Counting arithmetic objects

Notes taken by Daniel Miller

June 23–July 4, 2014

Contents

Editor's note

This is a private set of notes I took at the 2014 summer school "Counting Arithmetic Objects," at the Centre de Recherches Mathématiques Montreal. The summer school's homepage is http://www.crm.umontreal.ca/sms/2014/index_e.php. I posted the notes online in case anyone wanted to recall some of the material before an official proceedings is published. In case it is not clear, *these notes do not have the*

official backing of any of the speakers at the summer school. They are *not* a perfectly accurate replica of the talks, and no doubt many errors were introduced in the process of typing them up. Any and all errors should be blamed on me, the typist.

1 Introduction and perspective

1.1 Motivation

The main question we are interested in is: given a class $\mathcal{C}$ of objects "of arithmetic interest," how many objects are there in $\mathcal{C}$, up to isomorphism, having bounded invariants?

Example 1.1.1. The following are the main examples we're interested in:

$\mathcal{C}$	invariant
number fields of given degree	discriminant
class group elements of number fields of given degree	"
rational points on curves	height
elliptic curves weighted by rank	"
n-Selmer elements of Jacobians of curves	"

All of these will be defined precisely later on.

Given such a class of objects of arithmetic interest, how they are distributed (asymptotically) with respect to their basic invariants? Beyond the cases of degree 2 number fields and genus 0 curves, little was known at the beginning of the 20th century.

1.2 Strategy

Direct methods of counting arithmetic objects generally fail except in the "easy" cases of degree 2 number fields and genus 0 curves. The modern approach uses representation theory. We try to find a map

$$\mathcal{C}/\simeq \hookrightarrow G(\mathbf{Z})\backslash V(\mathbf{Z}),$$

where G is an algebraic group and V is a representation of G, both defined over $\mathbf{Z}$. More precisely, one finds such a map that sends the invariants of objects in $\mathcal{C}$ to the ring of fundamental polynomial invariants of the action of G on V. Good choices of such maps often come from algebraic geometry, but we have to work out the theory over $\mathbf{Z}$.

Example 1.2.1 (Gauss)**.** In his *Disquisitiones*, Gauss constructs a map

$$\left\{\begin{array}{c}\text{ideal classes of (orders}\\ \text{in) quadratic fields with}\\ \text{non-square discriminant}\end{array}\right\}/\simeq \xrightarrow{\sim} \mathrm{SL}_2(\mathbf{Z})\backslash\left\{\begin{array}{c}\text{integer binary quadratic}\\ \text{forms } ax^2+bxy+cy^2\end{array}\right\}$$

The map sends the discriminant of such an ideal class to b^2-4ac (the discriminant of the quadratic form, which is the unique polynomial invariant of quadratic forms). If $I=\langle\alpha,\beta\rangle$, then the corresponding quadratic form is $\mathrm{N}(\alpha x+\beta x)/\mathrm{N}(I)$.

Example 1.2.2 (Levi, Delone-Faddeev)**.** Recall that a *cubic ring* is a (commutative, unital) ring structure on $\mathbf{Z}^3$. They constructed a map

$$\{\text{cubic rings}\}/\simeq \xrightarrow{\sim} \mathrm{GL}_2(\mathbf{Z})\backslash \left\{ \begin{array}{c} \text{integer binary cubic forms} \\ ax^3+bx^2y+cxy^2+dy^3 \end{array} \right\},$$

which "preserves discriminant." The discriminant of a binary cubic form is $bc^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$. If $R = \langle 1, \alpha, \beta \rangle$ is a cubic ring, the associated binary cubic form is $\sqrt{\mathrm{Disc}(\alpha x + \beta y)/\mathrm{Disc}(R)} = [R : \mathbf{Z}[\alpha x + \beta y]]$. See [DF64] for Delone-Faddeev's original approach. Bhargava's thesis [Bha01] and Wood's thesis [Woo09] develop things further. See section 2 for more details.

Example 1.2.3. There are similar maps

$$\{\text{quartic (resp. quintic) fields}\}/\sim \to \{\text{representations}\dots\}.$$

These are also discriminant-preserving (in all these cases, the discriminant on the right is the unique polynomial invariant). See section 10 for more details.

Example 1.2.4 (Birch, Swinnerton-Dyer)**.** There is a map

$$\left\{ \begin{array}{c} \sigma \in E(\mathbf{Q})/2 : E \text{ of the} \\ \text{form } y^2 = x^3 + Ax + B \end{array} \right\} \hookrightarrow \mathrm{GL}_2(\mathbf{Q})\backslash \left\{ \begin{array}{c} \text{integer binary} \\ \text{quartic forms} \end{array} \right\}.$$

The program `mwrank` created by Cremona uses this. In fact, this map factors through the 2-Selmer group $\mathrm{Sel}_2(E)$. Write $E_{A,B}$ for the elliptic curve $y^2 = x^3 + Ax + B$. This map sends $E_{A,B}$ to the fundamental invariants I, J (of degree 2, 3 respectively) of binary quartics. See section 19 and section 22 for more.

Example 1.2.5 (Cremona-Fisher-Stoll)**.** For $n \in \{3, 4, 5\}$, there is a map

$$\{\sigma \in E(\mathbf{Q})/n : E \text{ of the form } E_{A,B}\} \hookrightarrow G(\mathbf{Z})\backslash V(\mathbf{Z}),$$

where A, B are send to the fundamental invariants I, J on the right-hand side.

Example 1.2.6. There is a non-injective, but still useful map

$$\left\{ \begin{array}{c} \text{rational points on odd hyperelliptic} \\ \text{curves } y^2 = x^{2g+1} + a_1x^{2g} + \cdots + a_{2g+1} \end{array} \right\} \to \mathrm{SO}_{2g+1}(\mathbf{Z})\backslash \mathrm{Sym}^2(\mathbf{Z}^{2g+1}).$$

This sends the a_i to the invariants on the right-hand side. See section 24 and section 29 for more.

Example 1.2.7. There is a (non-injective) map

$$\left\{ \begin{array}{c} \text{rational points on even hyperelliptic} \\ \text{curves } z^2 = a_0x^{2g+2} + \cdots + a_{2g+2}y^{2g+2} \end{array} \right\} \to \text{product of SL's}\backslash \cdots,$$

sending the a_i to invariants. See section 28, section 23, section 17, and section 12 for more.

Once we've found our map from arithmetic objects to orbits, the question becomes: how many orbits of $G(\mathbf{Z})$ on $V(\mathbf{Z})$ are there having bounded invariants? Gauss worked this out for binary quadratic forms. Let $h(D)$ be the number of $\mathrm{SL}_2(\mathbf{Z})$-orbits of integer binary quadratic forms of discriminant D.

Theorem 1.2.8 (Gauss, Lipschitz, Mertens)**.**

$$\sum_{0<-D<X} h(D) \sim \frac{\pi}{18} X^{3/2}.$$

Proof. Gauss shows that every integer binary quadratic form $ax^2 + bxy + cy^2$ with $D = b^2 - 4ac < 0$ has a unique $\mathrm{SL}_2(\mathbf{Z})$-equivalent form satisfying $|b| < a \leqslant c$ or $0 < b = a \leqslant c$.

We apply geometry of numbers to the counting problem

$$\sum_{0<-D<X} h(D) \sim \#\{(a,b,c) : 0 < 4ac - b^2 < X \text{ and } |b| < a \leqslant c\}.$$

Gauss's conjecture tells us that the number of points on the right is asymptotically the volume of the region. Proving this is tricky! In fact, the conjecture is false in general. If we consider the region $R = \{(a,b,c) : 0 \leqslant 4ac - b^2 < X \text{ and } |b| \leqslant a \leqslant c\}$, then $R \cap \mathbf{Z}^3$ contains infinitely points. □

One can attack the lattice-point counting problem explicitly by writing the count as a triple sum $\sum_{a,b,c}$, approximating the sum by a triple integral, and keeping track of error terms. This is how Lipschitz and Mertens proved the result. Working through this is a good exercise. Davenport developed some general principles for bounded regions. He used the principle to reprove Gauss' count of binary quadratics, and extended the argument to a count of binary cubic forms. This requires knowing explicit inequalities for the region.

A third approach uses zeta functions (or more generally L-functions). Siegel first applied this to binary quaratic forms. Goldfeld-Hoffstein, Shintani, and Datskovsky extended these methods. See section 6, section 13, and section 17 for details.

There is a hybrid method: average over a compact continuum of fundamental domains. It doesn't need explicit inequalities, but still uses elementary geometry of numbers. The method does adapt to situations where there are more than one invariant. For examples, it works on all the above examples. In particular, it gives a count of quartic and quintic fields, boundedness of average rank of elliptic curves, and produces lots of hyperelliptic curves with few rational points. See section 19 and section 22 for details.

What if we replace $\mathbf{Q}$ with another base field, like a number field or function field? Over a function field, one can use algebro-geometric and topological methods. Boundedness of average rank was proved by de Jong. Ellenberg has proved many other results of this type. Also, the "hybrid method" works over an arbitrary global field, as is worked out in section 31.

2 Basics of binary quadratic forms and Gauss composition

An official Beamer version of these notes can be found online at http://www.crm.umontreal.ca/sms/2014/pdf/AlgTalkSlides.pdf. These notes are essentially a transcription of the official version.

2.1 Sums of two squares

We start with a very old theorem, but proving it using some basic geometry of numbers.

Theorem 2.1.1. *Any prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares.*

Proof. Since $p \equiv 1 \pmod 4$, there is $i \in \mathbf{Z}$ such that $i^2 \equiv -1 \pmod p$. Consider the set of integers

$$\{m + ni : 0 \leqslant m, n \leqslant \lfloor \sqrt{p} \rfloor\}.$$

This set has cardinality $(\lfloor \sqrt{p} \rfloor + 1)^2 > p$, so by the pigeonhole principle, two are congruent modulo p; say

$$m + ni \equiv M + Ni \pmod p$$

where $0 \leqslant m, n, M, N \leqslant \lfloor \sqrt{p} \rfloor$ and $(m, n) \neq (M, N)$. Let $r = m - M$ and $s = N - n$ so that

$$r \equiv is \pmod p$$

where $|r|, |s| \leqslant \lfloor \sqrt{p} \rfloor < \sqrt{p}$, and r and s are both nonzero. Now

$$r^2 + s^2 \equiv (is)^2 + s^2 = s^2(i^2 + 1) \equiv 0 \pmod p$$

and $0 < r^2 + s^2 < \sqrt{p}^2 + \sqrt{p}^2 = 2p$. The only multiple of p between 0 and $2p$ is p, hence $r^2 + s^2 = p$. □

It is also well-known that a product of sums of squares is a sum of squares, via the identity

$$(a^2 + b^2)(c^2 + e^2) = (ac + be)^2 + (ae - bc)^2.$$

This has the easy generalization

$$(a^2 + db^2)(c^2 + de^2) = (ac + dbe)^2 + d(ae - bc)^2.$$

Gauss's perspective on this was as follows. A *binary quadratic form* is a polynomial of the form $f(x, y) = ax^2 + bxy + cy^2$. If we choose $f(x, y) = x^2 + dy^2$, then

$$f(a, b)f(c, e) = f(ac + dbe, ae - bc).$$

The latter values in f, namely $ac + dbe$ and $ae - bc$, are bilinear forms in a, b, c, e. Does this generalize to other "multiplications"?

2.2 Pell's equation

Pell asked whether there are integer solutions x, y to

$$x^2 - dy^2 = 1.$$

Solutions always exist, and can be found using the continued fraction expansion of $\sqrt{d}$. This was certainly known by Brahmagupta in 628 A.D. and probably by Archimedes much earlier. To solve his "Cattle Problem," one needs to find a solution to

$$x^2 - 4729494y^2 = 1.$$

The smallest solution has about $2 \cdot 10^6$ digits!

Theorem 2.2.1. *Let $d \geqslant 2$ be a non-square integer. Then there exist $x, y \in \mathbf{Z}$ such that $y \neq 0$ and*

$$x^2 - dy^2 = 1.$$

If (x, y) is the smallest positive solution, then all others are given by

$$x_n + y_n\sqrt{d} = (x + y\sqrt{d})^n.$$

A more interesting problem is to look for solutions to $x^2 - dy^2 = \pm 4$. Especially with a "$-$," this is very difficult, though there has great recent progress by Fouvry and Kluners.

Theorem 2.2.2. *Any quadratic irrational real number has an eventually periodic continued fraction.*

Here are some examples:

d	expansion of $\sqrt{d}$
2	$[1, \overline{2}]$
3	$[1, \overline{1, 2}]$
5	$[2, \overline{4}]$
6	$[2, \overline{2, 4}]$
7	$[2, \overline{1, 1, 1, 4}]$
8	$[2, \overline{1, 4}]$
10	$[3, \overline{6}]$
11	$[3, \overline{3, 6}]$
12	$[3, \overline{2, 6}]$
13	$[3, \overline{1, 1, 1, 1, 6}]$

If p_k/q_k are the convergents for $\sqrt{d}$, then $p_{n-1}^2 - dq_{n-1}^2 = (-1)^n$. Here are some examples of how far out one has to go to obtain a solution to Pell's equation:

d	expansion of $\sqrt{d}$	Pell's equation
2	$[1,\overline{2}]$	$1^2-2\cdot 1^2=-1$
3	$[1,\overline{1,2}]$	$2^2-3\cdot 1^2=1$
6	$[2,\overline{2,4}]$	$5^2-6\cdot 2^2=1$
7	$[2,\overline{1,1,1,4}]$	$8^2-7\cdot 3^2=1$
13	$[3,\overline{1,1,1,1,6}]$	$18^2-13\cdot 5^2=-1$
19	$[4,\overline{2,1,3,1,2,8}]$	$170^2-19\cdot 39^2=1$
22	$[4,\overline{1,2,4,2,1,8}]$	$197^2-22\cdot 42^2=1$
31	$[5,\overline{1,1,3,5,3,1,1,10}]$	$1520^2-31\cdot 273^2=1$
43	$[6,\overline{1,1,3,1,5,1,3,1,1,12}]$	$3482^2-43\cdot 531^2=1$
46	$[6,\overline{1,3,1,1,2,6,2,1,1,3,1,12}]$	$24335^2-46\cdot 3588^2=1$
76	$[8,\overline{1,2,1,1,5,4,5,1,1,2,1,16}]$	$57799^2-76\cdot 6630^2=1$

It is clear that the period of the continued fraction expansion grows quite quickly. Here is a table of the first few largest periods:

n	equation
16	$2143295^2-94\cdot 221064^2=1$
16	$4620799^2-124\cdot 4149602=1$
16	$2588599^2-133\cdot 2244602=1$
18	$77563250^2-139\cdot 65788292=1$
20	$1728148040^2-151\cdot 140634693^2=1$
22	$1700902565^2-166\cdot 132015642^2=1$
26	$278354373650^2-211\cdot 19162705353^2=1$
26	$695359189925^2-214\cdot 47533775646^2=1$
26	$5883392537695^2-301\cdot 339113108232^2=1$
34	$2785589801443970^2-331\cdot 153109862634573^2=1$
37	$44042445696821418^2-421\cdot 2146497463530785^2=-1$
40	$840560915469529337752-526\cdot 36650197573242955322=1$
42	$1811243550616307861302-571\cdot 7579818350628982587^2=1$

The length of the continued fractions here are around $2\sqrt{d}$, and the size of the fundamental solutions is around $10^{\sqrt{d}}$. It is believed that the smallest solution is of size $C^{\sqrt{d}}$ for some C, but very little has been proved. Understanding the distribution of sizes of the smallest solutions to Pell's equation is an outstanding open question in number theory.

There is one last fact we need on Pell's equation. Let $\epsilon_d = x_1 + y_1\sqrt{d}$ be the smallest solution with $x_1, y_1 \in \mathbf{N}$ to

$$x^2 - dy^2 = 1.$$

If $x^2 - dy^2 = n$ and $x, y \geqslant 0$, put $\alpha = x + y\sqrt{d} > \sqrt{n}$. If $\sqrt{n}\epsilon_d^k \leqslant \alpha < \sqrt{n}\epsilon_d^{k+1}$, let $\beta = \alpha\epsilon_d^{-k} = u + v\sqrt{d}$. Then $\sqrt{n} \leqslant \beta < \sqrt{n}\epsilon_d$, with $u, v \geqslant 1$ and $u^2 - dv^2 = n$.

2.3 Binary quadratic forms

Let $f(x, y) = ax^2 + bxy + cy^2$ be a binary quadratic form. The basic question is: what integers are represented by f? Namely, for what N are there coprime m, n so that

$N = f(m,n)$? We may assume $\gcd(a,b,c) = 1$. Completing the square, we obtain

$$4aN = (2am + bn)^2 d n^2$$

where the discriminant $d = b^2 - 4ac$, so d is congruent to either 0 or 1 modulo 4. If $d < 0$, the right side can only take positive values... things are easier when $d > 0$. If $a > 0$, the form is positive definite.

The quadratic forms $x^2 + y^2$ and $X^2 + 2XY + 2Y^2$ represent the same integers. If $N = n^2 + n^2$, then $N = (m-n)^2 + 2(m-n)n + 2n^2$. Conversely if $N = u^2 + 2uv + 2v^2$, then $N = (u+v)^2 + v^2$. More conceptually,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

transforms $x^2 + y^2$ into $X^2 + 2XY + 2Y^2$, and the transformation is invertible since $\det \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} = 1$.

Much more generally, define

$$\mathrm{SL}(2, \mathbf{Z}) = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} : \alpha, \beta, \gamma\delta \in \mathbf{Z} \text{ and } \alpha\delta - \beta\gamma = 1 \right\}.$$

Then $ax^2 + bxy + cy^2$ represents the same integers as $AX^2 + BXY + CY^2$ whenever $\begin{pmatrix} x \\ y \end{pmatrix} = \theta \begin{pmatrix} X \\ Y \end{pmatrix}$ with $\theta \in \mathrm{SL}(2, \mathbf{Z})$. When such a θ exists, we say the quadratic forms $ax^2 + bxy + cy^2$ and $AX^2 + BXY + CY^2$ are *equivalent*. Since this is an equivalence relation, it splits the set of binary quadratic forms into equivalence classes.

For a quadratic form $f(x,y) = ax^2 + bxy + cy^2$, we can write

$$ax^2 + bxy + cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

We define the *discriminant* of f to be

$$\mathrm{Disc}(f) = -\det \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} = \frac{1}{4}b^2 - ac.$$

If $g(X,Y) = AX^2 + BXY + CY^2$ is equivalent to $f(x,y) = ax^2 + bxy + cy^2$ via $\theta \in \mathrm{SL}(2, \mathbf{Z})$, we deduce that

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} = {}^{\mathrm{t}}\theta \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \theta,$$

so equivalent binary quadratic forms have the same discriminant.

Apparently complicated binary quadratic forms like $29X^2 + 82XY + 58Y^2$ can be equivalent to simple ones like $x^2 + y^2$.

Theorem 2.3.1 (Gauss)**.** *Every equivalence class of negative-definite binary quadratic forms contains a unique* reduced *representative, where a form f is reduced if*

$$-a < b \leqslant a \leqslant c, \text{ and } b \geqslant 0 \text{ whenever } a = c.$$

For a form f, put $d = \mathrm{Disc}(f)$. If $d < 0$ and f is reduced, we have $|d| = 4ac - |b|^2 \geqslant 3a^2$, hence

$$a \leqslant \sqrt{|d|/3}.$$

So for given $d < 0$, only finitely many a and b (since $|b| \leqslant a$), and hence $c = (b^2 - d)/4$ can exist with discriminant a. Let $h(d)$ be the (finite) number of equivalence classes of binary quadratic forms of discriminant d. The *class number* $h(d) \geqslant 1$ since we always have the *principal form*

$$\begin{aligned} x^2 - (d/4)y & \qquad d \equiv 0 \pmod 4 \\ x^2 + dy + \frac{1-d}{4}y^2 & \qquad d \equiv 1 \pmod 4. \end{aligned}$$

Theorem 2.3.2 (Gauss)**.** *Every positive definite binary quadratic form is equivalent to a reduced form.*

Proof. We proceed via "infinite descent" with the following algorithm.

1. If $c < a$ the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$ yields $(c, -b, a)$, which is equivalent to (a, b, c).

2. If $b > a$ or $b \leqslant -a$, let b' be the least residue, in absolute value, of $b \pmod{2a}$, so $-a < b' \leqslant a$; say $b' = b - 2ka$. Let $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -k \\ & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$. The resulting form (a, b', c') is equivalent to (a, b, c).

3. If $c = a$ and $-a < b < 0$, use the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & \\ 1 & \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$ to get the form $(a, -b, a)$. If the resulting form is not reduced, repeat.

The algorithm terminates after 3, since 2 is followed by 1 or 3, and since 1 reduces the size of a. □

If for example $f(x, y) = 76x^2 + 217xy + 155y^2$ of discriminant -31, the sequence of forms is

$$(76, 65, 14), (14, -65, 76), (14, -9, 2), (2, 9, 14), (2, 1, 4).$$

For the form $f(x, y) = 11x^2 + 49xy + 55y^2$ of discriminant -19, the sequence is

$$(11, 49, 55), (1, -5, 11), (1, 1, 5).$$

Proposition 2.3.3. *Suppose $d = b^2 - 4ac$ with $(, b, c) = 1$, and p a prime.*

1. *If $p = am^2 + bmn + cn^2$ for some integers m, n, then d is a sqare modulo $4p$.*
2. *If d is a square modulo $4p$, then there exists a binary quadratic form of discriminant d that represents p.*

Proof. 1. If $p \nmid 2ad$ and $p = am^2 + bmn + cn^2$, then $4ap = (2am + bn)^2 - dn^2$, so dn^2 is a square mod $4p$. Now $p \nmid n$ else $p \mid 4ap + dn^2 = (2am + bn)^2$, so that $p \mid 2am$ is impossible as $p \nmid 2a$ and $(m, n) = 1$. We deduce that d is a square mod p.

2. If $d \equiv b^2 \pmod{4p}$, then $d = b^2 - 4pc$ for some integer c, and so $px^2 + bxy + c^2$ is a quadratic form of discriminant d which represents p via $p = p \cdot 1^1 + b \cdot 1 \cdot 0 + c \cdot 0^2$. □

Theorem 2.3.4. *Suppose $h(d) = 1$. Then p is represented by a form of discriminant d if and only if d is a square modulo $4p$.*

We call the *fundamental discriminants* those such that if $q^2 \mid d$ then $q = 2$ and $d \equiv 8$ or $12 \pmod{16}$. The only fundamental $d < 0$ with $h(d) = 1$ are

$$-3, -4, -7, -8, -11, -19, -43, -67, -163.$$

This was proved originally by Heeger, then later by Baker and Stark.

Euler noticed that the polynomial x^2+x+41 is prime for $x = 0, \ldots, 39$, and found some similar polynomials. In fact, we have the following.

