

7.2.4 The Full conjecture

Combining everything that is known with very extensive machine computations, Robert Miller, ..., and Stein, proved the following theorem:

Theorem 7.2.7. *Suppose E is an elliptic curve over $\mathbb{Q}$ with conductor $N \leq 5000$ and the rank of E is either 0 or 1. Then Conjecture 7.1.1 is true for E.*

In theory one could verify that $r_{\mathrm{an}}(E) = r_{\mathrm{alg}}(E)$ for all curves of conductor less than 234446, since Cremona has enumerated all these curves and $r_{\mathrm{alg}}(E) \leq 3$ for these curves. The difficult part would be verifying that for each curve with algebraic rank 3 the Heegner point is torsion, which by the Gross-Zagier formula would imply that $L'(E,1) = 0$. Depending on how challenging this is, it might make a good short paper (or not – it really depends on the difficulty).

7.3 Computing with some examples

News flash: BSD is false!

```
From: Jennifer S Balakrishnan xxx
Date: Wed, Dec 4, 2013 at 12:01 PM
Subject: Magma elliptic curve rank bug
To: Mark Watkins xxx

Hi Mark,

Magma seems to think that this elliptic curve has \
   algebraic rank 1
```

```
(and analytic rank 2!):

E:=EllipticCurve([0, 1, 0, -39, -83]);
Rank(E);
1
P := Generators(E)[1];
P;
(-3 : 4 : 1)
AnalyticRank(E);
2 8.0466
```

Just kidding—in fact this is a bug in Magma.

```
E = EllipticCurve([0, 1, 0, -39, -83])
E.rank()
E.conductor().factor()
```

$$2$$

$$2^8 \cdot 59^2$$

```
E.conductor()
```

$$891136$$

7.3.1 The curve 11a of rank 0

```
E = EllipticCurve('11a'); E
```

$$y^2 + y = x^3 - x^2 - 10x - 20$$

```
L = E.lseries(); print(L)
Complex L-series of the Elliptic Curve defined by
y^2 + y = x^3 - x^2 - 10*x - 20 over Rational Field
```

```
L(2+5*I)
```

$$1.49538417460974 - 0.434431555743826i$$

```
L(1)
```

$$0.253841860855911$$

```
L.taylor_series(1, 5)      # about 1 so z=(s-1)
```

$$0.25 + 0.31z + 0.011z^2 - 0.037z^3 + 0.0088z^4 + 0.00076z^5 + O(z^6)$$

```
E.rank()
```

$$0$$

```
lstar = L(1); lstar
```

$$0.253841860855911$$

```
L(1+I)
```

$$0.252329844312245 + 0.345912342362107i$$

```
E.period_lattice().basis()
```

$$(1.26920930427955, 0.634604652139777 + 1.45881661693850i)$$

```
omega = E.period_lattice().omega(); omega
```

$$1.26920930427955$$

```
c11 = E.tamagawa_number(11); c11
```

$$5$$

```
Reg = E.regulator();  Reg   # exactly one, since rank\
    = 0
```

$$1.00000000000000$$

```
T = E.torsion_order(); T
```

$$5$$

```
sha = E.sha(); print(sha)
```

```
Tate-Shafarevich group for the Elliptic Curve defined by y^2 + y = x^3 - x^2 - 10*x - 2
over Rational Field
```

```
sha.bound()   # means only 2,3,5 can divide order
```

$$[2, 3, 5]$$

```
sha.an()     # conjectural order if you solve for it \
   in BSD.
```

$$1$$

```
S = 1
```

```
lstar
```

$$0.253841860855911$$

```
(omega * c11 * Reg * S) / T^2
```

$$0.253841860855911$$

7.3.2 The curve 37a of rank 1

```
E = EllipticCurve('37a'); E
```

$$y^2 + y = x^3 - x$$

```
E.rank()
```

$$1$$

```
L = E.lseries(); print(L)
```

```
Complex L-series of the Elliptic Curve defined by y^2 + y = x^3 - x over Rational Field
```

```
L.taylor_series(1,100,10)  # r_an = 1
```

$0.00000000000000000000000000000 + 0.30599977383405230182048368332z + 0.18654779726816196417381736878z^2 - 0.1$

```
E.analytic_rank()
```

$$1$$

```
E.period_lattice().basis()   # rectangular period \
   lattice
```

$$(2.99345864623196, 2.45138938198679i)$$

```
omega = E.period_lattice().omega(); omega
```

$$5.98691729246392$$

```
c37 = E.tamagawa_number(37); c37
```

$$1$$

```
E.gens()  # rank 1 mw group
```

$$[(0 : 0 : 1)]$$

```
Reg = E.regulator(); Reg
```

$$0.0511114082399688$$

```
E.gens()[0].height()
```

$$0.0511114082399688$$

```
sha = E.sha().an(); sha   # trivial sha
```

$$1$$

```
S = 1
```

```
T = E.torsion_order(); T
```

$$1$$

```
lstar = L.dokchitser().derivative(1,1); lstar    # \
    first derivative at 1
```