Theorem 2.3.5 (Rabinowiscz)**.** *The class number $h(1 - 4A) = 1$ for $A \geqslant 2$ if and only if $x^2 + x + A$ is prime for $x = 0, 1, \ldots, A - 2$.*

For some forms, whether or not a prime can be represented by that form can be detected by congruence conditions. For example:

form	condition for p to be rep. by f
$x^2 + y^2$	$(-1/p) = 1$
$x^2 + 2y^2$	$(-2/p) = 1$
$x^2 + xy + y^2$	$(-3/p) = 1$
$x^2 + xy + 2y^2$	$(-7/p) = 1$
$x^2 + xy + 3y^2$	$(-11/p) = 1$
$x^2 + xy + 5y^2$	$(-19/p) = 1$
$x^2 + xy + 11y^2$	$(-43/p) = 1$
$x^2 + xy + 17y^2$	$(-67/p) = 1$
$x^2 + xy + 41y^2$	$(-163/p) = 1$

When the class number is not one, sometimes you can still characterize representability of primes by congruence conditions. For example, $h(-20) = 2$, the two reduced forms are

$$x^2 + 5y^2, 2x^2 + 2xy + 3y^2.$$

A prime p is represented by $x^2 + 5y^2$ if and only if $p = 5$, or $p \equiv 1, 9 \pmod{20}$, and p is represented by $2x^2 + 2xy + 3y^2$ if and only if $p = 2$, or $p \equiv 3, 7 \pmod{20}$.

We cannot always distinguish which primes are represented by which binary quadratic form of discriminant d by congruence conditions. Euler found 65 such *idoneal numbers*. No more are known – at most one further idoneal number can exist, and only if GRH is false.

2.4 Structure of ideals

Any ideal I in a quadratic ring of integers

$$R = \{a + b\sqrt{d} : a, b \in \mathbf{Z}\}$$

can be generated by at most 2 elements. Indeed, if $I \subset \mathbf{Z}$, then I is principal. Otherwise there exists $r + s\sqrt{d} \in I$ with $s \neq 0$. Without loss of generality $s > 0$, and select such an element with s minimal.

2.5 Composition

2.6 Prehomogeneous vector spaces

2.7 Invariants of number fields

3 Algebraic groups, representation theory, and invariant theory

This lecture will consist mostly of a review of the basic terminology, as well as a little bit of Galois cohomology. An "official version" of the notes can be found online at http://www.math.mcgill.ca/goren/AlgberaicGroups.SMS2014.pdf.

3.1 Algebraic groups

For us, a *linear algebraic group* is a Zariski-closed subgroup of $\mathrm{GL}_N(\bar{k})$ for some integer $N \geqslant 1$, where k is a fixed field of characteristic zero, and $\bar{k}$ is the algebraic closure of k. Good general references for linear algebraic groups are the books [Bor91, GW09, Hum75, Spr09].

Example 3.1.1. The main example of a linear algebraic group is $G = \mathrm{GL}_N$. It contains several standard subgroups:

$$B = \begin{pmatrix} * & \cdots & * \\ & \ddots & \vdots \\ & & * \end{pmatrix} \qquad \text{"standard Borel"}$$

$$U = \begin{pmatrix} 1 & \cdots & * \\ & \ddots & \vdots \\ & & 1 \end{pmatrix} \qquad \text{"unipotent radical of } B\text{"}$$

$$T = \begin{pmatrix} * & & \\ & \ddots & \\ & & * \end{pmatrix} \qquad \text{"maximal torus"}$$

Example 3.1.2. Let q be a symmetric bilinear form corresponding to the matrix $(q_{ij})_{1 \leqslant i,j \leqslant N}$. Let

$$\mathrm{SO}(q) = \{g \in \mathrm{GL}_N : g q\, {}^{\mathrm{t}} g = q \text{ and } \det g = 1\}.$$

If q is the form $q(x_1, \ldots, x_{2n+1}) = \frac{1}{2}(x_1 x_{2n+1} + x_2 x_{2n} + \cdots + x_{n+1}^2$, then a maximal torus consists "anti-triangular" matrices with $(t_1, \ldots, t_n, 1, t_n^{-1}, \ldots, t_1^{-1})$.

An affine group G is determined by its coordinate ring $\bar{k}[G]$. The group operations $m : G \times G \to G$, $i : G \to G$, $e : 1 \to G$ correspond via the Yoneda lemma to $m^* : \bar{k}[G] \to \bar{k}[G] \otimes \bar{k}[G]$, $i^* : \bar{k}[G] \to \bar{k}[G]$, $e^* : \bar{k}[G] \to \bar{k}$. These give $\bar{k}[G]$ the structure of a Hopf algebra.

A *homomorphism* $f : G \to H$ of algebraic is a morphism of varieties that respects the group structures. It corresponds to a ring $f^* : \bar{k}[H] \to \bar{k}[G]$ that respects the comultiplication.

A *character* of G is a homomorphism $f : G \to \mathrm{GL}_1 = \mathbf{G}_\mathrm{m}$. This corresponds to a Hopf-algebra homomorphism $\phi : \bar{k}[t^{\pm 1}] \to \bar{k}[G]$. If we let f be the image of t under this map, then the fact that ϕ respects comultiplication tells us that $m^*(f) = f \otimes f$. We call such elements *grouplike*. Let $\mathrm{X}^*(G)$ be the group of characters of G; we have seen that $\mathrm{X}^*(G)$ is in bijection with the set of grouplike elements of $\bar{k}[G]$.

If $g \in G$, put $\operatorname{Int} g : G \to G$ for the action of g by inner automorphisms, i.e. $\operatorname{Int} g(x) = gxg^{-1}$. So we have a homomorphism $\operatorname{Int} : G \to \operatorname{Aut} G$.

3.2 Non-abelian cohomology

Let G be a topological group acting continuously on a discrete group M. Define

$$\mathrm{H}^0(G, M) = M^G = \{m \in M : gm = m \text{ for all } g \in G\}$$
$$\mathrm{H}^1(G, M) = \{\zeta : G \to M \text{ such that } \zeta(ab) = \zeta(a) \cdot a\zeta(b)\}/\sim$$

where $\zeta \sim \xi$ if there exists $m \in M$ such that $\zeta(a) = m^{-1}\xi(a) \cdot am$ for all $a \in G$. The set $\mathrm{H}^0(G, M)$ is naturally a group, but $\mathrm{H}^1(G, M)$ is only a pointed set. If $0 \to A \to B \to C \to 0$ is an exact sequence of G-groups, we get a long exact sequence

$$0 \to A^G \to B^G \to C^G \xrightarrow{\delta} \mathrm{H}^1(G, A) \to \mathrm{H}^1(G, B) \to \mathrm{H}^1(G, C) \to \mathrm{H}^2(G, A) \to \cdots$$

with the last map only existing if A is central in B. A good source for all of this is [Ser79].

We define δ directly. Given $c \in C^G$, lift c to $b \in B$, and let $\zeta = \delta(c)$ be $\zeta(g) = b^{-1} \cdot gb$. One can check that the class of ζ in $\mathrm{H}^1(G, A)$ is well-defined.

3.3 Forms

Suppose G is defined over k. A *k-form* of G is an algebraic group H over k together with an isomorphism $f : G_{\bar{k}} \xrightarrow{\sim} H_{\bar{k}}$. Let $\Gamma = \mathrm{Gal}(\bar{k}/k)$. For all $\sigma \in G$, we have $\sigma f : G_{\bar{k}} \xrightarrow{\sim} H_{\bar{k}}$. Then $f^{-1} \circ \sigma f \in \mathrm{Aut}_{\bar{k}}(G)$. An easy exercise in the definitions shows that this is a cocycle. In fact, we have

Theorem 3.3.1. *There is a natural isomorphism of pointed sets*

$$\{k\text{-forms of } G\}/\sim \xrightarrow{\sim} \mathrm{H}^1(\Gamma, \operatorname{Aut}_{\bar{k}} G).$$

If H corresponds to $\zeta : G \to M$, then $H(k) = G(\bar{k})^\Gamma$, where Γ now acts by $\tau \cdot g = \zeta(\tau)(\tau(g))$.

Example 3.3.2 (compact forms)**.** Let $\Gamma = \mathrm{Gal}(\mathbf{C}/\mathbf{R}) = \langle c \rangle$. Let $\theta(g) = {}^{\mathrm{t}}\bar{g}^{-1}$ be the *Cartan involution*. The cocycle ζ given by $\zeta(c) = \theta$ corresponds to the real form U_N of $\mathrm{GL}_N(\mathbf{C})$. It is defined by $\mathrm{U}_N(\mathbf{R}) = \{g \in \mathrm{GL}_N(\mathbf{C}) : \theta g = g\}$.

Theorem 3.3.3. *Any (connected) reductive algebraic group G over $\mathbf{R}$ has a unique compact form.*

All of the groups GL_n, SL_n, SO_n, Sp_{2n},... are reductive.

Example 3.3.4. If $G = \mathbf{G}_\mathrm{m}$, the compact form is $T(\mathbf{R}) = \{z \in \mathbf{C}^\times : z\bar{z} = 1\}$. We have

$$T \simeq \mathrm{SO}(2) = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} : a^2 + b^2 = 1 \right\},$$

via $\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mapsto a + bi$.

If M is a Γ-module, we often put $\mathrm{H}^i(k, M) = \mathrm{H}^i(\Gamma, M)$. Also, if G is an algebraic group defined over k, we put $\mathrm{H}^i(k, G) = \mathrm{H}^i(\Gamma, G(\bar{k}))$.

Example 3.3.5. Start with the exact sequence $1 \to \mathbf{G}_\mathrm{m} \to \mathrm{GL}_N \to \mathrm{PGL}_N \to 1$. The long exact sequence in cohomology is

$$1 \to k^\times \to \mathrm{GL}_N(k) \to \mathrm{PGL}_N(\bar{k})^\Gamma \to \mathrm{H}^1(k, \mathbf{G}_\mathrm{m}) \to \mathrm{H}^1(k, \mathrm{GL}_N) \to \mathrm{H}^2(k, \mathbf{G}_\mathrm{m}).$$

The famous *Hilbert Theorem 90* tells us that $\mathrm{H}^1(k, \mathrm{GL}_N) = 1$ for all $N \geqslant 1$, so we get

$$\mathrm{PGL}_N(\bar{k})^\Gamma = \mathrm{PGL}_N(k) = \mathrm{GL}_N(k)/\mathbf{G}_\mathrm{m}(k).$$

Moreover, $\mathrm{H}^1(k, \mathrm{PGL}_N) \hookrightarrow \mathrm{H}^2(k, \mathbf{G}_\mathrm{m})$. We call $\mathrm{Br}(k) = \mathrm{H}^2(k, \mathbf{G}_\mathrm{m})$ the *Brauer group* of k. Since $\mathrm{PGL}_N(\bar{k}) = \mathrm{Aut}_{\bar{k}\text{-}\mathsf{Alg}}(M_N(\bar{k}))$, we see that $\mathrm{H}^1(k, \mathrm{PGL}_N)$ classifies k-forms of $M_N(\bar{k})$, i.e. central simple algebras over K of rank N^2.

3.4 Jordan decomposition

Any $g \in \mathrm{GL}_N(\bar{k})$ has a unique decomposition $g = g_\mathrm{s} g_\mathrm{u}$, where g_s is simple (i.e. diagonalizable), g_u is unipotent, and $g_\mathrm{s} g_\mathrm{u} = g_\mathrm{u} g_\mathrm{s}$. One has $(g_\mathrm{u} - 1)^N = 0$. For example, in the two-dimensional case, a matrix $\begin{pmatrix} t_1 & u \\ & t_2 \end{pmatrix}$ is already diagonalizable if $t_1 \neq t_2$, or $u = 0$. If neither of those occur, we write it as

$$\begin{pmatrix} t & u \\ & t \end{pmatrix} = \begin{pmatrix} t & \\ & t \end{pmatrix} \begin{pmatrix} 1 & u/t \\ & 1 \end{pmatrix}.$$

The Jordan decomposition enjoys very strong rigidity properties. Namely, if $g \in G \subset \mathrm{GL}_N$, then also $g_\mathrm{s} \in G$ and $g_\mathrm{u} \in G$. If $f : G \to H$ is a homomorpism of algebraic groups, then we have $f(g_\mathrm{s}) = f(g)_\mathrm{s}$ and $f(g_\mathrm{u}) = f(g)_\mathrm{u}$.

3.5 Tori

A *torus* T is a form of $\mathbf{G}_\mathrm{m}^N$ for some N. Tori of rank N over k are classified by $\mathrm{H}^1(\Gamma, \mathrm{Aut}_{\bar{k}}(\mathbf{G}_\mathrm{m}^N)) = \hom(\Gamma, \mathrm{GL}_N(\mathbf{Z}))/\mathrm{conj}$.

If T is a orus, then its group of characters $\mathrm{X}^*(T_{\bar{k}}) \simeq \mathrm{X}^*(\mathbf{G}_\mathrm{m}^N) = \hom(\mathbf{G}_\mathrm{m}^N, \mathbf{G}_\mathrm{m}) = \mathbf{Z}^N$ has a continuous action of Γ, via $\sigma\chi(g) = \sigma(\chi(\sigma^{-1}(g)))$. Tori are completely classified by this action.

Theorem 3.5.1. *The functor* X^* *induces an anti-equivalence of categories*

$$\{\textit{tori over } k\} \xrightarrow{\sim} \{\textit{finite free } \mathbf{Z}\textit{-modules with continuous } \Gamma\textit{-action}\}.$$

A linear action of a torus T, namely $f : T \to \mathrm{GL}_N$, is simultaneously diagonalizable (every element of T is semi-simple). This follows from rigidity properties of the Jordan Decomposition.

All maximal tori in a linear algebraic group G are conjugate. The common dimension of these tori is called the *rank* of G, and written $\mathrm{rk}_{\bar{k}}(G)$.

Example 3.5.2. The standard torus of diagonal matrices $T \subset \mathrm{GL}_N$ is maximal, so $\mathrm{rk}(\mathrm{GL}_N) = N$.

3.6 Solvable groups

An algebraic group G is called *solvable* if it is solvable "in the usual sense." In other words, there exists a filtration $1 = G_0 \subset G_1 \subset \cdots \subset G_l = G$ such that each G_i a normal algebraic subgroup of G_{i+1}, and each G_{i+1}/G_i is abelian. The standard Borel $B \subset \mathrm{GL}_N$ is solvable. In fact, B is a maximal solvable subgroup.

Theorem 3.6.1 (Kolchin-Lie)**.** *If $G \subset \mathrm{GL}_N$ is solvable, then G can be conjugated into the standard Borel $B \subset \mathrm{GL}_N$.*

Proof. One uses the fact that if G acts on a projective space, then it has a fixed point. The standard representation $G \to \mathrm{GL}_N$ gives an action of G on $\mathbf{P}^{N-1}$, and a fixed point for G in $\mathbf{P}^{N-1}$ gives a line fixed by the action of G. □

Theorem 3.6.2 (Borel)**.** *If a solvable group G acts on a proper variety, then G has a fixed point.*

For a group G, let $\mathscr{R}(G)$ be the maximal connected normal solvable subgroup of G, and let $\mathscr{R}_{\mathrm{u}}(G)$ be the maximal connected normal unipotent subgroup of G. We say that G is *semisimple* if $\mathscr{R}(G) = 1$, and *reductive* if $\mathscr{R}_{\mathrm{u}}(G) = 1$. Clearly semisimple groups are reductive. The groups SL_n, SO_n, Sp_n are semisimple and GL_n, GSp_{2n}, GSpin_n are reductive.

For any G, the quotient $G/\mathscr{R}_{\mathrm{u}}(G)$ is reductive, and $G/\mathscr{R}(G)$ is semisimple. If G is reductive, then $G/Z(G)$ is semisimple. A *Levi subgroup* of a group G is a subgroup H such that $G = H \ltimes \mathscr{R}_{\mathrm{u}}(G)$. Such an H will be a maximal reductive subgroup of G.

A maximal connected solvable subgroup of G is called a *Borel subgroup*. The group B of upper-triangular matrices is a borel subgroup of $\mathrm{GL}(n)$. Every torus is contained in a Borel subgroup, and if G is a reductive group, then all Borel subgroups of G are conjugate.

A group G over $\mathbf{C}$ is reductive if and only if every representation $\rho : G \to \mathrm{GL}_N$ is semi-simple (a direct sum of irreducible representations). Alternatively, the ring $\mathbf{C}[\rho(G)]$ should be semi-simple.

Example 3.6.3. The group $\simeq U_1 = \begin{pmatrix} 1 & * \\ & 1 \end{pmatrix}$ is not reductive.

3.7 Parabolic subgroups

A subgroup P of a connected algebraic group G is called *parabolic* if the quotient G/P is projective. The basic theorem is that P is parabolic if and only if P contains

a Borel subgroup B. So in GL_N, a subgroup is parabolic if it contains a conjugate of the subgroup of upper-triangular matrices.

Example 3.7.1. Let k be an algebraically closed field. Recall that a *flag* in k^n is a collection of subspaces $F = (0 \subsetneq F_1 \subsetneq \cdots \subsetneq F_a = k^n)$. The *type* of F is $\boldsymbol{d} = (\dim F_i)_i$. The space of type $\boldsymbol{d}$ is a projective variety $\mathrm{Fl}_{\boldsymbol{d}}$ on which GL_n acts transitively. Let P be a stabilizer of a flag. Then $G/P \simeq \mathrm{Fl}_{\boldsymbol{d}}$ and P is parabolic. For example, if F_i is the span of $\{e_1, \ldots, e_{d_i}\}$, then

$$P = \left(1_{h_1}\right) [finish]$$

[finish]

4 Basic algebraic number theory

Essentially, this lecture will try to cover two semester-long courses (algebraic number theory and class field theory) in an hour. Hopefully, we'll focus on the theory of Hilbert class fields, and later on complex multiplication.

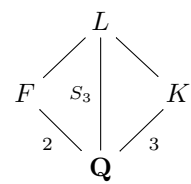
4.1 Number fields

A *number field* K is a finite field extension of $\mathbf{Q}$. An important invariant of K is its *class group*, $\mathrm{Cl}_K = I_K/P_K$, where I_K is the group of all fractional ideals in K and P_K is the group of principal fractional ideals.

Theorem 4.1.1. *The group* Cl_K *is finite.*

So we can define the *class number* h_K of K to be the cardinality of Cl_K. The class number $h_K = 1$ if and only if $\mathcal{O}_K$ is a principal ideal domain. The famous *Dirichlet unit theorem* says that $\mathcal{O}_K^\times$ is a finitely-generated abelian group, and gives a formula for $\operatorname{rk} \mathcal{O}_K^\times$ in terms of the number of real and complex places of K.

We would like to relate h_K with quadratic forms. Consider towers of fields



where L/F is an unramified (cubic) extension. Then counting the number of cubic fields K that are nowhere ramified with $|\operatorname{Disc} K| < X$ is equivalent to summing $\# h_3(F)$ for $|\operatorname{Disc} F| < X$. The average of $h_3(F)$ is $\frac{4}{3}$ for F real quadratic, and 2 for F imaginary quadratic. By the "average" we mean, for example,

$$\sum_{0 < |\operatorname{Disc} F| < X} h_3(F) \sim \frac{4}{3} X$$

in the case of real quadratic F. The point here is that averaging class numbers is equivalent to counting certain types of fields.

4.2 Starting point of class field theory

Class field theory is, in general, the study of abelian extensions of a field k. Let K be a number field, and let H/K be the maximal unramified abelian extension of K. One calls H the *Hilbert class field* of K. It is known that H is a number field. Moreover, the *Artin map* $I_K \to \mathrm{Gal}(H/K)$ determined by $\mathfrak{p} \mapsto \mathrm{fr}_\mathfrak{p} = (\mathfrak{p}, H/K)$ induces an isomorphism $\mathrm{Cl}_K \xrightarrow{\sim} \mathrm{Gal}(H/K)$. Recall that the *Frobenius element* $\mathrm{fr}_\mathfrak{p} \in \mathrm{Gal}(H/K)$ is characterized by $\mathrm{fr}_\mathfrak{p}(x) \equiv x^{N_{K/\mathbf{Q}}(\mathfrak{p})} \pmod{\mathfrak{P}}$ for all $\mathfrak{P} \mid \mathfrak{p}$ in H.

It is easy to show that $\mathfrak{p}$ splits compltely in H if and only if $\mathfrak{p}$ is principal. Somewhat harder is the principal ideal theorem:

Theorem 4.2.1. *Every ideal of K becomes principal in H.*

Proof. Let $G = \mathrm{Gal}(K^{\mathrm{ur}}/K)$, and let $G_1 \subset G$ correspond to the extension H/K. We have a commutative diagram:

$$\begin{array}{ccc} \mathrm{Cl}_H & \xrightarrow{\mathrm{Art}} & G_1^{\mathrm{ab}} \\ \uparrow & & \uparrow V \\ \mathrm{Cl}_K & \xrightarrow{\mathrm{Art}} & G^{\mathrm{ab}} \end{array}$$

where V is the *transfer map*. Since G_1 corresponds to H, $V = 0$, whence $\mathrm{Cl}_K \to \mathrm{Cl}_H$ is the trivial map. The result follows. □

If $K = \mathbf{Q}$, then $H = \mathbf{Q}$.

Assume from now on that $K = \mathbf{Q}(\sqrt{d})$ is an imaginary quadratic field ($d < 0$). There are nine possible values of $d < 0$ for which $h_{\mathbf{Q}(\sqrt{d})} = 1$, namely

$$-1, -2, -3, -7, -11, 19, -43, -67, -163.$$

It is known that $h_K \to \infty$ as $d \to -\infty$.

Example 4.2.2. Let $K = \mathbf{Q}(\sqrt{-23})$. Then H/K is cyclic of degree 3. It is known that $H = K(\alpha)$, where $\alpha^3 - \alpha + 1 = 0$.

4.3 Complex multiplication

One might ask if for any number field K, there is an explicit way of finding $\alpha \in \overline{\mathbf{Q}}$ such that $H = K(\alpha)$. The theory of complex multiplication describes how to do this explicitely whenever K is imaginary quadratic. The main theorem is the following:

Theorem 4.3.1. *Let E be an elliptic curve with CM by $\mathcal{O}_K$. Then $H = K(j(E))$.*

Recall that E has *complex multiplication* by $\mathcal{O}_K$ if $\mathrm{End}\, E \simeq \mathcal{O}_K$. If $E + y^2 = 4x^3 - g_2X - g_3$, then its *$j$-invariant* is given by

$$j(E) = 1728\frac{g_2^3}{g_2^3 - g_3^2}.$$

Alternatively, if $E = \mathbf{C}/\langle 1, \tau\rangle$, then $j(E) = \frac{1}{q} + 744 + 196884q^2 + \cdots$, where $q = e^{2\pi i\tau}$.

The rest of this lecture will be a sketch of a proof of this theorem.

4.4 Elliptic curves from fractional ideals

Let $\mathcal{E}_{\mathbf{C}}(K)$ be the set of isomorphism classes of elliptic curves over $\mathbf{C}$ with complex multiplication by $\mathcal{O}_K$. There is a bijection $\mathrm{Cl}_K \xrightarrow{\sim} \mathcal{E}_{\mathbf{C}}(K)$. Given $\mathfrak{a} \subset \mathcal{O}_K$, we have a canonical embedding $\mathfrak{a} \hookrightarrow \mathbf{C}$. Send $[\mathfrak{a}]$ to the elliptic curve $\mathbf{C}/\mathfrak{a}$. This gives us a simply transitive action of $\mathcal{O}_K$ on $\mathcal{E}_{\mathbf{C}}(K)$ via $[\mathfrak{a}] \cdot (\mathbf{C}/\mathfrak{b}) = \mathbf{C}/(\mathfrak{a}^{-1}\mathfrak{b})$.

4.5 Fields of definition

It is known that CM elliptic curves have rational models. In other words, the natural map $\mathcal{E}_{\overline{\mathbf{Q}}}(K) \to \mathcal{E}_{\mathbf{C}}(K)$ induced by an embedding $\overline{\mathbf{Q}} \hookrightarrow \mathbf{C}$ is a bijection. We know this because E is *always* defined over $\mathbf{Q}(j(E))$. But when E is CM, $j(E) \in \overline{\mathbf{Q}}$, so E is defined over $\overline{\mathbf{Q}}$. Indeed, if $\sigma \in \mathrm{Aut}\,\mathbf{C}$, then $j(E^\sigma)$ also has CM by K. So the orbit of E under $\mathrm{Aut}(\mathbf{C})$ lies in $\mathcal{E}_{\mathbf{C}}(K)$, a finite set. This tells us that $\mathcal{E}_{\overline{\mathbf{Q}}}(K) \twoheadrightarrow \mathcal{E}_{\mathbf{C}}(K)$. We leave injectivity as an exercise.

Because of this, we will write $\mathcal{E}(K)$ instead of $\mathcal{E}_{\overline{\mathbf{Q}}}(K)$ or $\mathcal{E}_{\mathbf{C}}(K)$.

4.6 Towards H

Fix $E \in \mathcal{E}(K)$. For each $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/K)$, there is a unique $[\mathfrak{a}] \in \mathrm{Cl}_K$ such that $E^\sigma = [\mathfrak{a}] \cdot E$. We define a map $F : \mathrm{Gal}(\overline{\mathbf{Q}}/K) \to \mathrm{Cl}_K$ by $F(\sigma) = [\mathfrak{a}]$.

Proposition 4.6.1. *1. F does not depend on the choice of E.*
2. F is a homomorphism.

Proof. 1. This is subtle.
2. Let $\sigma, \tau \in \mathrm{Gal}(\overline{\mathbf{Q}}/K)$. Then

$$\begin{aligned} F(\sigma\tau)E &\simeq E^{\sigma\tau} \\ &= (E^\sigma)^\tau && \text{part 1} \\ &= (F(\sigma)E)^\tau \\ &= F(\tau)(F(\sigma)E) \\ &= (F(\sigma)F(\tau))E && \mathrm{Cl}_K \text{ is abelian} \end{aligned}$$

□

Let L be the fixed field of $\ker(F)$. We claim that $L = K(j(E))$. Indeed,

$$\begin{aligned} \mathrm{Gal}(\overline{\mathbf{Q}}/L) &= \{\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/K) : E^\sigma \simeq F(\sigma)E \simeq E\} \\ &= \{\sigma : j(E)^\sigma = j(E)\}. \end{aligned}$$

Note that F is an injection $\mathrm{Gal}(L/K) \hookrightarrow \mathrm{Cl}_K$. So L/K is abelian. We will show that it is unramified of the right degree.

Let $\mathfrak{m}$ be the conductor of L/K. It is the greatest common divisor of all $\mathfrak{m} \subset \mathcal{O}_K$ such that $K_{\mathfrak{m},1} \subset \ker(\mathrm{Art}_{L/K})$, where $K_{\mathfrak{m},1} = \langle (\alpha) : \alpha \simeq 1 \pmod{\mathfrak{m}} \rangle$. One checks that the composite

$$I_K^{\mathfrak{m}} \xrightarrow{\mathrm{Art}} \mathrm{Gal}(L/K) \xrightarrow{F} \mathrm{Cl}_K$$

is the "identity map" $\mathfrak{a} \mapsto [\mathfrak{a}]$. It follows that F is surjective. So $F : \mathrm{Gal}(L/K) \xrightarrow{\sim} \mathrm{Cl}_K$, and $L = K(j(E))$.

If $F(((\alpha), L/K)) = 1$ then $(\alpha) \in I_K^{\mathfrak{m}}$ is principal. Indeed, it suffices to show $((\alpha), L/K) = 1$, and this follows from the injectivity of F. Since α was arbitrary, $\mathfrak{m} = 1$, so L/K is unramified. The rest is a simple dimension argument.

Theorem 4.6.2. *If E is CM, then $j(E)$ is an algebraic integer.*

As a corollary, one has the surprizing fact that $e^{\pi\sqrt{163}}$ is very close to an integer. The main theorem of complex multiplication has an analogue for ray class fields. One gets $\mathrm{RCF}(\mathfrak{m}) = K(j(E), h(E[\mathfrak{m}]))$, where $\mathfrak{m} \subset \mathcal{O}_K$ and h is the Weber function.

5 Geometric properties of curves

Here we treat those properties of curves which hold over an arbitrary field. Later on, in section 7, we will specialize to number fields.

5.1 Motivation

Throughout, k is a field. For simplicity we assume k has characteristic zero.

Definition 5.1.1. A *curve* over a field k is a smooth geometrically connected variety of dimension one over k.

Concretely, we think of equations like

$$1 = x^2 - Dy^2$$
$$y^2 = x^3 + ax + b$$
$$z^n = x^n + y^n.$$

The difference between the general definition and these concrete examples should be seen as analogous to the difference between the notion of an "abstract vector space" and concrete examples $\mathbf{R}^n$.

The key tool for passing from an abstract curve to a concrete representation is the *Riemann-Roch Theorem*. Let X be a proper curve. Zariski-open subsets of X are of the form $U = X \smallsetminus \{p_1, \ldots, p_s\}$, where the $p_i \in X(\bar{k})$ and $\{p_1, \ldots, p_s\}$ is stable under the action of $G_k = \mathrm{Gal}(\bar{k}/k)$. We have a sheaf $\mathscr{O} = \mathscr{O}_X$ of "regular functions" on X. For $U \subset X$, the ring $\mathscr{O}(U)$ consists of all regular functions $U \to \mathbf{A}^1$.

Our goal is to understand $\mathscr{O}(U)$ as a ring. Ideally, we would like to write $\mathscr{O}(U) = k[f_1, \ldots, f_n]/(p_1, \ldots, p_m)$. In other words, we want sections $f_1, \ldots, f_n \in \mathscr{O}(U)$ such that $(f_1, \ldots, f_n)$ induces an embedding $U \hookrightarrow \mathbf{A}^n$. We certainly can't do this with $U = X$, because $\mathscr{O}(X) = k$. If $k = \mathbf{C}$, this fact is known as Liouville's theorem, but it holds for arbitrary k.

5.2 Crude form of Riemann-Roch

We assume there is a point $\infty \in X(k)$. Put $U = X \smallsetminus \{\infty\}$. Define $\mathscr{O}(U; n\infty)$ to be the set of functions $f \in \mathscr{O}(U)$ such that $v_\infty(f) \geqslant -n$. This gives us a filtration $\mathscr{O}(X) \subset$

$\mathscr{O}(U;\infty) \subset \mathscr{O}(U;2\infty) \subset \cdots$ and $\bigcup \mathscr{O}(U,n\infty) = \mathscr{O}(U)$. Moreover, each successive quotient is at most one-dimensional, so $\dim \mathscr{O}(U,n\infty) \leqslant n+1$. The Riemann-Roch Theorem gives a lower bound for $\dim \mathscr{O}(U,n\infty)$.

Theorem 5.2.1 (Riemann-Roch; crude form)**.** *Let X be a proper curve over k and $\infty \in X(k)$. Then there is an integer $g \geqslant 0$, depending only on X, such that*

$$\dim \mathscr{O}(U;n\infty) \geqslant n+1-g$$

with equality if $n \gg 0$.

Idea of proof. Choose a local parameter t at ∞. Define the *principal part* $\mathrm{pp}_\infty : \mathscr{O}(U;n\infty) \to t^{-n}k[t]/k[t]$ in the obvious way. What are the obstructions to producing f with given principal part at ∞? The only obstruction comes from the Residue Theorem stated below. As a corollary, if $\omega \in \Omega^1(X)$ is a global regular differential, and if $f \in \mathscr{O}(U)$, then $\mathrm{res}_\infty(f\omega) = 0$. So global regular differentials provide obstructions to constructing f. Define $\mathrm{res}_\infty : t^{-n}k[t]/k[t] \to \Omega^1(X)^\vee$ by $\mathrm{res}_\infty(f)(\omega) = \mathrm{res}_\infty(f\omega)$. We have a 5-term sequence

$$0 \to k \to \mathscr{O}(U;n\infty) \xrightarrow{\mathrm{pp}_\infty} t^{-n}k[t]/t \xrightarrow{\mathrm{res}_\infty} \Omega^1(X)^\vee \to \Omega^1(X;-n\infty)^\vee \to 0.$$

The last term needs explanation. Define $\Omega^1(X;-n\infty)$ to be the set of $\omega \in \Omega^1(X)$ such that $v_\infty(\omega) \geqslant n$. The inclusion $\Omega^1(X;-n\infty) \hookrightarrow \Omega^1(X)$ induces the surjection in the sequence.

One can check that the sequence is exact. For $n \gg 0$, $\Omega^1(X;-n\infty) = 0$. We have in fact proved the "more precise form" below. □

Theorem 5.2.2 (residue theorem)**.** *If ω is a meromorphic differential on X, then $\sum_{x\in X} \mathrm{res}_x(\omega) = 0$.*

Let's recall what this means At each $x \in X(\bar{k})$, choose a uniformizing parameter t at x. One can write locally $\omega = (a_{-m}t^{-m} + \cdots)dt$; put $\mathrm{res}_x(\omega) = a_{-1}$. Surprizingly, this does not depend on our choice of t.

Theorem 5.2.3 (Riemann-Roch; more precise form)**.** *Let $g = \dim \Omega^1(X)$. Then $\dim \mathscr{O}(U;n\infty) - \dim \Omega^1(X;-n\infty) = n+1-g$.*

We call the integer g the *genus* of X.

5.3 Some vocabulary

A *divisor* of X is a formal finite linear combination of points in $X(\bar{k})$ with integer coefficients. So a typical divisor looks like

$$D = \sum_{x\in X(\bar{k})} n_x \cdot x,$$

where each $n_x \in \mathbf{Z}$ and $n_x = 0$ for all but finitely many x. Write $\mathrm{Div}(X_{\bar{k}})$ for the (abelian group) of divisors on X, and write $\mathrm{Div}(X) = \mathrm{Div}(X_{\bar{k}})^{G_k}$.

Recall the field of *rational functions* on X is $k(X) = \varinjlim_U \mathscr{O}(U)$. Define $\mathrm{div} : k(X)^\times \to \mathrm{Div}(X)$ by

$$\mathrm{div}(f) = \sum_{x \in X(\bar{k})} v_x(f) \cdot x.$$

Divisors of the form $\mathrm{div}(f)$ are called *principal divisors*. If D_1, D_2 are divisors, write $D_1 \geqslant D_2$ if $n_x(D_1) \geqslant n_x(D_2)$ for all $x \in X(\bar{k})$. For an arbitrary divisor D, define

$$\mathscr{L}(D) = \{f \in k(X) : \mathrm{div}(f) \geqslant -D\}.$$

Our space $\mathscr{O}(U; n\infty)$ earlier is just $\mathscr{L}(n\infty)$. Define the space of *meromorphic divisors* by $\Omega^1_{\mathrm{mer}}(X) = \varinjlim_U \Omega^1(U)$; this is a one-dimensional $k(X)$-vector space. Choose nonzero $\omega \in \Omega^1_{\mathrm{mer}}(X)$. We call $K = \mathrm{div}(\omega)$ the *canonical divisor*. It does not depend on the choice of ω.

It turns out that $\Omega^1(X)$ can be identified with $\mathscr{L}(K)$. Indeed, we have a natural isomorphism $\mathscr{L}(K) \xrightarrow{\sim} \Omega^1(X)$ defined by $f \mapsto f\omega$.

Finally, if D is a divisor, one often writes $\ell(D)$ for $\dim \mathscr{L}(D)$. We can now state the final form of the Riemann-Roch Theorem.

Theorem 5.3.1 (Riemann-Roch)**.** *For all divisors $D \in \mathrm{Div}(X)$, we have $\ell(D) - \ell(K - D) = \deg D + 1 - g$.*

5.4 Consequences of Riemann-Roch

We could set $D = 0$. Then the theorem specializes to $1 - \ell(K) = 1 - g$, so $\ell(K) = g$. Since $\ell(K) = \dim \Omega^1(X)$, this recovers our definition of the genus of X.

We could set $D = K$. Then the theorem tells us that $\ell(K) - 1 = \deg K + 1 - g$. We already know $\ell(K) = g$, so $g - 1 = \deg K + 1 - g$, so $\deg K = 2g - 2$. In other words, the number of zeros of a non-zero $\omega \in \Omega^1(X)$ is $2g - 2$.

Example 5.4.1 ($g = 0$)**.** If $X(k) \neq \varnothing$, choose a point $\infty \in X(k)$. Then Riemann-Roch says $\ell(n\infty) = n + 1$. In particular, $\mathscr{L}(\infty) = k \oplus kt$, $\mathscr{L}(2\infty) = k \oplus kt \oplus kt^2$, and in egneral $\mathscr{L}(n\infty) = k \oplus \cdots \oplus kt^n$. So $\mathscr{O}(U) = k[t]$, whence $U \simeq \mathbf{A}^1$ and $X \simeq \mathbf{P}^1$. Even if $X(k) = \varnothing$, X has a rational divisor of degree 2, namely $-K$. It must of be of the form $-K = p + p'$ for p, p' conjugates in a quadratic extension of $k(X)$. We know that $\mathscr{L}(-K) = k \oplus ku \oplus kv$, and that $\mathscr{L}(-2K)$ is spanned by $\{1, u, v, uv, v^2, u^2\}$. But $\ell(-2K) = 5$, so we must have a linear relation $a + bu + cv + duv + ev^2 + fu^2 + 0$. In particular, all curves of genus zero are conics.

Example 5.4.2 (elliptic curves)**.** A curve of genus 1 with chosen $\infty \in X(k)$ is called an *elliptic curve*. Since $\dim \Omega^1(X) = 1$, there is (up to homothety) a unique non-vanishing regular divisor $\omega \in \Omega^1(X)$. This gives an isomorphism $\mathscr{O}_X \xrightarrow{\sim} \Omega^1_X$. Apply Riemann-Roch to the spaces $\mathscr{L}(n \cdot \infty)$. We get

n	$\ell(n \cdot \infty)$	generators of $\mathscr{L}(n \cdot \infty)$
1	1	$\{1\}$
2	2	$\{1, x\}$
3	3	$\{1, x, y\}$
4	4	$\{1, x, y, x^2\}$
5	5	$\{1, x, y, x^2, xy\}$
6	6	$\{1, x, y, x^2, xy, y^2, x^3\}$

Since $\ell(6 \cdot \infty) = 6$, we must have $y^2 - x^3 \in \mathscr{L}(5\infty)$, so we get a relation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

Now elementary algebra gets rid of a_1 if 2 is invertible, and reduces further to an equation

$$y^2 = x^3 + ax + b$$

if 3 is also invertible.

If X is a curve of genus one with $X(k) = \varnothing$, the simplest case is when X has a rational divisor of degree 2. This is true for elements of $\mathrm{Sel}_2(E)$. We have

Theorem 5.4.3. *X has an equation of the form*

$$y^2 = ax^4 + bx^3 + cx^2 + dx + e.$$

Proof. Our rational divisor of degree 2 is of the form $p + p'$. We have

$$\begin{aligned}\mathscr{L}(p + p') &= k \oplus kx \\ \mathscr{L}(2p + 2p') &= k \oplus kx \oplus kx^2 \oplus ky \\ \mathscr{L}(4p + 4p') &= \dots\end{aligned}$$

The rest is easy. □

6 Basic analytic number theory

We are interested in asymptotics of $\sum_{n \leqslant x} a_n$ for various natural arithmetic sequences $\{a_n\}$. There are two main techniques: one is geometric, the other uses L-functions.

6.1 Geometric techniques

Example 6.1.1. Consider the constant sequence $a_n = 1$. We have $\lfloor x \rfloor = \sum_{1 \leqslant n \leqslant x} 1$. More precisely, $\sum_{n \leqslant x} 1 = x + O(1)$.

Example 6.1.2. We count lattice points inside a disk, i.e. look at the asymptotics of $\#\{(x, y) \in \mathbf{Z}^2 : x^2 + y^2 \leqslant T\}$. This is approximately the area of $\{x^2 + y^2 \leqslant T\}$. The error comes from the boundary of the region $\{x^2 + y^2 \leqslant T\}$. It will be a bounded multiple of the radious $\sqrt{T}$. So

$$\#\{(x, y) \in \mathbf{Z}^2 : x^2 + y^2 \leqslant T\} = \pi T + O(T^{1/2}).$$

We might hope for an error term of the form $O(T^{1/2-\epsilon})$ for some $\epsilon > 0$.

Questions like this become quite subtle if we are looking at intersections $\Lambda \cap r\Omega$, where $\Lambda \subset \mathbf{R}^n$ is a lattice and $\Omega \subset \mathbf{R}^n$ is a bounded region. If $\partial\Omega$ is smooth, things work as expected. If, however Ω has a "fractal-like" boundary, one has to be very careful.

Example 6.1.3. Let's count the number of lattice points inside an expanding triangle:

$$\#\{(x,y) \in \mathbf{Z}^2 : x, y > 0 \text{ and } y + \alpha x \leqslant T\}.$$

The area of the triangle $T\Delta = \{x, y > 0 \text{ and } y + \alpha x \leqslant T\}$ is $\frac{1}{2\alpha}T^2$. We get

$$\#(T\Delta \cap \mathbf{Z}^2) = \frac{1}{2\alpha}T^2 + O(T).$$

As with the circle, we could hope for an error term of the form $O(T^{1-\epsilon})$ for $\epsilon > 0$. By considering $\alpha = -1$, we can see that this is not possible. What if α is not rational? If we consider

$$\alpha = 1 + N^{-1} + 2^{-N} + 2^{2^N} + \cdots$$

then α is "almost rational," which leads to a "full error term" $O(T)$ for $\#(T\Delta \cap \mathbf{Z}^2)$. From this we see that diophantine approximation of transcendental numbers is relevant to these sorts of problems.

Example 6.1.4. Let d be the *divisor function* defined by

$$d(n) = \#\{(a,b) \in \mathbf{N} : ab = n\}.$$

We are interested in $\sum_{n \leqslant T} d(n)$. We can rewrite this as

$$\sum_{n \leqslant T} d(n) = \sum_{n \leqslant T} \sum_{\substack{x,y \geqslant 1 \\ xy = n}} 1 = \sum_{\substack{x,y \geqslant 1 \\ xy \leqslant T}} 1.$$

So we're trying to count lattice points in $T\Omega$, where

$$\Omega = \{(x,y) \in \mathbf{R}^2 : x, y > 0 \text{ and } xy \leqslant 1\}.$$

But Ω has infinite area and a pathological boundary. Instead, let's count lattice points in

$$\{(x,y) \in \mathbf{R}^2 : x, y > \frac{1}{2} \text{ and } xy \leqslant T\}.$$

This is bounded, so we're in good shape. It's area is

$$\begin{aligned} \int_{1/2}^{2T} \frac{T}{x}\, dx &= T \log(4T) \\ &= T \log T + O(T). \end{aligned}$$

Dirichlet has a beautiful trick for the asymptotics of the divisor function. We have

$$\begin{aligned} \sum_{\substack{a,b \geqslant 1 \\ ab \leqslant T}} 1 &= \sum_{T \geqslant a \geqslant 1} \sum_{1 \leqslant b \leqslant T/a} 1 \\ &= \sum_{1 \leqslant a \leqslant T} \left(\frac{T}{a} + O(1) \right) \\ &= Ta \sum_{1 \leqslant a \leqslant T} \frac{1}{a} + O(T) \end{aligned}$$

We know that $\sum_{n \leqslant N} \frac{1}{n} = \log N + \gamma + O(N^{-1})$. This also gives us $T \log T + O(T)$. We would like a power-saving error term. Dirichlet's insight was that when summing pairs (a, b) with $ab = n$, we can restrict to those with $a \leqslant b$. Write $m = \min\{a, b\}$ and $n = \max\{a, b\}$. We have

$$\sum_{\substack{ab \leqslant T \\ a,b \geqslant 1}} 1 = \sum_{1 \leqslant m \leqslant \sqrt{T}} \sum_{m < n \leqslant \frac{T}{m}} 1 + \sqrt{T},$$

yielding a sum

$$\sum_{n \leqslant T} d(n) = 2T \sum_{m=1}^{\lfloor \sqrt{T} \rfloor} \frac{1}{m} - T + O(\sqrt{T}) = T \log T + (2\gamma - 1)T + O(T^{1/2}).$$

6.2 L-functions

Example 6.2.1. Consider the identity

$$\int_0^1 e^{2\pi i n t} \, dt = \begin{cases} 0 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

This is a characteristic function for the integer $n = 0$. Suppose we wanted to attack Goldbach's conjecture, which says that each $2N$ can be written as $p + q$ for primes $p + q$. We could look at

$$\begin{aligned} \sum_{p,q \text{ prime}} \begin{cases} 1 & p + q - 2N = 0 \\ 0 & \neq 0 \end{cases} &= \sum_{p,q} \int_0^1 e^{2i\pi(p+q-2N)t} \, dt \\ &= \int_0^1 e^{4i\pi N t} \left(\sum_p e^{2i\pi t} \right)^2 dt. \end{aligned}$$

This is essentially the Hardy-Littlewood circle method.

It would be nice if instead of just characteristic functions of points, we could get characteristic functions of more general regions via integrals. One has the *Perron formula*

$$\frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} e^{sy} \frac{ds}{s} = \begin{cases} 1 & y > 0 \\ \frac{1}{2} & y = 0 \\ 0 & y < 0 \end{cases}$$

for $c > 0$. This is essentially Cauchy's residue theorem. One integrates over increasingly large squares with $\{\Re z = c\}$ as their right side.

Suppose $e^y = w$. Then we are integrating w^s/s, and the characteristic function is

for $w > 1$. Write

$$\begin{aligned}\sum_{n\leqslant x} a_n &= \sum_{n\geqslant 1} a_n \begin{cases} 1 & x/n > 1 \\ 0 & x/n < 1 \end{cases} \\ &= \sum_{n\geqslant 1} a_n \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} \left(\frac{x}{n}\right)^s \frac{ds}{s} \\ &= \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} A(s)x^s \frac{ds}{s}\end{aligned}$$

for $\Re c \gg 0$, where $A(s) = \sum_{n\geqslant 1} \frac{a_n}{n^s}$. Let's apply this approach to $\lfloor x\rfloor = \sum_{n\leqslant x} 1$. We get

$$\lfloor x\rfloor = \frac{1}{2i\pi}\int_{2-i\infty}^{2+i\infty} \zeta(s)x^s \frac{ds}{s}.$$

The function ζ is analytic except at $s = 1$, where it has a pole of order 1 with residue 1. Thus

$$\lfloor x\rfloor = x + \zeta(0) + \text{error} = x - \frac{1}{2} + \text{error}.$$

Our error term is the "sawtooth function" $x - \frac{1}{2} - \lfloor x\rfloor$.