$$0.305999773834052$$

```
(omega * c37 * Reg * S) / T^2
```

$$0.305999773834052$$

7.3.3 The curve 389a of rank 2

```
E = EllipticCurve('389a'); E
```

$$y^2 + y = x^3 + x^2 - 2x$$

```
E.rank()
```

$$2$$

```
L = E.lseries(); print(L)
Complex L-series of the Elliptic Curve defined by y^2 + y = x^3 + x^2 - 2*x over Ratio
Field
```

```
L.taylor_series(1,53)   # r_an = 2,  z=(s-1)
```

$$-2.69129566562797\times10^{-23}+\left(1.52514901968783 \times 10^{-23}\right) z+0.759316500288427z^2-0.430302337583362z$$

```
E.analytic_rank()
```

$$2$$

```
E.period_lattice().basis()    # rectangular period \
    lattice
```

$$(2.49021256085505, 1.97173770155165i)$$

```
omega = E.period_lattice().omega(); omega
```

$$4.98042512171011$$

```
c = E.tamagawa_number(389); c
```

$$1$$

```
E.gens()  # rank 2 mw group
```

$$[(-1 : 1 : 1), (0 : 0 : 1)]$$

```
Reg = E.regulator(); Reg
```

0.152460177943144

```
sha = E.sha().an(); sha    # trivial sha (?) -- we don\
   't know!  we don't know this is even finite!
```

1.00000000000000

```
S = 1
```

```
T = E.torsion_order(); T
```

1

```
lstar = L.dokchitser().derivative(1,2)/2; lstar    # \
   second derivative at 1, divided by 2
```

0.759316500288427

```
(omega * c * Reg * S) / T^2
```

0.759316500288426

7.3.4 The curve 5077a of rank 3

```
E = EllipticCurve('5077a'); E
```

$$y^2 + y = x^3 - 7x + 6$$

```
E.rank()
```

3

```
L = E.lseries(); print(L)
```

```
Complex L-series of the Elliptic Curve defined by y^2 + y = x^3 - 7*x + 6 over Rational
Field
```

```
L.taylor_series(1,5)   # r_an = 3
```

$$0.00 + \left(2.2 \times 10^{-7}\right) z + \left(-4.0 \times 10^{-7}\right) z^2 + 1.8z^3 - 3.2z^4 + 2.8z^5 + O(z^6)$$

```
E.analytic_rank()
```

3

```
E.period_lattice().basis()   # rectangular period \
    lattice
```

$$(2.07584399154347, 1.48054826824141i)$$

```
omega = E.period_lattice().omega(); omega
```

$$4.15168798308693$$

```
c = E.tamagawa_number(5077); c
```

$$1$$

```
E.gens()  # rank 2 mw group
```

$$[(-2:3:1), (-1:3:1), (0:2:1)]$$

```
Reg = E.regulator(); Reg
```

$$0.417143558758384$$

```
sha=1  # trivial sha (?) -- we don't know!  we don't \
    know this is even finite!
```

```
T = E.torsion_order(); T
```

$$1$$

```
lstar = L.dokchitser().derivative(1,3)/factorial(3); \
    lstar   # second derivative at 1, divided by 3!
```

$$1.73184990011930$$

```
(omega * c * Reg * S) / T^2
```

$$1.73184990011930$$

7.3.5 The rank 0 curve 681b with Sha of order 9

```
factor(681)
```

$$3 \cdot 227$$

```
E = EllipticCurve('681b'); E
```

$$y^2 + xy = x^3 + x^2 - 1154x - 15345$$

```
E.rank()
```

$$0$$

```
L = E.lseries(); print(L)
```

```
Complex L-series of the Elliptic Curve defined by y^2 + x*y = x^3 + x^2 - 1154*x - 15345
over Rational Field
```

```
L.taylor_series(1)  # r_an = 0
```

$$1.84481520612682-1.56198367532152z+1.27184630002575z^{2}+0.0355857020468455z^{3}-1.23910487517717z^{4}+1.561717$$

```
E.analytic_rank()
```

$$0$$

```
E.period_lattice().basis()   # rectangular period \
   lattice, yet again!
```

$$(0.409958934694849, 0.712395770891159i)$$

```
omega = E.period_lattice().omega(); omega
```

$$0.819917869389698$$

```
c3 = E.tamagawa_number(3); c3
c227 = E.tamagawa_number(227); c227
```

$$2$$

$$2$$

```
Reg = 1
```

```
sha = E.sha().an(); sha   # of order 9.
S = 9
```

$$9$$

```
T = E.torsion_order(); T
```

$$4$$

```
lstar = L(1); lstar
```

$$1.84481520612682$$

```
(omega * c3 * c227 * Reg * S) / T^2
```

$$1.84481520612682$$

```
E.sha().bound()
```

$$[2, 3]$$

```
EllipticCurve('681c').rank()
```

$$2$$

```
Visibility of Sha.
```

Bibliography

[BMSW07] Baur Bektemirov, Barry Mazur, William Stein, and Mark Watkins, *Average ranks of elliptic curves: tension between data and conjecture*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 2, 233–254 (electronic). MR 2291676

[BS66] Z.I. Borevich and I.R. Shafarevich, *Number theory*, Pure and applied mathematics, Academic Press, New York, 1966.