Now we consider the more complicated sum $\sum_{n\leqslant x} d(n)$. Our corresponding Dirichlet series is

$$D(s) = \sum_{n\geqslant 1} \frac{d(n)}{n^s} = \sum_{n\geqslant 1} n^{-s} \sum_{\substack{ab=n \\ a,b\geqslant 1}} 1 = \sum_{a,b\geqslant 1} \frac{1}{(ab)^s} = \zeta(s)^2.$$

Near $s = 1$, we have $\zeta(s) = (s-1)^{-1} + \gamma + c(s-1) + \cdots$. So

$$\begin{aligned}\zeta(s)^2\frac{x^s}{s} &= \left(\frac{1}{s-1} + \gamma + c_1(s-1)\right)^2 \cdot x \cdot (1 + (s-1)\log x + \ldots)\cdots \\ &= x\left(\frac{1}{(s-1)^2} + \frac{1}{s-1}(\log x + 2\gamma - 1) + \cdots\right).\end{aligned}$$

This recovers Dirichlet's formula for $\sum_{n\leqslant x} d(n)$.

Finally, let's review Riemann's original application of the zeta function. From the Euler product $\zeta(s) = \prod(1-p^{-s})^{-1}$ valid for $\Re s \geqslant 1$, we compute

$$-\frac{\zeta'(s)}{\zeta(s)} = \sum_{\substack{p \text{ prime} \\ m\geqslant 1}} \frac{\log p}{p^{ms}}.$$

It follows that

$$\sum_{p^m\leqslant x} \log p = \frac{1}{2i\pi}\int_{2-i\infty}^{2+i\infty} -\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}\,ds.$$

The poles of ζ' are easy to analyze. The other poles of the integrand come from zeros of ζ. We get

$$\sum_{p^m\leqslant x} \log p = x - \frac{\zeta'(0)}{\zeta(0)} - \sum_{\zeta(\rho)=0} \frac{x^\rho}{\rho}.$$

It is trickier to count lattice points inside a circle using zeta functions. We have

$$\sum_{a^2+b^2\leqslant T} 1 = \sum_{n\leqslant T} R(n),$$

where $R(n) = \#\{(a,b)\in \mathbf{Z}^2 : n = a^2+b^2\} = 4r(n)$. The Dirichlet series has an Euler product

$$\left(1-\frac{1}{2^s}\right)^{-1} \prod_{p\equiv 1 \pmod 4} \left(1-\frac{1}{p^s}\right)^{-2} \prod_{p\equiv 3 \pmod 4} \left(1-\frac{1}{p^{2s}}\right)^{-1}$$

If $\chi = \left(\frac{-4}{\cdot}\right)$, then

$$L(s,\chi) = \prod_{p\equiv 1 \pmod 4} \left(1-\frac{1}{p^s}\right)^{-1} \prod_{p\equiv 3 \pmod 4} \left(1+\frac{1}{p^s}\right)^{-1}.$$

It follows that our Dirichlet series is $\zeta(s)L(s,\chi)$.

6.3 Sieving

Let's try to count square-free integers. We have

$$\begin{aligned}
\sum_{\substack{n\leqslant x\\ n \text{ squarefree}}} 1 &= \lfloor x\rfloor - \sum_p \#\{n\leqslant x : p^2 \mid n\} + \sum_{p,q} \#\{n\leqslant x : p^2q^2 \mid n\} \\
&= \lfloor x\rfloor - \sum_p \left\lfloor \frac{x}{p^2}\right\rfloor + \sum_{p,q} \left\lfloor \frac{x}{p^2q^2}\right\rfloor \\
&= x + O(1) - \sum_{p\leqslant x} \left(\frac{x}{p^2} + O(1)\right) + \sum_{p,q} \left(\frac{x}{p^2q^2} + O(1)\right) \\
&= x\prod_p \left(1-\frac{1}{p^2}\right) + \text{error} \\
&= \frac{6}{\pi^2}x + \text{error}
\end{aligned}$$

A less risky approach (one that does not have as many error terms) is to write

$$\sum_{\substack{n\leqslant x\\ n \text{ squarefree}}} 1 = \sum_{\substack{n\leqslant x\\ p^2\mid n \text{ for all } p\leqslant y}} 1 + \text{error},$$

where $y = \log x$ and

$$|\text{error}| \leqslant \sum_{y<p<\sqrt{x}} \#\{n\leqslant x : p^2 \mid n\}.$$

A similar problem is

$$\#\{n\leqslant x : n^2+1 \text{ is squarefree}\} = \sum_{\substack{n\leqslant x\\ p^2\nmid n^2+1 \text{ for } p\leqslant y}} 1 + O\left(\sum_{y<p<x} \#\{n\leqslant x : p^2 \mid n^2+1\}\right)$$

The summand inside the big-O is bounded above by $2\left(\frac{x}{p^2}+1\right)$.

Conjecture 6.3.1. *For all $\varepsilon > 0$, there is a constant κ_ε such that whenever $a+b=c$ with $(a,b)=1$, then*

$$\prod_{p|abc} p > \kappa_\varepsilon \max\{|a|,|b|\}^{1-\varepsilon} = \kappa H(a,b)^{1-\varepsilon}.$$

A remarkable article of Noam Elkies relates the *abc*-conjecture to Belyi maps. The *abc* conjecture implies that if $F(x,y) \in \mathbf{Z}[x,y]$ is a homogeneous polynomial, then

$$\prod_{p|F(a,b)} p > \kappa H(a,b)^{\deg F - 2 - \varepsilon}$$

Given f, set $F(x,y) = y^{d+1} f(x/y)$. Then a consequence of *abc* is

$$\prod_{p|f(n)} p > \kappa_\varepsilon |n|^{\deg f + 1 - \varepsilon}.$$

See [Elk91] for details.

We could also consider $4a^3 + 27b^2$ divisible by p^2.

7 Diophantine properties of curves

Let X be a curve over a number field k. The main diophantine questions we are interested in are:

- What is $X(k)$?
- Is $X(k)$ finite?
- What is $\#X(k)$ for "typical" X?

We would like to phrase questions in a way that allow for us to talk about integral points on a curve – e.g. equations like the Pell equation $x^2 - dy^2 = 1$. If X is projective, then $X(\mathbf{Z}) = X(\mathbf{Q})$, so there is no limitation in studying rational points. More generally, if X/k is projective, then $X(\mathcal{O}_k) = X(k)$. If X is affine, we can choose an embedding $X \hookrightarrow \mathbf{A}^n$ and put $X(\mathbf{Z}) = X(\mathbf{Q}) \cap \mathbf{A}^n(\mathbf{Z})$. With this definition $X(\mathbf{Z})$ depepnds on the chosen equations for X, but hopefully the "main features" of $X(\mathbf{Z})$ do not depend on this embedding.

So our question is: if k is a number field, S is a finite set of places of S and $X/\mathcal{O}_{k,S}$ is a curve, what is $X(\mathcal{O}_{k,S})$?

To X we can attach some numerical invariants. The curve X will be of the form $\widetilde{X} \smallsetminus \{x_1, \ldots, x_s\}$ where $\widetilde{X}$ is proper. For g the genus of $\widetilde{X}$, we define the *Euler characteristic* of X by

$$\chi(X) = 2 - 2g - s \in \mathbf{Z}.$$

A lot of the diophantine behavior of X is governed by $\chi(X)$. The fundamental trichotomy comes from whether $\chi > 0$, $\chi < 0$, or $\chi = 0$.

7.1 Positive Euler characteristic

Theorem 7.1.1. *If $\chi(X) > 0$, then $X(\mathcal{O}_{k,S})$ is either empty or infinite.*

Proof. If X is affine, then $g = 0$ and $s = 1$, so $X = \mathbf{A}^1$, whence $X(\mathcal{O}_{k,S}) = \mathcal{O}_{k,S}$. If X is projective, then $g = s = 0$. Then X either has a rational point, in which case it is $\mathbf{P}^1$, or X is a conic with $X(k) = \varnothing$. □

Theorem 7.1.2 (Hasse-Minkowski)**.** *Let X be a curve over $\mathbf{Q}$ of genus zero. Then $X(\mathbf{Q}) \neq \varnothing$ if and only if $X(\mathbf{Q}_p) \neq \varnothing$ for all p and $X(\mathbf{R}) \neq \varnothing$.*

7.2 Negative Euler characteristic

Theorem 7.2.1 (Siegel,Faltings)**.** *If $\chi(X) < 0$, then $\#X(\mathcal{O}_{k,S}) < \infty$.*

The affine case was proven by Siegel in 1932. The prototypical examples are:

g	s	X
0	3	$\mathbf{P}^1 \smallsetminus \{0, 1, \infty\}$
1	1	$E \smallsetminus \{\infty\}$

The coordinate ring of $\mathcal{O}_{\mathbf{P}^1 \smallsetminus \{0,1,\infty\}}$ is $\mathbf{Z}[x, \frac{1}{x}, \frac{1}{1-x}]$, and

$$(\mathbf{P}^1 \smallsetminus \{0, 1, \infty\})(\mathcal{O}_{k,S}) = \{v \in \mathcal{O}_{k,S}^\times : v - 1 \in \mathcal{O}_{k,S}^\times\}.$$

This is an *S-unit equation*, and Siegel proved that such equations have only finitely many solutions.

If $g = 1$, $s = 1$, then the result amounts to showing that elliptic curves have only finitely many integral points. Since integral points are torsion, this follows from the Mordell-Weil Theorem.

In the projective case, $g > 1$, and the finiteness result is Faltings' Theorem, originally known as the Mordell Conjecture.

7.3 Zero Euler characteristic

This is the most interesting case.

Theorem 7.3.1 (Dirichlet,Mordell-Weil)**.** *If $X(\mathcal{O}_{k,S})$ is non-empty, then it is naturally an abelian group, and as such is finitely-generated.*

In the affine case $g = 0, s = 2$, if $X(k) \neq \varnothing$, then (for the sake of illustration) $X = \mathbf{P}^1 \smallsetminus \{0, \infty\} = \mathbf{G}_\mathrm{m}$, so $X(\mathcal{O}_{k,S}) = \mathcal{O}_{k,S}^\times$. The famous *Dirichlet Unit Theorem* tells us this group is finitely generated.

In the projective case $g = 1, s = 0$, X is an elliiptic curve which we will denote by E. The Mordell-Weil Theorem says that $E(k)$ is finitely generated.

7.4 Ranks

In the affine case, the rank of $\mathcal{O}_{k,S}^{\times}$ is easily determined. Dirichlet's theorem says that

$$\mathrm{rk}_{\mathbf{Z}}(\mathcal{O}_{k,S}^{\times}) = r + s - 1 + \#S,$$

where r is the number of real places and s is the number of complex places of k.

In the projective case, the rank is much more subtle. If $X = \mathbf{P}^1 \smallsetminus \{p, p'\}$ for p, p' conjugates in a quadratic extension $\mathbf{Q}(\sqrt{D})$, at least if $k = \mathbf{Q}$. We are led to the equation $x^2 - Dy^2 = 1$. This has rank 0 if $D < 0$, and rank 1 if $D > 0$.

For elliptic curves over $\mathbf{Q}$, little is known.

Conjecture 7.4.1. *For E ranging over elliptic curves defined over $\mathbf{Q}$, is $\mathrm{rk}\, E = \mathrm{rk}_{\mathbf{Z}}\, E(\mathbf{Q})$ bounded?*

Conjecture 7.4.2. *As E ranges over elliptic curves defined over $\mathbf{Q}$, $\mathrm{rk}\, E$ is 0 and 1 with probability $\frac{1}{2}$ each.*

Bhargava and Shankar have proved that there is a positive density set of elliptic curves having rank 0 and 1.

7.5 Proof of Mordell-Weil

The proof has two main ingredients. The first is a height function $h : E(\mathbf{Q}) \to \mathbf{R}$ satisfying the property that for each X, the set $\{x \in E(\mathbf{Q}) : h(x) < X\}$ is finite. Moreover, $h(n \cdot x) = n^2 h(x)$ and $h(x + y) + h(x - y) = 2h(x) + 2h(y)$. The second ingredient is the *weak Mordell-Weil theorem*:

Theorem 7.5.1. *For some $n > 1$, the group $E(\mathbf{Q})/n$ is finite.*

Proving Mordell-Weil from these two ingredients is a very old idea, going back to Fermat at least. Let $\{p_1, \ldots, p_r\}$ be a set of representatives for $E(\mathbf{Q})/n$. Choose $X \gg h(p_j)$, and let $S = \{p_1, \ldots, p_r\} \cup \{p : h(p) < X\}$. We claim that S generates $E(\mathbf{Q})$. Let p be a point not in $\langle S \rangle$ with minimal height with respect to this property. There exists some j such that $p - p_j = n \cdot q$. One sees that $h(q) < h(p)$, so $q \in \langle S \rangle$. This implies $p \in \langle S \rangle$, a contradiction.

7.6 Proof of weak Mordell-Weil

We do this for $n = 2$. Assume $E[2]$ is defined over $\mathbf{Q}$, i.e. E is of the form $y^2 = (x - a)(x - b)(x - c)$. Given $P \in E(\mathbf{Q})$, choose some $\widetilde{P} \in E(\overline{\mathbf{Q}})$ such that $2\widetilde{P} = P$. Define a function $\delta(P) : G_{\mathbf{Q}} \to E[2]$ by $\delta(P)(\sigma) = \sigma(\widetilde{P}) - \widetilde{P}$.

The function $\delta(P)$ is actually a continuous homomorphism $G_{\mathbf{Q}} \to E[2]$. Moreover, $\delta(P_1) = \delta(P_2)$ if and only if $P_1 - P_2 \in 2E(\mathbf{Q})$. So δ is an injection $E(\mathbf{Q})/2 \hookrightarrow \hom(G_{\mathbf{Q}}, E[2])$. This doesn't solve our problem because $\hom(G_{\mathbf{Q}}, E[2])$ is infinite. The necessary property of δ is the following. Let $L = \mathbf{Q}(\sqrt{\ell} : \ell \mid 2(a-b)(b-c)(a-c))$. Then $\delta(P)$ factors through $\mathrm{Gal}(L/\mathbf{Q})$. Indeed, if $P = (x, y)$, then $\widetilde{P}$ is defined over $\mathbf{Q}(\sqrt{x-a}, \sqrt{x-b}, \sqrt{x-c})$. It is easy to check that if $P \in E[2]$, then $y = 0$, and this implies $\widetilde{P}$ is defined over L.

To conclude, δ is an injection $E(\mathbf{Q})/2 \hookrightarrow \hom(\mathrm{Gal}(L/\mathbf{Q}), E[2])$, the latter being a finite set. Thus $E(\mathbf{Q})/2$ is finite.

Let's give a more "highbrow" proof using Galois cohomology. Let $n > 1$ be an integer. We have an exact sequence

$$0 \to E[n] \to E(\overline{\mathbf{Q}}) \xrightarrow{n} E(\overline{\mathbf{Q}}) \to 0.$$

Take $G_{\mathbf{Q}}$-invariants and we get an exact sequence

$$0 \to E(\mathbf{Q})/n \xrightarrow{\delta} \mathrm{H}^1(G_{\mathbf{Q}}, E[n]) \to \mathrm{H}^1(G_{\mathbf{Q}}, E)[n] \to 0.$$

The middle set is still infinite. Repeat the process for each place of $\mathbf{Q}$:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(\mathbf{Q})/n & \xrightarrow{\delta} & \mathrm{H}^1(G_{\mathbf{Q}}, E[n]) & \longrightarrow & \mathrm{H}^1(G_{\mathbf{Q}}, E)[n] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E(\mathbf{Q}_\ell)/n & \longrightarrow & \mathrm{H}^1(G_{\mathbf{Q}_\ell}, E[n]) & \longrightarrow & \mathrm{H}^1(G_{\mathbf{Q}_\ell}, E)[n] & \longrightarrow & 0 \end{array}$$

Define the *n-Selmer group* and *Tate-Shafarevich group* of E by

$$\mathrm{Sel}_n(E) = \ker\left(\mathrm{H}^1(G_{\mathbf{Q}}, E[n]) \to \bigoplus_v \mathrm{H}^1(G_{\mathbf{Q}_v}, E)\right)$$
$$Ш(E) = \ker\left(\mathrm{H}^1(G_{\mathbf{Q}}, E) \to \bigoplus_v \mathrm{H}^1(G_{\mathbf{Q}_v}, E)\right).$$

There is a canonical exact sequence

$$0 \to E(\mathbf{Q})/n \to \mathrm{Sel}_n(E) \to Ш(E)[n] \to 0.$$

An elementary argument using the Hermite-Minkowski theorem shows that $\mathrm{Sel}_n(E)$ is finite. Since $E(\mathbf{Q})/n \hookrightarrow \mathrm{Sel}_n(E)$, we're done.

7.7 Geometric interpretation of $\mathrm{Sel}_n(E)$

In general, we know that $\mathrm{H}^1(G_{\mathbf{Q}}, \mathrm{Aut}\, X)$ classifies $\mathbf{Q}$-forms of X. We would like to find an object whose automorphism group is $E[n]$. Consider the isogeny $E \xrightarrow{n} E$. The automorphisms of this cover of E are exactly elements of $E[n]$.

Definition 7.7.1. An *n-cover* of E is a curve C of genus 1, equipped with a $\mathbf{Q}$-rational map $\widetilde{n} : C \to E$ and a $\overline{\mathbf{Q}}$-isomorphism $\varphi : C \xrightarrow{\sim} E$ such that the following diagram commutes:

$$\begin{array}{ccc} C & \xrightarrow{\varphi} & E \\ \downarrow{\scriptstyle \widetilde{n}} & & \downarrow{\scriptstyle n} \\ E & = & E \end{array}$$

Similarly, $\mathrm{H}^1(\mathbf{Q}, E)$ can be identified with the set of isomorphism classes of curves of genus one such that $\operatorname{Jac} C \simeq E$ over $\mathbf{Q}$. The map $\mathrm{H}^1(\mathbf{Q}, E[n]) \to \mathrm{H}^1(\mathbf{Q}, E)$ comes from "forgetting $\widetilde{n}$." It follows that $\mathrm{Sel}_n(E)$ can be identified with the set of isomorphism classes of n-covers $\widetilde{n} : C \to E$ such that $C(\mathbf{Q}_\ell) \neq \varnothing$ for all ℓ and $C(\mathbf{R}) \neq \varnothing$. Similarly $Ш(E)$ consists of isomorphism classes of genus-one curves C such that $\operatorname{Jac} C \simeq E$, and such that $C(\mathbf{Q}_v) \neq \varnothing$ for all places v.

Theorem 7.7.2 (Swinnerton-Dyer)**.** *If $\widetilde{2} : C \to E$ is an element of $\mathrm{Sel}_2(E)$, then C has a $\mathbf{Q}$-rational positive divisor of degree 2.*

Proof. Degree 2 divisors of C correspond to rational points on $\operatorname{Sym}^2(C)$. Define a rational morphism $\varphi : \operatorname{Sym}^2(C) \to E$ by $(P, Q) \mapsto P + Q = \widetilde{2}(P) - (Q - P)$. (Recallhere that $E = \operatorname{Jac} C$.) Then $X = \varphi^{-1}(0)$ is a curve. Over $\overline{\mathbf{Q}}$, the map ϕ can be identified with the addition map $\operatorname{Sym}^2(E) \to E$. So $X_{\overline{\mathbf{Q}}} = \{(P, -P) : P \in E(\mathbf{Q})\}$. In fact, $X_{\overline{\mathbf{Q}}} = (E/-1)_{\overline{\mathbf{Q}}} = \mathbf{P}^1_{\overline{\mathbf{Q}}}$. The same reasoning, replacing $\overline{\mathbf{Q}}$ by $\mathbf{Q}_\ell$ and using the fact that $C \simeq E$ over each $\mathbf{Q}_v$, tels us that X has a rational point in each $\mathbf{Q}_v$. By Hasse-Minkowski, $X \simeq \mathbf{P}^1$, whence the result. □

Corollary 7.7.3. *If $\widetilde{2} : C \to E$ is an element of $\mathrm{Sel}_2(E)$, then C has an equation of the form $y^2 = f(x)$, where $\deg f = 4$.*

This gives us the dictionary between elements of a 2-Selmer group and binary quartic forms over $\mathbf{Q}$. Bhargava and Shankar show that there is a positive proportion of E with $\mathrm{Sel}_n E = 0$, and also a positive proportion of E with $\mathrm{Sel}_n E = dZ/n$. Both of these only hold for $n \in \{2, 3, 4, 5\}$.

Theorem 7.7.4. *1. If $\mathrm{Sel}_n E = 0$, then $\mathrm{rk}\, E = 0$.*
2. If $\mathrm{Sel}_n E = \mathbf{Z}/n$, then $\mathrm{rk}\, E = 1$.

Part 1 is trivial. Part 2 is incredibly deep. It uses in a crucial way the connection between elliptic curves and L-functions. This is a special case of the Birch and Swinnerton-Dyer conjecture. A big ingredient is the theory of complex multiplication discussed in section 4.

8 More algebraic groups, representation theory and invariant theory

The notes for this lecture can be found online at http://www.math.mcgill.ca/goren/AlgberaicGroups.SMS2014.pdf. Good references on representation theory include [Hum78, FH91]. For invariant theory, see [MFK94, DC70, GW09].

8.1 Lie algebra

Let G be a reductive group over an algebraically closed field k of characteristic zero. The *Lie algebra* of G is $\mathrm{Lie}(G) = T_1 G$, the tangent space at the identity of 1. We can identify $\mathrm{Lie}\, G$ with the space of left-invariant derivations of G. This identification gives $\mathrm{Lie}(G)$ the structure of a Lie algebra via $[X, Y] = X \circ Y - Y \circ X$. The operation $G \mapsto$

$\mathrm{Lie}(G)$ is a functor. In particular, the operation of G on itself by inner automorphisms induces $\mathrm{ad} : G \to \mathrm{GL}(\mathfrak{g})$.

The fundamental example is $G = \mathrm{GL}_N$, $\mathfrak{g} = \mathfrak{gl}_n$, where $\mathrm{ad}(g)(X) = gXg^{-1}$. For $H \subset G$, $\mathfrak{h} = \mathrm{Lie}\, H$ is found by "1st-order approximations to the equations defining H." If q is the Euclidean bilinear form on k^n, then $\mathrm{O}(q) = \{x : x\,{}^{\mathrm{t}}x = 1\}$. If we write $x = 1 + x'$, then

$$(1 + x')\,{}^{\mathrm{t}}(1 + x') = 1 + x' + {}^{\mathrm{t}}x' + x'\,{}^{\mathrm{t}}x' = 1$$

so infinitesimally,

$$\mathfrak{o} = \mathfrak{so}(q) = \{X \in \mathfrak{gl}_n : X + {}^{\mathrm{t}}X = 0\}.$$

Similarly, $\mathrm{SL}_N = \{g \in \mathrm{GL}_N : \det g = 1\}$. One easily computes $\mathfrak{sl}_n = \{X \in \mathfrak{gl}_n : \mathrm{tr}\, X = 0\}$.