[CL84] H. Cohen and H. W. Lenstra, Jr., *Heuristics on class groups of number fields*, Number theory, Noordwijkerhout 1983 (Noordwijkerhout, 1983), Lecture Notes in Math., vol. 1068, Springer, Berlin, 1984, pp. 33–62. MR 756082 (85j:11144)

[Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997, `http://www.warwick.ac.uk/~masgaj/book/fulltext/`.

[DD06] Jacques Dubrois and Jean-Guillaume Dumas, *Efficient polynomial time algorithms computing industrial-strength primitive roots*, Information processing letters **97** (2006), no. 2, 41–45.

[Elk91] Noam D Elkies, *Abc implies mordell*, International Mathematics Research Notices **1991** (1991), no. 7, 99–109.

[GM84] Rajiv Gupta and M Ram Murty, *A remark on artin's conjecture*, Inventiones mathematicae **78** (1984), no. 1, 127–130.

[HB86] DR Heath-Brown, *Artin's conjecture for primitive roots*, The Quarterly Journal of Mathematics **37** (1986), no. 1, 27–38.

[Hec36] Erich Hecke, *Über die bestimmung dirichletscher reihen durch ihre funktionalgleichung*, Mathematische Annalen **112** (1936), no. 1, 664–699.

[Hec59] ______, *Mathematische werke: herausgegeben im auftrage der akademie der wissenschaften zu göttingen*, Vandenhoeck und Ruprecht, 1959.

[Hee52] Kurt Heegner, *Diophantische analysis und modulfunktionen*, Mathematische Zeitschrift **56** (1952), no. 3, 227–253.

[Ich98] Humio Ichimura, *A note on greenbergs conjecture and the abc conjecture*, Proceedings of the American Mathematical Society **126** (1998), no. 5, 1315–1320.

[LT58] S. Lang and J. Tate, *Principal homogeneous spaces over abelian varieties*, Amer. J. Math. **80** (1958), 659–684.

[Maz77] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. (1977), no. 47, 33–186 (1978), `http://archive.numdam.org/article/PMIHES_1977__47__33_0.pdf`.

[RW03] Herman te Riele and Hugh Williams, *New computations concerning the cohen-lenstra heuristics*, Experimental Mathematics **12** (2003), no. 1, 99–113.

[Sho92] Victor Shoup, *Searching for primitive roots in finite fields*, Mathematics of Computation **58** (1992), no. 197, 369–380.

[Sta69] Harold M Stark, *On the gap in a theorem of heegner*, Journal of Number Theory **1** (1969), no. 1, 16–27.

[Ste02] W. A. Stein, *There are genus one curves over* $\mathbf{Q}$ *of every odd index*, J. Reine Angew. Math. **547** (2002), 139–147. MR 2003c:11059

[Ste07] William Stein, *Modular Forms, A Computational Approach*, Graduate Studies in Mathematics, vol. 79, American Mathematical Society, Providence, RI, 2007, With an appendix by Paul E. Gunnells, `http://wstein.org/books/modform/`. MR 2289048

[Ste12] ______, *Algebriac Number Ttheory: a computational approach*, 2012, `http://wstein.org/books/ant/`.

[SW02] William Stein and Mark Watkins, *A database of elliptic curves—first report*, Algorithmic number theory (Sydney, 2002), Lecture Notes in Comput. Sci., vol. 2369, Springer, Berlin, 2002, `http://wstein.org/ecdb`, pp. 267–275. MR 2041090 (2005h:11113)

[SW12] William Stein and Christian Wuthrich, *Computations About Tate-Shafarevich Groups Using Iwasawa Theory*, Mathematics of Computation (2012), `http://wstein.org/papers/shark/`.

[Was97] Lawrence C. Washington, *Introduction to cyclotomic fields*, second ed., Graduate Texts in Mathematics, vol. 83, Springer-Verlag, New York, 1997. MR 1421575 (97h:11130)

[Wei67] André Weil, *Über die bestimmung dirichletscher reihen durch funktionalgleichungen*, Mathematische Annalen **168** (1967), no. 1, 149–156.

[Wet97] Joseph Loebach Wetherell, *Bounding the number of rational points on certain curves of high rank*, Ph.D. thesis, University of California, Berkeley, 1997.

[Wil95] A. J. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551, `http://users.tpg.com.au/nanahcub/flt.pdf`.