8.2 Root systems

Let $T \subset G$ be a maximal torus and $\mathfrak{h} = \mathrm{Lie}\, T$. Since T is "universally semisimple," we have

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha,$$

where $\mathfrak{g}_\alpha = \{X \in \mathfrak{g} : \mathrm{ad}(t)X = \alpha(t)X\}$ and $\Phi = \Phi(G,T) = \{\alpha \in \mathrm{X}^*(T) \smallsetminus 0 : \mathfrak{g}_\alpha \neq 0\}$. The space $\mathrm{X}^*(T)_{\mathbf{R}}$ is equipped with a canonical inner product coming from the Killing form on $\mathfrak{g}$. The pair $(\mathrm{X}^*(T), \Phi)$ is a *root system*, i.e. $\Phi = -\Phi$, $s_\alpha(\Phi) = \Phi$ for all $\alpha \in \Phi$. Here $s_\alpha(v) = v - 2\frac{\langle \alpha, v\rangle}{\langle \alpha, \alpha\rangle}\alpha$ is the reflection induced by α.

A *Weyl chamber* W is a connected component of $\mathrm{X}^*(T)_{\mathbf{R}} \smallsetminus \bigcup_{\alpha\in\Phi} \alpha^\perp$. It determines an ordering of the roots:

$$\alpha > 0 \text{ if } \langle \alpha, v\rangle > 0 \text{ for all } r \in W.$$

A positive root is called *simple* if it is not a non-trivial sum of other positive roots. Let $\Delta \subset \Phi$ be the set of simple roots. Any root can be written as $\sum_{\delta\in\Delta} a_\delta \delta$, where the a_δ are either all positive or all negative.

The choice of a borel $B \supset T$ is equivalent to the choice of a Weyl chamber W. Given B, we put $\mathfrak{b} = \mathrm{Lie}\, B = \sum_{\alpha \geqslant 0} \mathfrak{g}_\alpha$. The subalgebra $\mathfrak{u} = \sum_{\alpha>0} \mathfrak{g}_\alpha$ is the Lie algebra of the unipotent radical of B. We have $B = T \cdot U$ with $\mathrm{Lie}\, U = \mathfrak{u}$.

Given a subset $\Theta \subset \Delta$, set $S_\Theta = \left(\bigcap_{\alpha\in\Theta} \ker(\alpha)\right)^\circ$; this is a torus of rank $\mathrm{rk}\, G - \#\Theta$. Define $P_\Theta = Z(S_\Theta) \cdot U$; this is a *standard parabolic subgroup* of G.

Theorem 8.2.1. *There is a bijection between parabolic subgroups of G (up to conjugation) and standard parabolics. There are exactly $2^{\#\Delta}$ such standard parabolics.*

Example 8.2.2. If $G = \mathrm{GL}_N$ and T is the diagonal torus, define $\lambda_i \in \mathrm{X}^*(T)$ by $\lambda_i(t_{jj}) = t_{ii}$. Then $\mathrm{X}^*(T) = \bigoplus \mathbf{Z} \cdot \lambda_i$. The Lie algebra $\mathfrak{g} = \mathrm{Lie}\, G$ is just $\mathfrak{gl}_n = M_n$ with basis $\{E_{ij}\}_{i,j}$ of matrices with a single 1 in the (i,j) entry and zeros elsewhere. One can check that $t \in T$ acts as

$$tE_{ij}t^{-1} = \frac{t_{ii}}{t_{jj}}E_{ij} = (\lambda_i - \lambda_j)(t)E_{ij}.$$

So $\Phi = \{\lambda_i - \lambda_j : i \neq j\}$. The standard Borel of upper-triangular matrices induces the ordering $\lambda_i \geqslant \lambda_j$ if and only if $i \leqslant j$. Thus $\Phi^+ = \{\lambda_i - \lambda_j : i < j\}$. The corresponding set of simple roots is $\Delta = \{\lambda_1 - \lambda_2, \lambda_2 - \lambda_3, \ldots, \lambda_{n-1} - \lambda_n\}$.

Example 8.2.3 ($n = 2$)**.** [do this]

Example 8.2.4 ($n = 3$)**.** [do this]

8.3 Weights

Let G be a semisimple group, and fix $T \subset B$ a maximal torus contained in a Borel subgroup. Let $\rho : G \to \mathrm{GL}(V)$ be a representation of G. Then $V = \bigoplus_{\alpha \in \mathrm{X}^*(T)} V_\alpha$, where

$$V_\alpha = \{v \in V : \rho(t)(v) = \alpha(t) \cdot v \text{ for all } t \in T\},$$

is the *weight space* of α. The weights of ρ are $\{\alpha : V_\alpha \neq 0\}$.

The *weight lattice* Λ_{w} is the smallest subgroup of $\mathrm{X}^*(T)$ containing all weights of linear representations of G. The *root lattice* of G, Λ_{r}, is the $\mathbf{Z}$-span of Φ. It is known that $[\Lambda_{\mathrm{w}} : \Lambda_{\mathrm{r}}] < \infty$.

If V is an irreducible representation of G, then there is a unique maximal weight α in the weights of V. This "highest weight" determines the representation. The set of possible α as highest weight vectors is $\Lambda_{\mathrm{w}} \cap W$. For each such weight α, let U_α be the unique irreducible representation of G with highest weight α.

Given any representation V of G, one can decompose V into irreducible representations "by hand." Find a maximal weight α of V. Then $V = U_\alpha \oplus V'$ for some V'. Continue the process.

8.4 Hilbert's invariants theorem

As before, let G be a reductive group and $\rho : G \to \mathrm{GL}(V)$ a linear representation. Let $\mathrm{Sym}(V^\vee)$ be the ring of polynomial function on V. Indeed, if $\{e_1, \ldots, e_n\}$ is a basis for V and $\{x_1, \ldots, x_n\}$ the dual basis, then $\mathrm{Sym}(V^\vee) = k[x_1, \ldots, x_n]$. Put $R = \mathrm{Sym}(V^\vee)$. Then G acts on R by substitutions: $(g \cdot f)(r) = f(g^{-1}r)$. The fundamental problem of invariant theory is to give a presentation for the k-algebra R^G consisting of G-invariant polynomial functions on V.

Theorem 8.4.1 (Hilbert)**.** *If V is a representation of a reductive group G, then $\mathrm{Sym}(V^\vee)^G$ is a finitely generated k-algebra.*

Example 8.4.2. Consider $G = \mathrm{SL}_2(k)$ acting on symmetric bilinear forms $q = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ by $g \cdot q = gq\,{}^{\mathrm{t}}g$. Then $\det q = ad - b^2 = \mathrm{Disc}(q)$ is an invariant. Is it the only one? Are the invariants sufficient to classify (separate) the orbits? What is the nature of the map $\operatorname{Spec} R \to \operatorname{Spec}(R^G)$ corresponding to $V \to V/G$? These questions, when V is replaced by an arbitrary variety is the subject of *Geometric invariant theory.*

Write $\mathrm{Sym}(V^\vee) = \bigoplus_{d \geqslant 0} \mathrm{Sym}^d(V^\vee)$. The group G acts on each $\mathrm{Sym}^d(V^\vee)$, which is a finite-dimensional k-vector space. Decompose each of these into highest-weight

representations. We get $R = \bigoplus_\alpha R_\alpha$, where R_α is the isotypical component of type α. Any $f \in R$ can be written as a sum $f = \sum f_\alpha$ of isotypical components. Each f_α is a sum of homogeneous polynomials, all in R_α.

The *Reynolds operator* $f \mapsto f^\natural$ sends f to its "trivial isotypical component." Then $(-)^\natural : R \to R^G$ is a projection. This map is k-linear and satisfies $(\varphi f)^\natural = \varphi f^\natural$ for $\varphi \in R^G$. Indeed, it is enough to show this for $f \in R[d]_\alpha$. Then $\varphi \cdot R[d]_\alpha \xrightarrow{\sim} k\varphi \otimes_k R[d]_\alpha$ via $\varphi \otimes f \mapsto \varphi \cdot f$ is an isoorphism of G-modules, and the right-hand side is clearly of type α. In fact, $(-)^\natural : R \to R^G$ is a homomorphism of R^G-algebras.

Let $R^G_+ = \bigoplus_{d>0} R[d]^G$. Consider the ideal $R \cdot R^G_+$ of R. By Hilbert's basis theorem there exist $f_1, \ldots, f_N \in R^G_+$ such that $R \cdot R^G_+ = \langle f_1, \ldots, f_N \rangle$. Without loss of generality each f_i is homogeneous of positive degree. We claim that $R^G = k[f_1, \ldots, f_N]$. Let $\varphi \in R^G$. To show that $\varphi \in k[f_1, \ldots, f_N]$ we may assume φ is homogeneous. We argue by induction on $\deg \varphi$, the trivial case $\deg = 0$ being obvious. Write $\varphi = \sum a_i f_i$, where without loss of generality $a_i \in R$ homogeneous with $\deg(a_i f_i) = \deg(\varphi)$. Then $\varphi = \varphi^\natural = \sum a_i^\natural f_i^\natural = \sum a_i^\natural f_i$. Without loss of generality the $a_i^\natural$ are homogeneous with $\deg(a_i^\natural f_i) = \det(\varphi)$. As $\deg(a_i^\natural) < \deg \varphi$, each $a_i^\natural \in k[f_1, \ldots, f_N]$, whence the result.

Example 8.4.3. As before, SL_2 acts on symmetric bilinear forms $q = \begin{pmatrix} a & b \\ b & d \end{pmatrix}$ by $g \cdot q = g q\, {}^{\mathrm{t}}g$. The orbits are represented by $\begin{pmatrix} * & \\ & 1 \end{pmatrix}$, $\begin{pmatrix} & \\ & 1 \end{pmatrix}$, and 0. The function $\det(q)$ is the only invariant. In other words,

$$R^G = k[\det(q)] \simeq k[x].$$

So $k^3/\mathrm{SL}_2 \simeq \mathbf{A}^1$, but the orbits are not in bijection with points of $\mathbf{A}^1$. So there are "not enough" invariant functions to separate orbits.

A similar phenomenon occurs for the action of SL_N on symmetric bilinear forms in N variables. There is only one invariant (the discriminant) but many orbits.

Theorem 8.4.4 (Chevalley-Iwahori-Nagata)**.** *The set of orbits always surjects onto V^G (for any action of a reductive group on an affine algebraic variety). t is bijective if and only if each orbit is Zariski-closed.*

In our example above, we have

$$\mathrm{SL}(2) \cdot \begin{pmatrix} & \\ & 1 \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ b & d \end{pmatrix} : ad - b^2 \neq 0 \right\}$$

which is not Zariski-closed.

9 Cubic rings

Recall that the goal of this conference is "counting arithmetic objects." The first sort of objects we might try to count are number fields. To do this, we would like a parameterization of all number fields of a given degree. For degree 2, this is easy. All quadratic fields are of the form $\mathbf{Q}(\sqrt{D})$, where D is a square-free integer. In general,

a degree n number field is of the form $\mathbf{Q}(\theta)$. Let f be the minimal polynomial of θ; it will be of the form $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbf{Z}[x]$. The problem is: determining whether two monic irreducible polynomials yield the same field is quite difficult.

Rather than counting number fields, we will try to count their maximal orders instead.

9.1 Some definitions

Definition 9.1.1. A *rank-n ring* is a (commutative, unital) ring R such that $R \simeq \mathbf{Z}^n$ as a $\mathbf{Z}$-module.

For $n = 2, 3, 4, 5$, we call these rings quadratic, cubic, quartic, and quintic. Typical cubic rings are maximal orders $R = \mathcal{O}_K$ in cubic number fields K. But non-maximal orders in cubic number fields are also cubic rings. For example, $\mathbf{Z}[3\sqrt[3]{2}] \subset \mathbf{Q}(2^{1/3})$ is a perfectly good cubic ring, with $\mathbf{Z}$-basis $\{1, 3\sqrt[3]{2}, 9\sqrt[3]{4}\}$. A still more pathological example is $\mathbf{Z}[X]/X^3$, which has $\mathbf{Z}$-basis $\{1, X, X^2\}$. If R is any quadratic ring, then $\mathbf{Z} \times R$ is another "pathological" cubic ring.

In the end, we'll count number fields by counting their maximal orders, which we will count by including them into a larger class of rank-n rings.

Definition 9.1.2. Let R be a rank-n ring. The *trace map* $\mathrm{tr} : R \to \mathbf{Z}$ is defined by $\mathrm{tr}(r) = \mathrm{tr}(\cdot r : R \to R)$.

If $\alpha_1, \ldots, \alpha_n$ is a $\mathbf{Z}$-basis of R, then the *discriminant* of R is $\mathrm{Disc}(R) = \det(\mathrm{tr}(\alpha_i\alpha_j)_{i,j}$

9.2 Quadratic rings

We know that quadratic rings (up to isomorphism) are in bijection with $\{D \in \mathbf{Z} : d \equiv 0, 1 \pmod 4\}$. Given a quadratic ring R, the corresponding integer is $D = \mathrm{Disc}(R)$. (It is an old theorem that $\mathrm{Disc}(R)$ is a square modulo 4.) Given such an integer D, the corresponding ring is

$$\mathbf{Z}[\tau]/\left(\tau^2 - D\tau + \frac{D^2 - D}{4}\right)$$

If, for example $D = 0$, the corresponding ring is $\mathbf{Z}[\tau]/\tau^2$.

From our parameterization of quadratic rings, it is easy to count them!

9.3 Cubic rings

Let R be a cubic ring with $\mathbf{Z}$-basis $\{1, W, T\}$. A key invariant is $WT = q + rW + sT$ for some $q, r, s \in \mathbf{Z}$. Choose a new $\mathbf{Z}$-basis $\{1, \omega, \theta\}$ with $\omega = W - s$ and $\theta = T - r$. In our new basis, $\omega\theta = n$ for some $n \in \mathbf{Z}$. We call such a basis *normalized*. Every basis of $R/\mathbf{Z}$ has a unique normalized lift to R.

We also write $\omega^2 = m - b\omega + a\theta$ and $\theta^2 = \ell - d\omega + c\theta$. The fact that multiplication on R is associative forces some conditions on $\{m, b, a, \ell, d, c\}$. The associativity relations

are exactly:

$$n = -ad$$
$$\ell = -bc$$
$$m = -ac.$$

This induces a bijection between isomorphism classes of cubic rings with choice basis of $R/\mathbf{Z}$ and quadruples $(a, b, c, d) \in \mathbf{Z}^4$. Clearly $\mathrm{GL}_2(\mathbf{Z})$ acts on cubic rings with basis of $R/\mathbf{Z}$; orbits of this action are isomorphism classes of cubic rings. The only thing here that is mysterious is the $\mathrm{GL}_2(\mathbf{Z})$-action on $\mathbf{Z}^4$.

If we write $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ and $g \in \mathrm{GL}_2(\mathbf{Z})$, then

$$(gf)(x, y) = \frac{1}{\det(g)} f\left(\begin{pmatrix} x & y \end{pmatrix} \cdot g\right).$$

In other words, if we identify $\mathbf{Z}^4$ with the space of cubic forms in two variables, then the action of $\mathrm{GL}_2(\mathbf{Z})$ is the natural one (with a twist by $\det^{-1}$). This parameterization of cubic rings goes back to [DF64]. It was also used in [DH69], and saw its first modern formulation in [GGS02].

Example 9.3.1. The cubic ring corresponding to $(a, b, c, d) = (0, 0, 0, 0)$ is pretty pathological – namely $\mathbf{Z}[\omega, \theta]/(\omega, \theta)^2$.

[. . . wasn't able to take notes to the end. . .]

10 Quartic and quintic rings

The plan is to first review a bit of the theory of cubic rings. We'll spend most of the time on quartic rings, then give a brief treatment of quintic rings.

10.1 Cubic rings revisited

We are interested in passing from cubic rings to forms geometrically. For simplicity, we assume $R = \mathcal{O}_k$ is the maximal order in a cubic number field. Consider the affine scheme $\mathrm{Spec}(\mathcal{O}_k)$; we want to embed this into $\mathbf{P}^1_{\mathbf{Z}}$. In general, a map $X \to \mathbf{P}^1_{\mathbf{Z}}$ is determined by a line bundle $\mathscr{L}$ on X together with two global sections that generate $\mathscr{L}$. For $\mathrm{Spec}(\mathcal{O}_k)$, this consists of an ideal $\mathfrak{a} \subset \mathcal{O}_k$ together with two elements generating $\mathfrak{a}$. We choose the inverse different $\mathfrak{D}^{-1} \subset \mathcal{O}_k$, and two elements of $\mathfrak{D}^{-1}$ having trace zero as our global sections. It turns out that the image of $\mathrm{Spec}(\mathcal{O}_k) \to \mathbf{P}^1_{\mathbf{Z}}$ is the zero-set of a binary cubic form.

If $\mathcal{O}_k$ is the maximal order in a number field k with $[k : \mathbf{Q}] = n$, the same construction embeds $\mathrm{Spec}(\mathcal{O}_k)$ into $\mathbf{P}^{n-2}_{\mathbf{Z}}$. It turns out that for any maximal order, a basis of $\ker(\mathrm{tr} : \mathfrak{D}^{-1} \to \mathbf{Z})$ will always generate $\mathfrak{D}^{-1}$ as a $\mathcal{O}$-module.

10.2 Quartic rings

Start as we did with cubic rings. Given a quartic ring Q, write down a $\mathbf{Z}$-basis $\{1, \alpha_1, \alpha_2, \alpha_3\}$ for Q. Just as with cubic rings, we can encode the multiplication of Q

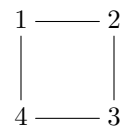
by

$$\alpha_i \alpha_j = c_{ij}^0 + \sum_{k=1}^{3} c_{ij}^k \alpha_k.$$

We could "shift" some of the basis elements in order to make some of the $c_{ij}^k = 0$. Associativity gives us some conditions on the c_{ij}^k. We're left with a "moduli scheme" for quartic rings with basis, but the polynomial relations between the c_{ij}^k are complicated enough that a direct, explicit approach, does not get you very far.

In the paper [WY92], Write and Yukie showed that quartic fields are parameterized by pairs of ternary and quadratic forms, moduli the natural action of $\mathrm{GL}_2(\mathbf{Q}) \times \mathrm{GL}_3(\mathbf{Q})$. Unfortunately this approach isn't very useful either. More recently, in the paper [Bha04c], Bhargava realized that the problem needed to be understood over $\mathbf{Z}$, and that cubic resolvent are essential.

We start with cubic resolvent fields $K/\mathbf{Q}$ with $[K : \mathbf{Q}] = 4$. Let $\widetilde{K}$ be the Galois closure of K; we assume $\mathrm{Gal}(\widetilde{K}/\mathbf{Q}) \simeq S_4$. The group S_4 has a canonical subgroup D_4 of index 3, consisting of permutations which are symmetries of the square



Galois theory gives us a subfield $K_3 \subset \widetilde{K}$ such that $[K_3 : \mathbf{Q}] = 3$; this is the *cubic resolvent* of K. For $k \in K$, let $k^{(1)}, k^{(2)}, k^{(3)}, k^{(4)}$ be the Galois conjugates of K. Then $k^{(1)}k^{(3)} + k^{(2)}k^{(4)}$ is an element of K_3. Put $\phi_{4,3}(k) = k^{(1)}k^{(3)} + k^{(2)}k^{(4)}$; this is a discriminant-preserving map $K \to K_3$.

Next we define the cubic resolvent of a ring. Let Q be a quartic ring, which for simplicity we assume is an order in a S_4-quartic field K.

Definition 10.2.1. A *cubic resolvent ring* of Q is a cubic ring R in the resolvent field K_3 such that

1. $\mathrm{Disc}(R) = \mathrm{Disc}(Q)$
2. for all $q \in Q$, $\phi_{4,3}(q) \in R$.

We have a quadratic map $\phi_{4,3} : Q \to R$. It descends to a map $\phi_{4,3} : Q/\mathbf{Z} \to R/\mathbf{Z}$; here the quotients are taken as $\mathbf{Z}$-modules, not as rings. The $\mathbf{Z}$ in the quotient is the sub-$\mathbf{Z}$-module generated by the multiplicative unit in the ring.

Exercise. Show that $\phi_{4,3}$ descends to a map $Q/\mathbf{Z} \to R/\mathbf{Z}$.

An element of $Q/\mathbf{Z}$ can be written as $\ell\alpha_1 + m\alpha_2 + n\alpha_3$ for $\ell, m, n \in \mathbf{Z}$. It's image under $\phi_{4,3}$ will be a linear combination of ω and θ, i.e.

$$\phi_{4,3}(\ell\alpha_1 + m\alpha_2 + n\alpha_3) = A(\ell, m, n)\omega + B(\ell, m, n)\theta.$$

The functions A and B are ternary quadratic forms with coefficients in $\mathbf{Z}$.

It is not clear whether every quartic ring even has a cubic resolvent. Even if it does, how many? Luckily, every quartic ring does have a cubic resolvent, but this resolvent is not necessarily unique. But maximal quartic rings (which include maximal quartic orders) have a unique cubic resolvent.

We want to parameterize isomorphism classes of pairs (Q, R), where Q is a quartic ring and R is a cubic resolvent of Q. The main result of [Bha04c] is that these are in bijection with $\mathrm{GL}_2(\mathbf{Z}) \times \mathrm{GL}_3(\mathbf{Z})$-classes of pairs of ternary quadratic forms. This bijection preserves discriminants, and allows you to detect prime splitting, automorphism groups, ... as the parameterization of cubic rings.

To be completely explicit, a *ternary quadratic form* is of the form

$$A(\ell, m, n) = a_{11}\ell^2 + a_{12}\ell m + a_{13}\ell n + \cdots + a_{33}n^2.$$

The GL_2 and GL_3 action can be written down explicitly. We identify the form A with a 3×3 matrix

$$\begin{pmatrix} a_{11} & a_{12}/2 & a_{13}/2 \\ a_{12}/2 & a_{22} & a_{23}/2 \\ a_{13}/2 & a_{23}/2 & a_{33} \end{pmatrix}$$

and similarly for B. The quantity $4\det(Ax + By)$ is a binary cubic form with coefficients in $\mathbf{Z}$. The corresponding cubic ring is the cubic resolvent ring in the pair (Q, R). What we haven't done is describe how to construct the quartic ring Q from A and B.

10.3 Geometric perspective

Write $A = a_{11}x^2 + a_{12}xy + a_{13}xz + \cdots$ and similarly for B. These cut out a subscheme of $\mathbf{P}^2$. Over $\mathbf{Q}$, we would expect this subscheme to have four points (over $\overline{\mathbf{Q}}$). The Galois conjugates of any such point p are among the four intersection points, so p is defined over a (at most) quartic extension of $\mathbf{Q}$.

Let's do this over $\operatorname{Spec} \mathbf{Z}$. A good reference is [Woo11b]. The two forms A, B over $\mathbf{Z}$ cut out a subscheme $V_{A,B}$ of $\mathbf{P}^2_{\mathbf{Z}}$. The functions on $V_{A,B}$ recover the quartic ring corresponding to A, B.

If $\mathcal{O}$ is the maximal order in a quartic field K, then using the inverse different we can embed $\operatorname{Spec}(\mathcal{O}) \hookrightarrow \mathbf{P}^2_{\mathbf{Z}}$. From [CE96], such subschemes of $\mathbf{P}^2_{\mathbf{Z}}$ are cut out by pairs of ternary quadratic forms.

10.4 Quintic rings

In [Bha08], it is proved that there is a bijection between isomorphism classes of pairs (R, S), where R is a quintic ring and S is a sextic resolvent of R, and $\mathrm{GL}_4(\mathbf{Z}) \times \mathrm{SL}_5(\mathbf{Z})$-orbits of quadruples of 5×5 skew-symmetric matrices (alternatively, quinary alternating forms).

We'll briefly describe the sextic resolvent of a quintic ring. At the level of Galois groups of fields, we're looking for an index 6 subgroup of S_5. Make a pentagon out of $\{1, 2, 3, 4, 5\}$. There are 6 ways to put them into two disjoint 5-cycles. The permutations that fix this decomposition into 2-cycles gives us such a subgroup.

The resolvent associated to (R, S) is a map $\Lambda^2 S^\vee \to R^\vee$. From a geometric perspective: over $\mathbf{Q}$, in $\mathbf{P}^3_{\mathbf{Q}}$, we have $A = A_1 x + A_2 y + A_3 z + A_4 w$, where A_1 is an alternating $5 \times t$ matrix. The matrix A has five skew-symmetric 4×4 minors. The determinant of such a minor is a a square, so the Pfaffian $\sqrt{\det(\text{minor})}$ is a quadratic form in 4 variables. So from A we get forms $Q_1, \ldots, Q_5$. These cut out a subvariety of $\mathbf{P}^3_{\mathbf{Q}}$, which is the analogue of the ring R.

We should expect the quadratic forms $Q_1, \ldots, Q_5$ to have no common zeros! But when a 5-tuple of quadratic forms come from Pfaffians of an alternating matrix, the subscheme they cut out is non-empty.

We would have $\mathrm{Spec}(\mathcal{O}) \subset \mathbf{P}^3_{\mathbf{Z}}$. There is a geometric analogue of this in [BE77] which shows that such a subvariety of $\mathbf{P}^3$ has to be cut out by 5 quadratic forms which come from Pfaffians of an alternating matrix.

11 How to count rings and fields I

In this lecture, we'll use the parameterization of cubic rings discussed in section 9 to count cubic rings. Recall that there is a bijection

$$\{\text{cubic rings}/\sim\} \xrightarrow{\sim} \{\text{int. bin. cubic forms}\}/\operatorname{GL}_2(\mathbf{Z})$$

which preserves discriminant. So counting cubic rings by discriminant is equivalent to counting cubic forms by discriminant. The corresponding problem for binary quadratic forms was solved by Gauss and Lipschitz (and first done rigorously by Mertens and Siegel).

11.1 Lipschitz principle and generalizations

Lipschitz's principle is that if R is some region in the Euclidean plane with area T, then the number of lattice points in R is approximately T. Better,

$$\#\{\text{lattice points in } R\} = T + O(T^{1/2}).$$

The implied constant in $O(T^{1/2})$ will depend on R. Here we assume the region R is "homogeneously expanding" via homothety. This principle, though completely elementary, was sufficient to count binary quadratic forms.

Davenport realized that in order to attack binary cubic forms, one needs a version of the Lipschitz principle where the region R is fixed (not necessarily expanding homogeneously).

Theorem 11.1.1 (Davenport)**.** *Let R be a bounded semi-algebraic region in $\mathbf{R}^n$ defined by at most k inequalities each of degree at most ℓ. Then the number of lattice points in R is $\mathrm{Vol}(R) + O_{k,\ell}(\max\{\mathrm{Vol}(\overline{R}), 1\})$, where $\mathrm{Vol}(\overline{R})$ denotes the greatest d-dimensional volume of a projection of R onto a d-dimensional coordinate hyperplane, $1 \leqslant d \leqslant n-1$.*

Proof. See the papers [Dav51, Dav64]. □

The following is Davenport's cubic version of Gauss' formula for the number of binary quadratic forms.

Theorem 11.1.2 (Davenport)**.** *Let $H(D)$ be the number of irreducible integer binary cubic forms, up to $\mathrm{GL}_2(\mathbf{Z})$-equivalence, having discriminant D. Then*

$$\sum_{0<D<X} H(D) = \frac{\pi^2}{72}X + O(X^{15/16})$$
$$\sum_{0<-D<X} H(D) = \frac{\pi^2}{24}X + O(X^{15/16}).$$

The proof uses Davenport's refined Lipschitz principle, along with analysis of the cusps using explicit inequalities.

Attempts to mimic Davenport's methods to count quartic forms fail because the inequalities involved are far too complicated to be analyzed explicitly. We will reprove Davenport's theorem using Davenport's principle, but without having to write down explicit inequalities.

11.2 Proof of Davenport's theorem

Let V be the space of binary cubic forms, $G = \mathrm{GL}(2)$. So $V(\mathbf{R})$ is the real vector space of binary cubic forms, and $V(\mathbf{Z})$ is the lattice of integer binary cubic forms. Similarly for $G(\mathbf{R})$ and $G(\mathbf{Z})$. We have a representation $G \to \mathrm{GL}(V)$ defined over $\mathrm{Spec}(\mathbf{Z})$.

Let E be a fundamental domain for the action of $G(\mathbf{Z})$ on $V(\mathbf{R})$. We can write

$$\begin{aligned}\sum_{0<D<X} H(D) &= \#\{x \in E \cap V(\mathbf{Z})^{\mathrm{irr}} : 0 < |\operatorname{Disc}(x)| < X\} \\ &= \#\{x \in E \cap V^+(\mathbf{Z})^{\mathrm{irr}} : |\operatorname{Disc}(x)| < X\}.\end{aligned}$$

We start by constructing any fundamental domains E. Fix $v \in V^+(\mathbf{R})$, and let $\mathcal{F}$ be a fundamental domain for the left action of $G(\mathbf{Z})$ on $G(\mathbf{R})$. Then the multiset $\mathcal{F}v = \{x = gv : g \in \mathcal{F}\}$ is the union of six fundamental domains for the action of $G(\mathbf{Z})$ on $V(\mathbf{R})$. Indeed,

$$(G(\mathbf{Z})\backslash G(\mathbf{R})) \cdot \big(G(\mathbf{R})\backslash V^+(\mathbf{R})\big) = G(\mathbf{Z})\backslash V^+(\mathbf{R}).$$

A brief consideration of stabilizers yields the number of fundamental domains.

A nice $\mathcal{F}$ to take was defined by Gauss:

$$\mathcal{F} = \{g \in \mathrm{GL}_2(\mathbf{R}) : g \cdot i \in \Omega\},$$

where $\mathrm{GL}_2(\mathbf{R})$ acts on the upper half-plane $\mathfrak{H} = \{z \in \mathbf{C} : \Im z > 0\}$ as usual, and

$$\Omega = \left\{ z \in \mathfrak{H} : -\frac{1}{2} \leqslant \Re z \leqslant \frac{1}{2} \text{ and } |z| \geqslant 1 \right\}.$$

Davenport's fundamental domain for $G(\mathbf{Z})$ on $V(\mathbf{R})$ occurs when $v = x^2y - xy^2$. Now we apply the Taski-Seidenberg theorem to conclude that $\mathcal{F}v$ is semi-algebraic.

This was generalized by Borel and Harish-Chandra in [BHC62] to arbitrary reductive groups over number fields. Note that $\mathcal{F} = N'A'K\Lambda$, where

$$\begin{aligned} N' &= \left\{ \begin{pmatrix} 1 & \\ n & 1 \end{pmatrix} : |n| \leqslant \frac{1}{2} \right\} \\ A' &= \left\{ \begin{pmatrix} t & \\ & t^{-1} \end{pmatrix} : t \geqslant \frac{\sqrt[4]{3}}{\sqrt{2}} \right\} \\ K &= \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : 0 \leqslant \theta < 2\pi \right\} \\ \Lambda &= \{\lambda \in \mathbf{R}^\times : \lambda > 0\}. \end{aligned}$$

Let B be a compact set in $V^+(\mathbf{R})$ that is the closure of a non-empty open set, on which $\operatorname{Disc} \geqslant 1$. We will allow v to vary in B.

Next, we need estimates on reducibility.

Lemma 11.2.1. *Let $R_X(v) = \mathcal{F}v \cap \{|\operatorname{Disc}| < X\}$. Then the number of reducible integral forms $ax^3 + \cdots + dy^3$ in $R_X(v)$ with $a \neq 0$ is $O(X^{3/4+\epsilon})$.*

Proof. In $R_X(v)$, we have $a = O(X^{1/4})$, $b = O(X^{1/4})$, $abc = O(X^{3/4})$, $abd = O(X^{3/4})$, ... If $d = 0$, at most $O(X^{3/4+\epsilon})$ such forms in $R_X(v)$. If $d \neq 0$, then $rx + sy \mid f(x,y) = ax^3 + \cdots + dy^3$, which implies $r \mid a, s \mid d$. Fixing a, b, d (for which there are $O(X^{3/4+\epsilon})$ choices) there are $O(X^\epsilon)$ choices for r and s. The fact that $f(-s, r) = 0$ determines c. The result follows. □

Finally we average. Recall that $|\operatorname{Disc}(v)|^{-1}\,dv$ is the unique (up to scalar) $G(\mathbf{R})$-invariant measure on $V(\mathbf{R})$. This lets us compute

$$\begin{aligned} N^+(X) &= \sum_{0<D<X} H(D) \\ &= \frac{\displaystyle\int_B \#\{x \in \mathcal{F}v \cap V(\mathbf{Z})^{\mathrm{irr}} : |\operatorname{Disc}(x)| < X\} \cdot |\operatorname{Disc}(v)|^{-1}\,dv}{\displaystyle 6\int_B |\operatorname{Disc}(v)|^{-1}\,dv} \\ &= \frac{1}{M}\int_{\mathcal{F}} \#\{x \in gB \cap V(\mathbf{Z})^{\mathrm{irr}} : |\operatorname{Disc}(x)| < X\}\,dg \\ &= \frac{1}{6}\operatorname{Vol}(R_X(v)) + O(X^{5/6}), \end{aligned}$$

the last equality coming from a uniform application of Davenport's inequality. Note that we have proved a stronger version of Davenport's result, namely one with an error term of $O(X^{5/6})$ instead of $O(X^{15/16})$. In [BST13], a more careful version of this proof shows that there is a second term of the form $cX^{5/6} + O(X^{3/4+\epsilon})$. So the error term is in many ways as good as possible.

12 Rings associated to binary n-ic forms, composition of $2 \times n \times n$ boxes and class groups

We'll focus especially on the case of binary quartic forms, and the parameterization of ideal classes. Good references are [Nak89, Sim01, Sim03, Sim05, Woo11c].

12.1 The construction

Let $f = a_0x^n + a_1x^{n-1}y + \cdots + a_ny^n$ be a binary n-ic form with the $a_i \in \mathbf{Z}$. We will construct a rank n ring R_f.

First we'll give an explicit construction. If $a_0 \neq 0$, consider the ring $\mathbf{Q}[\theta]/f(\theta, 1)$. This contains elements

$$\begin{aligned} 1 \\ a_0\theta \\ a_0\theta^2 + a_1\theta \\ \cdots \\ a_0\theta^{n-1} + a_1\theta^{n-2} + \cdots + a_{n-2}\theta. \end{aligned}$$

Let R_f be the $\mathbf{Z}$-module generated by these. It turns out that R_f is closed under multiplication, so it is a rank-n ring. For example,

$$\begin{aligned} (a_0\theta)^2 &= a_0^2\theta^2 \\ &= a_0(a_0\theta^2 + a_1\theta) - a_1(a_0\theta), \end{aligned}$$

so $(a_0\theta)^2 \in R_f$. The same phenomenon occurs for all of the additive generators of R_f.

A more highbrow construction is as follows. Inside $\mathbf{P}^1_{\mathbf{Z}}$ we have a closed subscheme V_f cut out by $f = 0$. The ring R_f is just $\mathrm{H}^0(V_f, \mathscr{O})$, the ring of regular functions on V_f (at least if not all the $a_i = 0$). When $a_0 \neq 0$, these functions are determined by their restriction to $\mathbf{A}^1_{\mathbf{Z}} \cap V_f$. The regular functions on $\mathbf{A}^1_{\mathbf{Z}}$ are polynomials in $\frac{x}{y}$, i.e. $\mathrm{H}^0(\mathbf{A}^1_{\mathbf{Z}}, \mathscr{O}) = \mathbf{Z}[\frac{x}{y}]$. We have for example

$$a_0\frac{x}{y} = -\left(a_1 + a_2\frac{y}{x} + \cdots\right).$$

We claim that the $\mathbf{Z}$-span of $1, a_0\frac{x}{y}, a_0(\frac{x}{y})^2 + a_1\frac{x}{y}, \ldots$ is the whole ring of regular functions on V_f. This recovers our explicit definition of R_f.

Proposition 12.1.1. *Let f be a binary n-ic form. Then*

- *If R_f is a domain, its fraction field is $\mathbf{Q}[\theta]/f(\theta, 1)$.*
- $\mathrm{Disc}(R_f) = \mathrm{Disc}(f)$.
- *R_f is a domain if and only if f is irreducible in $\mathbf{Q}[x, y]$*
- *R_f is a maximal order if and only if certain conditions modulo p^2 for every p are satisfied.*

- *If R_f is maximal, then a prime p splits in R_f as $f(x,y)$ factors modulo p.*

The special case $a_0 = 1$ yields monogenic rings $\mathbf{Z}[\alpha]$. The bad news is that for $n > 3$ we do not obtain all rank n rings (or all rank n maximal orders). Heuristically, this works as follows. Given a form f we can produce $V_f \subset \mathbf{P}^1_{\mathbf{Z}}$. A map to $\mathbf{P}^1$ is determined by a line bundle with two generating global sections. In the maximal order case, this is an ideal class of the ring with two generating elements. If we think of $R_f \subset \mathbf{Q}[\theta]/f(\theta,1)$, then the ideal class is $I_f = \langle 1, \theta \rangle$, which is a fractional ideal class unless $a_0 = 1$.

The general story goes as follows. There is a bijection between binary n-ic forms up to $\mathrm{GL}_2(\mathbf{Z})$-equivalence and rank-$n$ rings with an "ideal class" satisfying some conditions on the ideal class. We don't get all rank n rings because not all rings have ideal classes satisfying the conditions. For $n = 2$, the conditions on the ideal class are trivial, so we get all quadratic rings and all "ideal classes." (We put ideal class in quotation marks because we haven't defined such things for rings that are not integral domains.)

12.2 Geometric story of Gauss composition

Start with a binary quadratic form $f(x,y) = ax^2 + bxy + cy^2$ for $a, b, c \in \mathbf{Z}$. To this we can associate a quadratic ring with an ideal class. Geometrically, $\{f = 0\}$ cuts out a subscheme V_f of $\mathbf{P}^1_{\mathbf{Z}}$. The ring $\mathrm{H}^0(V_f, \mathscr{O})$ of regular functions on V_f is the associated ring, and the invertible sheaf coming from $V_f \hookrightarrow \mathbf{P}^1$ is the ideal class.

We work with pathological rings (having zero-divisors and nilpotents) because we want to be able to talk about the behavior of a form when it is reduced modulo p. Even if things are well-behaved over $\mathbf{Z}$, their reduction modulo p can be pathological.

We define ideal classes in general rings. Let C be an order in a quadratic field. If $\mathfrak{a}, \mathfrak{b}$ are ideals in the same class, then $\mathfrak{a} = k\mathfrak{b}$ for some $k \in C \otimes \mathbf{Q}$. The map $k : \mathfrak{a} \to \mathfrak{b}$ is an isomorphism of C-modules. Moreover, the converse holds: ideals that are isomorphic as C-modules are in the same ideal class. So we can replace the notion of an "ideal class" with an "isomorphism class of modules." But certainly not all C-modules come from ideal classes. For example, C^2 is not an ideal class. However, all modules isomorphic to $\mathbf{Z}^2$ as $\mathbf{Z}$-modules are ideal classes.

The reference for what follows is [Woo11a].

Theorem 12.2.1. *There is a bijection between $\mathrm{GL}_2(\mathbf{Z}) \times \mathrm{GL}_1(\mathbf{Z})$-classes of $ax^2 + bxy + cy^2$ and isomorphism classes of (C, M), where C is a quadratic ring and M is a C-module such that M is $\simeq \mathbf{Z}^2$ as a $\mathbf{Z}$-module and $\mathrm{tr} : C \to \mathbf{Z}$ is the same using multiplication on C or M.*

Recall that for $n = 2$ we get all ideal classes. For $n = 3$, the conditions on our "ideal class" force I_f to be the inverse different. So we get precisely isomorphism classes of cubic rings. For $n > 3$, the conditions on the ideal class become non-trivial. In particular, we don't get all maximal orders. For example, when $n = 4$, we get exactly the quartic rings with a monogenic cubic resolvent. A good reference is [Woo12b].

12.3 Parameterization of ideal classes of R_f

References here are [Bha04a, Bha04b, Woo14]. Let $A \in \mathbf{Z}^2 \otimes \mathbf{Z}^n \otimes \mathbf{Z}^n$, the space of "pairs (A_1, A_2) of $n \times n$ matrices with coefficients in $\mathbf{Z}$." Put $f = \det(A) = \det(A_1 x_1 + A_2 x_2)$; this is a binary n-ic form. The group $G \subset \mathrm{GL}_2(\mathbf{Z}) \times \mathrm{GL}_n(\mathbf{Z})$ consisting of (g, h) with $\det(g)\det(h) = 1$ acts on $\mathbf{Z}^2 \otimes \mathbf{Z}^n \otimes \mathbf{Z}^n$ by left and right multiplication on (A_1, A_2):

$$(g, h) \cdot (A_1, A_2) = (gA_1h, gA_2h).$$

Theorem 12.3.1. *For primitive, irreducible binary n-ic forms f with coefficients in $\mathbf{Z}$, there is a bijection between ideal classes of R_f and G-classes of $A \in \mathbf{Z}^2 \otimes \mathbf{Z}^n \otimes \mathbf{Z}^n$ with $\det(A) = f$.*

The ideal classes appearing in this theorem are not necessarily invertible. If we unravel this when f is monic, the theorem generalizes the classical result parameterizing ideal classes of monogenic orders by conjugacy classes of matrices. We will not cover the proof of this. We will describe the map from forms to ideal classes by restricting to the case when R_f is a maximal order.

Instead of viewing $\mathbf{Z}^2 \otimes \mathbf{Z}^n \otimes \mathbf{Z}^n$ as pairs of $n \times n$ matrices, we can view it as n-tuples $(a_1, \dots, a_n)$ of $2 \times n$ matrices. Given such a tuple, $a_1y_1 + a_2y_2 + \cdots + a_n$ is a $2 \times n$ matrix. Let $g_1, \dots, g_{\binom{2}{n}}$ be the determinants of maximal minors of the matrix $a_1y_1 + \cdots + a_ny_n$. Each g_i is a quadratic form in n variables. Let $V_{\boldsymbol{g}} \subset \mathbf{P}^{n-1}_{\mathbf{Z}}$ be the subscheme cut out by the g_i. If the g_i were generic, we would expect $V_{\boldsymbol{g}} = \varnothing$. In our case, the ring of functions on $V_{\boldsymbol{g}}$ is R_f. The map $V_{\boldsymbol{g}} \hookrightarrow \mathbf{P}^{n-1}_{\mathbf{Z}}$ comes from a line bundle with n global sections. The line bundle gives us an ideal class in R_f, which can be any ideal class in R_f.

When $n = 2$, there are three ways of "slicing up" an element of $\mathbf{Z}^2 \otimes \mathbf{Z}^2 \otimes \mathbf{Z}^3$. The same quadratic form f produces a ring R_f with ideal classes M_A, N_A, I_f. It turns out that $M_A N_A = I_f^{n-3}$, where I_f is the "standard" ideal coming from f. Since $n = 2$, $M_A N_A I_f = 1$.

Let M be an ideal in the ring R_f, and let N be an ideal such that $MN = I_f^{n-3}$. Then with the multiplication map $\mathbf{Z}^n \otimes \mathbf{Z}^n = M \otimes_{\mathbf{Z}} N \to I_f^{n-3} = \mathbf{Z}^n$, forget all but the last two coordinates. This gives us two $n \times n$ matrices, which define M and N to begin with. For $n = 2$, this can all be done quite explicitly.

13 The zeta functions attached to prehomogeneous vector spaces

13.1 Introduction

The first main example is the space of binary cubics. Let $G = \mathrm{GL}_2$, and let V be the space of binary cubics. We consider the standard twisted action of G on V:

$$(g \cdot x)(u, v) = \frac{1}{\det g} x\left(\begin{pmatrix} u & v \end{pmatrix} \cdot g\right).$$

We know that $\mathrm{Disc}(g \cdot x) = (\det g)^2 \mathrm{Disc}(x)$.

Definition 13.1.1 (Shintani)**.** Define the function $\xi^{\pm} : \{\Im z > 1\} \to \mathbf{C}$ by

$$\xi^{\pm}(s) = \sum_{\substack{x \in G(\mathbf{Z})\backslash V(\mathbf{Z}) \\ \pm \operatorname{Disc}(x) > 0}} \frac{\#\operatorname{Stab}(x)^{-1}}{\#\operatorname{Disc}(x)^s} = \sum_{\substack{R \text{ cubic ring} \\ \pm \operatorname{Disc}(R) > 0}} \frac{\#\operatorname{Aut}(R)^{-1}}{\operatorname{Disc}(R)^s}.$$

Theorem 13.1.2. *1. The function $\xi^{\pm}$ has an analytic continuation to $\mathbf{C}$. The function $(s-1)^2(s-\frac{5}{6})(s-\frac{7}{6})\xi^{\pm}(s)$ is entire.*

2. Indeed, $\xi^{\pm}(s)$ is holomorphic except for simple poles at $s = 1, \frac{5}{6}$, with explicit residue formulas.

3. There is a functional equation between $\xi^{\pm}(1-s)$ and $\widehat{\xi}^{\pm}(s)$.

Proof. 1,3. These follow from the general theory of prehomogeneous vector spaces.

2. This needs some careful analysis. □

As an application [see Frank's talk] if we write

$$\xi^{\pm}(s) = \sum_{n \geqslant 1} \frac{a_{\pm n}}{n^s},$$

then

$$\sum_{0<n<X} a_{\pm n} = r_1^{\pm} X + r_{5/6}^{\pm} \frac{X^{5/6}}{5/6} + O(X^{3/5+\epsilon}).$$

In [Shi75], Shintani separated the contributions of irreducible and reducible representations.

13.2 A proof of analytic continuation and functional equation for ζ

Let $f \in \mathcal{S}(\mathbf{R})$ be a smooth function that decays rapidly. The *Fourier transform* of f is

$$\widehat{f}(y) = \int_{\mathbf{R}} f(x) e^{2\pi i x y}\, dx.$$

One proof of the analytic continuation of the Riemann ζ function uses the Poisson summation formula:

$$\sum_{x \in \mathbf{Z}} f(x) = \sum_{y \in \mathbf{Z}} \widehat{f}(y).$$

A simple variation is that for $t \in \mathbf{R}^{\times}$, we have

$$\sum_{x \in \mathbf{Z}} f(tx) = |t|^{-1} \sum_{y \in \mathbf{Z}} \widehat{f}(t^{-1} y).$$

Simply let $f_t(x) = f(tx)$, and show using a change of variables that $\widehat{f_t}(y) = |t|^{-1} \widehat{f}(t^{-1} y)$. Applying Poisson summation to f_t yields the formula.

Definition 13.2.1 (local zeta)**.** Define

$$\Phi(f,s) = \int_{\mathbf{R}} |x|^{s-1} f(x)\, dx,$$

where f ranges over $\mathcal{S}(\mathbf{R})$ and $s \in \mathbf{C}$.

For $\Re s > 0$, the map $\Phi(-,s)$ is a functional on $\mathcal{S}(\mathbf{R})$.

Proposition 13.2.2. *1. $\Phi(f,s)$ has a meromorphic continuation to $\mathbf{C}$. Moreover, $\Gamma(s)^{-1}\Phi(f,s)$ is entire.*

2. $\Phi(\widehat{f},s) = c(s)\Phi(f,1-s)$, where $c(s) = (2\pi)^{-s}\Gamma(s)(e^{i\pi s/2} + e^{-i\pi s/2})$.

Let's use this proposition to prove the functional equation for ζ. We have

$$\begin{aligned}
\zeta(s)\Phi(f,s) &= \sum_{n\geqslant 1} \frac{1}{n^s} \int_{\mathbf{R}} |x| s f(x) \frac{dx}{|x|} \\
&= \sum_{n\geqslant 1} \int_{\mathbf{R}} \left|\frac{x}{n}\right|^s f(x) \frac{dx}{|x|} \\
&= \int_{\mathbf{R}} |x|^s \sum_{n\geqslant 1} f(nx)\, \frac{dx}{|x|} \\
&= \int_0^\infty |x|^s \sum_{n\in \mathbf{Z}\smallsetminus 0} f(nx)\, \frac{dx}{x}.
\end{aligned}$$

Recall that $d^\times t = \frac{dt}{t}$ is an invariant measure on $\mathbf{R}_+^\times$. Define

$$\begin{aligned}
Z(f,s) &= \int_0^\infty t^s \sum_{x\in \mathbf{Z}\smallsetminus 0} f(tx)\, d^\times t \\
Z_+(f,s) &= \int_1^\infty t^s \sum_{x\in \mathbf{Z}\smallsetminus 0} f(tx)\, d^\times t.
\end{aligned}$$

Lemma 13.2.3. *The function $Z_+(f,s)$ is entire.*

Proof. Since $t \geqslant 1$, the convergence is better when $\Re s$ is smaller. □

We compute:

$$\begin{aligned}
Z(f,s) - Z_+(f,s) &= \int_0^1 t^s \sum_{x\in \mathbf{Z}\smallsetminus 0} f(tx)\, d^\times t \\
&= \int_0^1 t^s \left(t^{-1} \sum_{y\in \mathbf{Z}\smallsetminus 0} \widehat{f}(t^{-1}y) + t^{-1}\widehat{f}(0) - f(0) \right) d^\times t \\
&= \int_0^1 t^{s-1} \sum_{y\in \mathbf{Z}\smallsetminus 0} f(t^{-1}x)\, d^\times t + \widehat{f}(0) \int_0^1 t^{s-1}\, d^\times t - f(0) \int_0^1 t^s\, d^\times t \\
&= Z_+(\widehat{f}, 1-s) + \frac{\widehat{f}(0)}{s-1} + \frac{f(0)}{s}.
\end{aligned}$$

We have shown that

$$\begin{aligned} Z(f,s) &= Z_+(\widehat{f}, 1-s) + \frac{\widehat{f}(0)}{s-1} + \frac{f(0)}{s} \\ &= Z(\widehat{f}, 1-s). \end{aligned}$$

The functional equation follows:

$$\begin{aligned} \zeta(1-s)\Phi(f,1-s) &= Z(f,1-s) \\ &= Z(\widehat{f}, s) \\ &= \zeta(s)\Phi(\widehat{f}, s) \\ &= \zeta(s)c(s)\Phi(f,1-s), \end{aligned}$$

the last equality following from the previous proposition. Canceling the $\Phi(f,1-s)$ yields the functional equation we're looking for.

We can also derive the residue of ζ:

$$\operatorname{res}_{s=1} Z(f,s) = \widehat{f}(0) = \int_{\mathbf{R}} f(x)\,dx.$$

But $Z(f,s) = \zeta(s)\Phi(f,s)$, and

$$\Phi(f,1) = \int_{\mathbf{R}} |x|^{1-1} f(x)\,dx = \int_{\mathbf{R}} f(x)\,dx.$$

It follows that $\operatorname{res}_{s=1}\zeta(s) =$.

13.3 Outline of proof properties of $\xi^{\pm}$

Define

$$\Phi_1(f,s) = \int_0^\infty x^{s-1} f(x)\,dx.$$

Our "main tool" is integration by parts, using $x^{s-1} = \left(\frac{x^s}{s}\right)'$. We compute:

$$\begin{aligned} \frac{1}{\Gamma(s)}\Phi_1(f,s) &= \frac{1}{s\Gamma(s)} \int_0^\infty (x^s)' f(x)\,dx \\ &= \frac{1}{\Gamma(s+1)} \left(x^s f(x)|_0^\infty - \int_0^\infty x^s f'(x)\,dx \right) \\ &= \frac{-1}{\Gamma(s+1)} \Phi_1(f', s+1). \end{aligned}$$

Repeat this n times, obtaining

$$\frac{1}{\Gamma(s)}\Phi_1(f,s) = \frac{(-1)^n}{\Gamma(s+n)}\Phi_1(f^{(n)}, s+n),$$

where the right-hand side is holomorphic for $\Re s > -n$. This proves part 1 of our main proposition.

For part 2 of the main proposition, compute

$$\begin{aligned}\Phi(f_t, s) &= \int_{\mathbf{R}} |x|^s f(tx)\, dx \\ &= |t|^{-s}\Phi(f, s).\end{aligned}$$

It follows that

$$\Phi(\widehat{f}_t, s) = |t|^{-1}\Phi(\widehat{f_{t^{-1}}}, s) = |t|^{s-1}\Phi(\widehat{f}, s).$$

The distributions $f \mapsto \Phi(\widehat{f}, s)$ and $f \mapsto \Phi(f, 1-s)$ satisfy the same transformation properties. There is a general theorem of uniqueness of relatively invariant distributions on homogeneous spaces. It implies that the two distributions coincide up to a constant $c = c(s)$. The theory of prehomogeneous vector spaces provides a way to generalize this.

There is a simpler proof of the functional equation that comes from a clever choice of f. Pick $f_0 \in C_c^\infty(\mathbf{R} \smallsetminus 0)$ and put $f = \frac{d}{dx} f_0 = f_0'$. By an elementary argument, we can prove

$$\widehat{f}(y) = \widehat{f_0}(y) = y\widehat{f_0}(y).$$

This implies $f(0) = \widehat{f}(0) = 0$. From the Poisson summation formula, we get

$$\sum_{x \in \mathbf{Z} \smallsetminus 0} f(x) = \sum_{y \in \mathbf{Z} \smallsetminus 0} \widehat{f}(y).$$

This implies $Z(f, s) = Z_+(f, s) + Z_+(\widehat{f}, 1-s)$, which is entire. Similarly

$$Z(f, s) = \zeta(s)\Phi(f, s) = \zeta(s)(s-1)\Phi(f_0, s-1).$$

For all s, there exists f_0 such that $\Phi(f_0, s-1) \neq 0$, hence $\zeta(s)(s-1)$ is entire.

13.4 Generalizing the result

Definition 13.4.1 (Sato)**.** Let G be an algebraic group over a field k. A finite-dimensional representation V of G is a *prehomogeneous vector space* if there exists $x \in V_{\bar{k}}$ such that the orbit $G_{\bar{k}} \cdot x \subset Z_{\bar{k}}$ is Zariski-open.

We say that a non-constant $P \in k[V]$ is a *relative invariant polynomial* if there exists $\chi \in \mathrm{X}^*(G)$ such that $P(g \cdot x) = \chi(g)P(x)$ for all $g \in G$, $x \in V$.

Let (G, V) be a prehomogeneous vector space defined over $\mathbf{R}$. Sato proved that if $P \in \mathbf{R}[V]$ is a relative invariant, then

$$\Phi^{(i)}(f, s) = \int_{V_{\mathbf{R}}^{(i)}} |P(x)|^s f(x)\, dx$$

has analytic continuation and satisfies a functional equation. In [SS74], Sato and Shintani proved that there is a zeta-function associated to (G, V).

The following are basic examples of prehomogeneous vector spaces:

Example 13.4.2. Let $G = \mathrm{GL}_1$, $V = \mathbf{A}^1$. Then the natural action of G on V has Zariski-open orbits. The function $P(x) = x$ leads to the standard Riemann zeta function $\zeta(s)$.

Example 13.4.3. If (G, V) is GL_2 acting in binary cubics and $P(x) = \mathrm{Disc}(x)$, then the associated zeta function is $\xi^{\pm}(s)$.

13.5 Shintani zeta function

Recall that for $G = \mathrm{GL}_2$ and V the space of binary cubics, the associated relative invariant polynomial is $P(x) = \mathrm{Disc}(x)$. Let

$$V' = \{x \in V : P(x) \neq 0\};$$

this consists of $x \in V$ having no multiple roots in $\mathbf{P}^1$. Let $S = \{x \in V : P(x) = 0\} = V \smallsetminus V'$. Then $V'_{\mathbf{C}}$ is a single $G_{\mathbf{C}}$-orbit in $V_{\mathbf{C}}$. Over the real numbers, $V'_{\mathbf{R}}$ breaks up into two orbits $V^+_{\mathbf{R}} \cup V^-_{\mathbf{R}}$ corresponding to $P(x) > 0$ and $P(x) < 0$.

Define a local zeta function by

$$\Phi^{\pm}(f, s) = \int_{V^{\pm}_{\mathbf{R}}} |P(x)|^{s-1} f(x)\, dx;$$

this converges when $\Re s > 1$. For $x, y \in V_{\mathbf{R}}$, define

$$\langle x, y\rangle = x_1 y_4 - \frac{1}{3} x_2 y_3 + \frac{1}{3} x_3 t_2 - x_4 y_1.$$

This bilinear form is invariant in the sense that $\langle gx, g^* y\rangle = \langle x, y\rangle$, where $g^* = (\det g)^{-1} g$. We can use this form to identify $V_{\mathbf{R}}$ with its dual, and define a Fourier transform

$$\widehat{f}(y) = \int_{V_{\mathbf{R}}} f(x) e^{2\pi i \langle x, y\rangle}\, dx.$$

Proposition 13.5.1. *1. The function*

$$\frac{1}{\Gamma(s)^2 \Gamma(s - \frac{1}{6}) \Gamma(s + \frac{1}{6})} \Phi^{\pm}(f, s)$$

is entire.

2. There exists $M(s)$ such that the following functional equations hold:

$$\Phi^+(\widehat{f}, s) = M(s) \Phi^+(f, 1 - s)$$
$$\Phi^-(\widehat{f}, s) = M(s) \Phi^-(f, 1 - s).$$

Proof. There exists a differential operator $Q(\frac{\partial}{\partial x})$ such that

$$Q\left(\frac{\partial}{\partial x}\right) e^{\langle x, y\rangle} = P(y) e^{\langle x, y\rangle}$$
$$Q\left(\frac{\partial}{\partial x}\right) P(x)^s = b(s) P(x)^{s-1},$$

where $b(s) = s^2 (s - \frac{1}{6})(s + \frac{1}{6})$. The rest is routine. □

13.6 Proof of analytic continuation and functional equation for $\xi^{\pm}(s)$

We define a "extended zeta function" by

$$\begin{aligned} Z(f,s) &= \int_{G(\mathbf{R})/G(\mathbf{Z})} |\det g|^{2s} \sum_{\substack{x\in V(\mathbf{Z})\\ P(x)\neq 0}} \Phi(gx)\,dg \\ &= \sum_{\substack{x\in G(\mathbf{Z})\backslash V(\mathbf{Z})\\ P(x)\neq 0}} \int_{G(\mathbf{R})} |P(gx)|^s f(gx)\,dg. \end{aligned}$$

Because of the invariance of the measure dg, the integrand on the second line depends only on the $G(\mathbf{R})$-orbit of x. But there are only two such orbits!

In general, if φ is a function on $V(\mathbf{R})$, we have a formula

$$\int_{G(\mathbf{R})} \varphi(gy)\,dy = \frac{m_{\pm}}{2\pi} \int_{G(\mathbf{R})\cdot x} \varphi(y)\,\frac{dy}{|P(y)|},$$

where $m_{\pm} \in \{2,6\}$ is the degree of the covering $G_{\mathbf{R}} \to V_{\mathbf{R}}^{\pm}$ defined by $g \mapsto gx$. Returning to our definition of $Z(f,s)$, we see that

$$\begin{aligned} \int_{G(\mathbf{R})} |P(gx)|^s f(gx)\,dg &= \frac{m_{\pm}}{2\pi} \int_{G(\mathbf{R})\cdot x} |P(y)|^s f(y)\,\frac{dy}{|P(y)|} \\ &= \frac{m_{\pm}}{2\pi} \Phi^{\pm}(f,s). \end{aligned}$$

Thus we have

$$Z(f,s) = \begin{pmatrix} \xi^{+}(s) & \xi^{-}(s) \end{pmatrix} \begin{pmatrix} \frac{3}{\pi}\Phi^{+}(f,s) \\ \frac{1}{\pi}\Phi^{-}(f,s) \end{pmatrix}.$$

Now choose $f_0 \in C_c^{\infty}(V_{\mathbf{R}}')$ and set $f = Q(\frac{\partial}{\partial x})f_0$. We get $\widehat{f}(y) = P(y)\widehat{f_0}(y)$, which implies $f|_{S_{\mathbf{R}}} = \widehat{f}|_{S_{\mathbf{R}}} = 0$, where S is the singular set. It follows that

$$Z(f,s) = Z_{+}(f,s) + Z_{+}(\widehat{f}, 1-s) = Z(\widehat{f}, 1-s),$$

an entire function. From an earlier proposition, we get the functional equation. By a similar argument, $\xi^{\pm}(s)b(s-1)$ is an entire function. Since $b(s-1) = (s-1)^2(s-\frac{5}{6})(s-\frac{7}{6})$, this recovers our main proposition on $\xi^{\pm}$.

The hard part is to analyze the integral

$$\begin{aligned} &X(f,s) - Z_{+}(f,s) - Z_{+}(\widehat{f}, 1-s) \\ &\quad = \int_{\substack{G(\mathbf{R})/G(\mathbf{Z})\\ |\det|\leqslant 1}} |\det g|^{2s} \left(|\det g|^{-2} \sum_{y\in V_{\mathbf{Z}}^{*}\cap S} \widehat{f}(g^{*}y) - \sum_{x\in V_{\mathbf{Z}}\cap S} f(gx) \right) dg. \end{aligned}$$

There are three types of singular points. Namely,

$$S = \{0\} \cup \{\text{triple root}\} \cup \{\text{distinct double root and single root}\}.$$

We can compute

$$X(f,s) - Z_+(f,s) - Z_+(\widehat{f}, 1-s)$$
$$= \left(\frac{\widehat{f}(0)}{2s-2} - \frac{f(0)}{2s} \right) \mathrm{Vol}(G^1_{\mathbf{R}}/G_{\mathbf{Z}}) + \int \cdots \left(\sum_{y \neq 0} \cdots - \sum_{x \neq 0} \cdots \right) dg.$$

In the cubic case, this computation is carried out in [Shi72], and the quartic case is done in [Yuk92]. The general case remains open. We can impose congruence conditions, obtaining:

$$\sum_{x \in x_0 + NV_{\mathbf{Z}}} f(x) = \frac{1}{N^4} \sum_{y \in \frac{1}{N} V^*_{\mathbf{Z}}} e^{2\pi i \langle x_0, y \rangle} \widehat{f}(y).$$

14 How to count rings and fields II

In section 11, we proved that $N^+(X) = \frac{1}{6} \mathrm{Vol}(\mathcal{F}v \cap \{|\operatorname{Disc}| < X\}) + O(X^{5/6})$. Recall that $N^+(X)$ is the number of irreducible positive binary cubic forms with $0 < \operatorname{Disc} < X$ up to $\mathrm{GL}_2(\mathbf{Z})$-equivalence.

14.1 The general theorem

Recall that we defined $R_X(v) = \mathcal{F}v \cap \{|\operatorname{Disc}| < X\}$.

Proposition 14.1.1. *Let $f \in C^0(V^+(\mathbf{R}))$, and $v_0 \in V^+(\mathbf{R})$. Then*

$$\int_{\mathcal{F}} f(gv_0)\, \mathrm{d}g = \frac{1}{2\pi} \int_{\mathcal{F}_{v_0}} f(x) |\operatorname{Disc}(x)|^{-1}\, \mathrm{d}x.$$

Proof. Because of uniqueness of invariant measures, this comes down to a simple Jacobian calculation. □

Now let $f = |\operatorname{Disc}|$ when $|\operatorname{Disc}| < X$, and 0 elsewhere. Then

$$\begin{aligned} \frac{1}{6} \mathrm{Vol}(R_X(v)) &= \frac{2\pi}{6} \int_1^{X^{1/4}} \lambda^4\, \mathrm{d}^\times \lambda \int_{\mathrm{GL}_2(\mathbf{Z}) \backslash \mathrm{GL}_2^{\pm}(\mathbf{R})} \mathrm{d}g \\ &= \frac{2\pi}{6} \frac{X}{4} \cdot \frac{\zeta(2)}{\pi} \\ &= \frac{\pi^2}{72} X. \end{aligned}$$

This explains the coefficient of X in our formula for $N^+(X)$. If we wanted to compute $N^-(X)$, then everything would be the same except that the $\frac{1}{6}$ would be replaced by $\frac{1}{3}$, so the coefficient of X would be $\frac{\pi^2}{24}$.

More generally, we've proved the following theorem:

Theorem 14.1.2. *Let $S \subset V(\mathbf{Z})$ be a $G(\mathbf{Z})$-invariant subset defined by congruence conditions modulo finitely many prime powers. If we let $N^+(S;X)$ be the number of positive-discriminant irreducible integer binary cubic forms in S with $|\operatorname{Disc}| < X$, up to $\mathrm{GL}_2(\mathbf{Z})$-equivalence, then*

$$N^+(S;X) = \frac{\pi^2}{72}\prod_p \mu_p(S)\cdot X + O(X^{5/6}).$$

where $\mu_p(S)$ is the p-adic density of S in $V(\mathbf{Z})$.

The implied constant in $O(X^{5/6})$ will depend on the set S. See [BST13] for details. What if S is defined by infinitely many congruence conditions? For example,if we want to count cubic fields via their maximal orders, maximality is determined by congruence conditions modulo p^2 for every p. So we need to know how the constant in $O(X^{5/6})$ changes as we vary S.

Let W_p be the set of integer binary cubic forms f such that R_f is *not* maximal at p, i.e. $R_f \otimes \mathbf{Z}_p$ is not a maximal order in $R_f \otimes \mathbf{Q}_p$.

Proposition 14.1.3. *With notation as above,*

$$N(W_p, X) = O(X/p^2),$$

with the implied constant is independent of p.

Proof. If $f \in W_p$ is a multiple of p, then f/p has discriminant $\operatorname{Disc}(f)/p^4$. The number of such f is $O(X/p^4)$. If $f \in W_p$ is not a multiple of p, we use the following lemma to replace f with an equivalent form satisfying $p \mid c$ and $p^2 \mid d$. Let $f' = (ap, b, c/p, d/p)$; this is $\mathrm{GL}_2(\mathbf{Q})$-equivalent to f via the matrix $\begin{pmatrix} 1 & \\ & p^{-1} \end{pmatrix}$. Moreover, $\operatorname{Disc}(f') = \operatorname{Disc}(f)/p^2$. This gives at most $O(X/p^2)$ such f, because $[f] \mapsto [f']$ is at most 3-to-1 when $f' \not\equiv 0 \pmod p$. □

Lemma 14.1.4. *Let $f \in W_p$ not be a multiple of p. Then there exists a binary cubic form $ax^3 + \cdots + dy^3$ that is $\mathrm{GL}_2(\mathbf{Z})$-equivalent to f such that $p \mid c$ and $p^2 \mid d$.*

Proof. Let $\mathcal{O}$ be an order strictly containing R_f. Then there exist bases $1, \omega, \theta$ and $1, \omega', \theta'$ of R_f and $\mathcal{O}$ respectively, such that $\omega = p^i\omega'$ and $\theta = p^j\theta'$ and $i, j \geqslant 0$ are distinct. If we write down the multiplication table for R_f, we get $p \mid b$ and $p^2 \mid a$, or $p \mid c$ and $p^2 \mid d$ depending on which of i and j are bigger. □

Theorem 14.1.5 (Davenport-Heilbronn)**.** *The number of cubic fields of positive discriminant $< X$ is $\frac{1}{12\zeta(3)}X + o(X)$. The number of cubic fields with negative discriminant $< X$ is $\frac{1}{4\zeta(3)}X + o(X)$.*

Proof. Let $U \subset V(\mathbf{Z})$ (resp. $U_p \subset V(\mathbf{Z}_p)$) be the set of binary cubic forms f such that R_f is maximal (resp. maximal at p). Then $U = \bigcap_p U_p$. We know that

$$N^+\left(\bigcap_{p<Y} U_p; X\right) = \frac{\pi^2}{72}\cdot\prod_{p<Y}\mu_p(U_p)\cdot X + o(X).$$

An elementary lemma gives $\mu_p(U_p) = \frac{(p^2-1)(p^3-1)}{p^5}$, so we have

$$\begin{aligned}\limsup_{X\to\infty} \frac{N^+(U;X)}{X} &\leqslant \lim_{X\to\infty} \frac{1}{X} N^+\left(\bigcap_{p<Y} U_p; X\right) \\ &= \frac{\pi^2}{72} \prod_{p<Y} \frac{(p^2-1)(p^3-1)}{p^5} \\ &= \frac{1}{12\zeta(3)}.\end{aligned}$$

The latter bit coming from $Y \to \infty$. Now we want to show the lim inf is sufficiently large. Since $U \subset \bigcap_{p<Y} U_p \subset U \cup \bigcup_{p\geqslant Y} W_p$,

$$\liminf_{X\to\infty} \frac{N^+(U;X)}{X} \geqslant \lim_{X\to\infty} \frac{1}{X} N^+\left(\bigcap_{p<Y} U_p; X\right) - \sum_{p\geqslant Y} O(1/p^2)$$

Let $Y \to \infty$ and the sum tends to zero. Since the limit approaches $1/12\zeta(3)$, we get the lower bound. □

What if we wanted to sieve to square-free discriminants, rather than just maximal orders? If a cubic ring has squarefree discriminant, it is maximal. But the converse does not hold. If any prime p totally ramifies in a cubic extension $K/\mathbf{Q}$, then $p^2 \mid \mathrm{Disc}(K)$. This is the only way $\mathrm{Disc}(K)$ can be divisible by a square. So counting square-free discriminant forms is equivalent to counting cubic fields in which no prime totally ramifies. On the side of forms, if R_f is maximal, the only way $\mathrm{Disc}(f)$ is not squarefree is if f is a constant times a cube of some linear form modulo p for some p. If R_f is not maximal, then $\mathrm{Disc}(R_f)$ is automatically divisible by a square. This is a condition modulo p^2, whereas $\mathrm{Disc}(R_f)$ being non-squarefree is a modulo p condition.

14.2 The geometric sieve

This is also known as the *closed point sieve*, or the *Ekedahl sieve*.

Theorem 14.2.1. *Let B be a bounded region in $\mathbf{R}^n$ with finite volume. Let Y be a subscheme of $\mathbf{A}^n_{\mathbf{Z}}$ of codimension k. Let $r, M > 0$. Then*

$$\#\{a \in rB \cap V(\mathbf{Z}) : a \in Y(\mathbf{F}_p) \text{ for some } p > M\} = O\left(\frac{r^n}{M^{k-1}\log M} + r^{n-k+1}\right).$$

Sketch of proof. We treat the case $k = 2$. Suppose Y is defined by integer polynomials $f(x_1,\ldots,x_n)$ and $g(x_1,\ldots,x_n)$. We can assume that f does not involve x_n (for example, by using elimination theory). Then the number of lattice points $(a_1,\ldots,a_n) \in rB$ where $f = 0$ or $g = 0$ is $O(r^{n-1})$. Suppose f and g are nonzero on a. Let's count all bad pairs (a,p). For $p \leqslant r$, this easy: the number of such (a,p) is $O(r^n/p2)$. Fix $a_1,\ldots,a_{n-1}$. If $p > r$, then $f(a_1,\ldots,a_n)$ has at most $O(1)$ of prime factors $p > r$. For each such prime p, there are $O(1)$ values of a_n such that $g(a_1,\ldots,a_n) \equiv 0 \pmod p$. So there are $O(r^{n-1})$ such $a = (a_1,\ldots,a_n)$. □

15 Heuristics for number field counts and applications to curves over finite fields

We'll discuss three things: local (p-adic) densities, applications to curves over finite fields, and heuristics for counting number fields. The motivating question is: how many number fields are there?

15.1 Local densities

What proportion of degree n number fields (ordered by $|\operatorname{Disc}|$) have 7 split completely? Of course, a similar question can be asked for any prime p. Note that we are fixing the prime and letting number fields vary, as opposed to the other way around. If we fix a number field and let primes vary, the splitting is controlled by the Čebotarev Density Theorem.

To be more precise, we are interested in

$$\lim_{X\to\infty} \frac{\#\{[K:\mathbf{Q}]=n, |\operatorname{Disc} K|<X : 7 \text{ splits completely}\}/\sim}{\#\{[K:\mathbf{Q}]=n : |\operatorname{Disc} K|<X\}/\sim}.$$

It is not clear that this limit exists, and it is not known if $n > 5$. It matters that we order by discriminant – if we ordered by some other invariant, the limit would change.

Let $K/\mathbf{Q}$ be a degree n number field. Write $K_7 = K \otimes \mathbf{Q}_7$; if $K = \mathbf{Q}[\theta]/f$ this is $\mathbf{Q}_7[\theta]/f$. If $(7) = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}$ in $\mathcal{O}_K$ and each $\mathfrak{p}_i$ has inertia degree f_i, then $K_7 = K_{\mathfrak{p}_1}\times\cdots\times K_{\mathfrak{p}_r}$, where each $K_{\mathfrak{p}_i}/\mathbf{Q}_7$ is a field extension of degree e_if_i. In the ring of integers of $K_{\mathfrak{p}_i}$, $(7) = \mathfrak{p}_i^{e_i}$. So K_7 carries all the splitting data (e_i, f_i) as well as more information about how 7 ramifies.

Example 15.1.1. The field $\mathbf{Q}_2$ has a unique unramified extension of degree 2 (in which $e = 1$, $f = 2$). It has six different ramified extensions of degree 2 (in which $e = 2$, $f = 1$). So there are six different ways a quadratic extension $K/\mathbf{Q}$ can ramify at 2.

Definition 15.1.2. An *étale $\mathbf{Q}_p$-algebra* is a finite direct product of finite field extensions of $\mathbf{Q}_p$. The *degree* of an étale $\mathbf{Q}_p$-algebra is its dimension as a $\mathbf{Q}_p$-vector space. If L is an étale $\mathbf{Q}_p$-algebra, define $\mathcal{O}_L$ as usual. The *discriminant* of L is $\operatorname{Disc}_{\mathbf{Z}_p}(\mathcal{O}_L) = \langle\det(\operatorname{tr}(\alpha_i\alpha_j))\rangle$. Put $|\operatorname{Disc}_{\mathbf{Z}_p}(\mathcal{O}_L)| = \#(\mathbf{Z}_p/\operatorname{Disc}(\mathcal{O}_L))$.

Let's look at all degree n étale $\mathbf{Q}_p$-algebras. There are only finitely many of these (for fixed n and p). In fact, if $p > n$, they are well-understood (and have an easy classification). We can ask: how often does each étale $\mathbf{Q}_p$-algebra occur as $K_p = K \otimes \mathbf{Q}_p$ for a random degree n number field K? The most naive guess would be a uniform distribution.

Example 15.1.3. Consider $p = 5, n = 2$. There are four étale $\mathbf{Q}_5$-algebras of degree 2. These are $\mathbf{Q}_5\times\mathbf{Q}_5$, the unique unramified extension of degree 2, and two ramified extensions.

Example 15.1.4. Consider $p = 5, n = 3$. Here there are six étale algebras of this type: $\mathbf{Q}_5\times\mathbf{Q}_5\times\mathbf{Q}_5$, $\mathbf{Q}_5\times$ (any quadratic extension), and one ramified and one unramified extension of degree 3.

Our naive "equidistribution" guess is wrong. First, ramified algebras are rare, because lots of ramification corresponds to a large discriminant. Also, objects occur "in nature" inversely proportionally to cardinality of their automorphism groups. For example, if our objects are isomorphism classes of cubic fields and "nature" is $\overline{\mathbf{Q}}$, then a Galois cubic field occurs once in $\overline{\mathbf{Q}}$, whereas non-Galois cubic fields occur three times, in keeping with the respective orders of automorphism groups. This principle is the basis for the Cohen-Lenstra Heuristics.

We will construct a measure on the set of étale $\mathbf{Q}_p$-algebras of degree n. Namely,

$$\mu_p(\{L\}) = \frac{1}{\#\operatorname{Aut}(L)|\operatorname{Disc} L|}.$$

This is *not* a probability measure. Let $\widetilde{\mu}_p$ be the normalized version of μ_p so that $\widetilde{\mu}_p$ is a probability measure on the set of étale $\mathbf{Q}_p$-algebras of degree n.

Heuristic 15.1.5. *A $\mathbf{Q}_p$-algebra L of degree n occurs as K_p for a random degree n number field K with probability $\widetilde{\mu}_p$.*

References for this are [Bha07, Mal02, Mal04]. It is known to hold when $n = 2, 3, 5$. For $n = 2, 3$, this follows from results of Davenport-Heilbront. For $n = 5$, this was done by Bhargava via his count of quintic extensions. For $n = 4$, the heuristic fails. About 16% of quartic fields have Galois closure with group D_4, and about 83% have Galois closure with group S_4. We can recover the heuristic for $n = 4$ by restricting to S_4-quartic extensions.

15.2 Points on curves over finite fields

There is a well-known analogy between number fields (e.g. $\mathbf{Q} \supset \mathbf{Z}$) and function fields over finite fields (e.g. $\mathbf{F}_q(X) \supset \mathbf{F}_q[X]$). For example, we should think of a quadratic field $\mathbf{Q}(\sqrt{D})$ as being analogous to an extension $\mathbf{F}_q(X)(\sqrt{X^3+1})/\mathbf{F}_q(X)$. We can talk about things like splitting of primes... in both cases.

The heuristic above corresponds to a conjecture for the proportion of curves over $\mathbf{F}_q$ with a degree n map to $\mathbf{P}^1$ that have k points for each k.

Theorem 15.2.1. *For $n = 3$, the conjecture is true. That is, when ordered by genus, the average number of points on a trigonal curve is $q + 2 - \frac{1}{q^2+q+1}$.*

See [Woo12a] for a proof of this.

15.3 Zeta functions

Let's return to the question: how many degree n number fields are there? As is so common number theory, we define a zeta function:

$$Z(s) = \prod_p \left(\sum_{\substack{K \text{ étale } \mathbf{Q}_p\text{-algebra} \\ [K:\mathbf{Q}_p]=n}} \frac{1}{\#\operatorname{Aut}(K)|\operatorname{Disc} K|^s} \right) = \sum_{n \geqslant 1} a_n n^{-s}.$$

The a_n don't literally count anything. But heuristically, the number of degree n number fields with $|\operatorname{Disc} K| < X$ should asymptotically be $\sum_{1 \leqslant n \leqslant X} a_n$. There are conjectures due to Malle and Bhargava on hour many degree n number fields there are with fixed Galois group. These conjectures are (at last naively) false. However, when interpreted more loosely (up to $O(X^\epsilon)$) they are known, e.g. for nilpotent groups, ...

16 Moduli space of rings

All rings in this lecture are commutative with unit. Fix an integer $n \geqslant 0$. The main question is: is there a scheme $\mathcal{A}_n$ such that $\mathcal{A}_n(\mathbf{Z})$ is in bijection with the set of isomorphism classes of rings A such that $A \simeq \mathbf{Z}^n$ as $\mathbf{Z}$-modules. Of course there is such a scheme! The set of rank-n rings is countable, and there are lots of schemes with $\aleph_0$ points defined over $\mathbf{Z}$.

A good source for what follows is the paper [Poo08].

16.1 Fine moduli space

A better question would be to ask for the existence of a scheme $\mathcal{A}_n$ such that for all rings k, the set $\mathcal{A}_n(k)$ is naturally in bijection with the set of isomorphism classes of k-algebras A such that $A \simeq k^n$ as a k-module. What do we mean by "natural"? For every ring homomorphism $k \to L$, we require the following diagram to commute:

$$\begin{array}{ccc} \mathcal{A}_n(k) & \longrightarrow & \{\text{rank-}n\ k\text{-algebras}\}/\sim \\ \downarrow & & \downarrow{\scriptstyle -\otimes_k L} \\ \mathcal{A}_n(L) & \longrightarrow & \{\text{rank-}n\ L\text{-algebras}\}/\sim \end{array}$$

If such a $\mathcal{A}_n$ exists, it follows from Yoneda's lemma that $\mathcal{A}_n$ is unique up to unique isomorphism. For, the functor

$$k \mapsto \{\text{isomorphism cases of rank-}n\ k\text{-algebras}\}$$

would be representable, and the standard argument shows that the representing object is unique.

Unfortunately, for $n \geqslant 2$ no such scheme exists. We'll show why in the case $n = 2$. The map $\mathcal{A}_2(\mathbf{R}) \to \mathcal{A}_2(\mathbf{C})$ must be injective. That is, the map "extension of scalars" from rank-2 $\mathbf{R}$-algebras to rank-2 $\mathbf{C}$-algebras should be injective on isomorphism classes. The rings $\mathbf{R}[x]/(x^2 - 1) \simeq \mathbf{R} \times \mathbf{R}$ and $\mathbf{R}[x]/(x^2 + 1) \simeq \mathbf{C}$ are certainly not isomorphic over $\mathbf{R}$, but they are both isomorphic to $\mathbf{C} \times \mathbf{C}$ when tensored with $\mathbf{C}$.

Essentially, what is going on here is that twists of an $\mathbf{R}$-algebra A are in bijection with $\mathrm{H}^1(\mathbf{R}, \operatorname{Aut} A_{\mathbf{C}})$. Since $A_{\mathbf{C}}$ can have nontrivial automorphisms (for example when $A = \mathbf{R}[x]/(x^2 + 1)$, there is no way that $\mathcal{A}_2(\mathbf{R}) \to \mathcal{A}_2(\mathbf{C})$ can be injective. This fits into the slogan that "objects with nontrivial automorphisms have no coarse moduli scheme." The canonical example is that elliptic curves have isomorphisms, preventing the existence of a fine moduli scheme for all elliptic curves. Just as with elliptic curves, the way to remedy the situation is to add structure.

16.2 Moduli space of based algebras

Unlike elliptic curves, it is very easy to prove that the moduli space of "based algebras" exists.

Theorem 16.2.1. *There exists a scheme $\mathcal{B}_n$ representing the functor*

$$k \mapsto \{(A, \boldsymbol{e})\}/\sim,$$

where A ranges over k-algebras (abstractly) isomorphic to k^n and $\boldsymbol{e} = (e_1, \ldots, e_n)$ is a k-basis for A.

Proof. A k-algebra A with basis $\boldsymbol{e} = (e_1, \ldots, e_n)$ is just a k-module $\bigoplus_i ke_i$ with multiplication table

$$e_i e_j = \sum_l c_{ij}^l e_l.$$

together with $1 = \sum d_i e_i$. The $n^3 + n$ elements $\{c_{ij}^l, d_i\}$ determine A, but commutativity, associativity, and unit force certain conditions on the c's and d's. These impose polynomial conditions. So

$$\mathcal{B}_n = \operatorname{Spec}\left(\mathbf{Z}[c_{ij}^l, d_m : 1 \leqslant i, j, l, m \leqslant n]/\text{relations}\right).$$

That is, $\mathcal{B}_n$ is the subscheme of $\mathbf{A}^{n^3+n}$ cut out by the polynomial relations encoding associativity, commutativity, and existence of unit. □

The rest of this lecture will be concerned with understanding the geometry of $\mathcal{B}_n$.

16.3 $\mathrm{GL}(n)$-basis

Reinterpret our scheme $\mathcal{B}$ as representing the functor that assigns to a ring k the set of isomorphism classes of pairs (A, ϕ), where A is a k-algebra and $\phi : A \xrightarrow{\sim} k^n$ is an isomorphism of k-modules. The group $\mathrm{GL}_n(k)$ on $\mathcal{B}_n(k)$ by

$$g \cdot (A, \phi) = (A, g \circ \phi).$$

This action is nicely functorial, so it gives an action of the group scheme $\mathrm{GL}(n)$ on the scheme $\mathcal{B}_n$.

Proposition 16.3.1. *For (A, ϕ) and (A', ϕ') in $\mathcal{B}_n(k)$, we have*

$$\mathrm{Isom}_{k\text{-}\mathsf{Alg}}(A, A') = \{g \in \mathrm{GL}_n(k) \text{ mapping } (A, \phi) \text{ to } (A', \phi')\}.$$

Proof. An isomorphism $\alpha : A \xrightarrow{\sim} A'$ corresponds to $g \in \mathrm{GL}_n(k)$ if $g \circ \phi = \phi' \circ \alpha$. □

Corollary 16.3.2. *For each k, there is a natural isomorphism*

$$\{\text{algebras of rank } n \text{ over } k\}/\sim = \mathrm{GL}_n(k)\backslash\mathcal{B}_n(k).$$

Corollary 16.3.3. *For any $(A, \phi) \in \mathcal{B}_n(k)$, there is a natural isomorphism* $\mathrm{Aut}_{k\text{-}\mathsf{Alg}}(A) \simeq \mathrm{Stab}_{\mathrm{GL}_n(k)}(A, \phi)$.

So the problem of classifying rank-n algebras comes down to understanding the scheme $\mathcal{B}_n$ together with its $\mathrm{GL}(n)$-action.

16.4 Example: $\mathcal{B}_3$ over $\mathbf{C}$

Let's classify rank 3 algebras A over $\mathbf{C}$. If A is a finite-dimensional $\mathbf{C}$-algebra, it will be artinian. As such, it will be a finite product of local artinian rings. Any finite-dimensional local artinian $\mathbf{C}$-algebra A has (by definition) a unique maximal ideal $\mathfrak{m}$, which is nilpotent. If $x_1, \ldots, x_d \in A$ map to a basis of $\mathfrak{m}/\mathfrak{m}^2$, then we will have a surjection $\mathbf{C}[x_1, \ldots, x_d]/(x_1, \ldots, x_d)^r \twoheadrightarrow A$ for some $r \gg 0$.

dim. of factors	algebra	Aut	dim(Aut)	dim(GL_3-orbit)
1,1,1	$\mathbf{C} \times \mathbf{C} \times \mathbf{C}$	S_3	0	9
2,1	$\mathbf{C}[x]/x^2 \times \mathbf{C}$	$x \mapsto ax$	1	8
3	$\mathbf{C}[x]/x^3$	$x \mapsto ax + bx^2$	2	7
	$\mathbf{C}[x, y]/(x, y)^2$	GL(2)	4	5

So $\mathcal{B}_3$ is 9-dimensional, and contains an open dense orbit isomorphic to $\mathrm{GL}(3)/S_3$. Each orbit is in the closure of a higher-dimensional orbit. We can see this by watching families of algebras degenerate. Consider $\mathbf{C}[x]/(x(x-t)(x-1))$. When $t \notin \{0, 1\}$, this algebra is étale over $\mathbf{C}$, namely $\mathbf{C} \times \mathbf{C} \times \mathbf{C}$. When $t = 0$, this algebra lies in the eight-dimensional orbit. Similar arguments yield the rest.

16.5 $\mathcal{B}_n(\mathbf{C})$ for all n

Consider the following table:

n	$\dim \mathcal{B}_n$
1	1
2	4
3	9
$\vdots$	
11	$\geqslant 129$
$\vdots$	
n	$\sim \frac{2}{27} n^3$

What's going on here? For small n, there is a large open orbit isomorphic to $\mathrm{GL}(n)/S_n$. Once n gets sufficiently large, there are high-dimensional components outside the étale locus.

The scheme $\mathcal{B}_n$ is smooth if and only if $n \leqslant 3$. The set $\mathrm{GL}_n(\mathbf{C})\backslash\mathcal{B}_n(\mathbf{C})$ is finite if and only if $n \leqslant 6$, as was proved in [Cha54, Dym66, Sup56]. The scheme $\mathcal{B}_n$ is irreducible if and only if $n \leqslant 7$; this is shown in [CEVV09]. The fact that $\dim(\mathcal{B}_n) \sim \frac{2}{27} n^3$ is due to [Ner87, Poo08].

These schemes can be used to construct lots of finite rings. Start with

$$\mathbf{Z}_2[x_1, \ldots, x_d]/(2, x_2, \ldots, x_d)^3$$

and then quotient out by a vector space. It is conjectured that asymptotically, all finite rings are of this form, and have "characteristic 8."

17 Zeta function methods

17.1 Motivation

The question is: what are zeta functions good for? Let $N_3^{\pm}(X)$ be the number of cubic fields K with $0 < \pm \operatorname{Disc}(K) < X$. Define

$$C^{\pm} = \begin{cases} 1 & + \\ 3 & - \end{cases}$$

$$K^{\pm} = \begin{cases} 1 & + \\ \sqrt{3} & - \end{cases}$$

Theorem 17.1.1. *The following holds:*

$$N_3^{\pm}(X) = C^{\pm} \frac{1}{12\zeta(3)} X + K^{\pm} \frac{4\zeta(1/3)}{5\Gamma(2/3)^3\zeta(5/3)} X^{5/6} + O(X^{2/3+\epsilon}).$$

We would like to understand how zeta functions can be used to provide such good error terms.

17.2 Definitions

As we have done before, put

$$V(\mathbf{Z}) = \{au^3 + bu^2v + cuv^2 + dv^3 : a, b, c, d \in \mathbf{Z}\}$$
$$\widehat{V}(\mathbf{Z}) = \{\cdots : 3 \mid b, c\}.$$

The group $\mathrm{GL}_2(\mathbf{Z})$ acts on both of these via

$$(\gamma \cdot f)(u, v) = \frac{1}{\det \gamma} f\left(\begin{pmatrix} u & v \end{pmatrix} \cdot \gamma\right).$$

Theorem 17.2.1. *There is a natural bijection*

$$\mathrm{GL}_2(\mathbf{Z}) \backslash V(\mathbf{Z}) \xrightarrow{\sim} \{\textit{cubic rings}\}/\sim .$$

Definition 17.2.2 (Shintani)**.** Put

$$\xi^{\pm}(s) = \sum_{x \in \mathrm{GL}_2(\mathbf{Z}) \backslash V^{\pm}(\mathbf{Z})} \frac{1}{\#\operatorname{Stab}(x)} |\operatorname{Disc}(x)|^{-s}$$
$$= \sum_{\substack{R \text{ cubic ring} \\ \pm \operatorname{Disc}(R) > 0}} \frac{1}{\#\operatorname{Aut}(R)} |\operatorname{Disc}(R)|^{-s}.$$

Theorem 17.2.3 (Shintani)**.** *The functions $\xi^{\pm}(s)$ have analytic continuation to $\mathbf{C}$ except for poles at $s = 1, \frac{5}{6}$, explicit residue formulas at these poles, and a functional equation*

$$\begin{pmatrix} \xi^+(1-s) \\ \xi^-(1-s) \end{pmatrix} = \Gamma\left(s - \frac{1}{6}\right) \Gamma(s)^2 \Gamma\left(s + \frac{1}{6}\right) \frac{3^{6s-2}}{2\pi^{4s}} \begin{pmatrix} \sin(2\pi s) & \sin(\pi s) \\ 3\sin(\pi s) & \sin(2\pi s) \end{pmatrix} \begin{pmatrix} \widehat{\xi}^+(s) \\ \widehat{\xi}^-(s) \end{pmatrix}$$

where $\widehat{\xi}^{\pm}$ are defined in terms of $\widehat{V}(\mathbf{Z})$ instead of $V(\mathbf{Z})$.

This was proved in [Shi72].

17.3 How analytic number theorists count

We'll see explicitly how analytic properties of $\xi^{\pm}$ translate into asymptotic estimates for the number of cubic rings.

Principle 17.3.1 (Perron's formula)**.** *Given any Dirichlet series $B(s) = \sum_{n \geqslant 1} b(n) n^{-s}$ which is absolutely convergent for $\Re s = 2$, then*

$$\sum_{n \leqslant X} b(n) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} B(s) X^s \frac{\mathrm{d}s}{s}.$$

For the $b(n)$ completely arbitrary, this is not very helpful. The idea is: for specific $b(n)$, shift the contour of this integral. For example, if we are trying to count integers less than X, apply Davenport's lemma to conclude that there are $X + O(1)$ integers between 0 and X. We define

$$\zeta(s) = \sum_{n \geqslant 1} n^{-s}.$$

By Perron's formula, we get

$$\sum_{1 \leqslant n < X} 1 = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \zeta(s) X^s \frac{\mathrm{d}s}{s}.$$

See if you can spot the mistake in the following computation:

$$\begin{aligned} \sum_{n \leqslant X} 1 &= \operatorname{res}_{s=1} + \operatorname{res}_{s=0} \left(\zeta(s) X^s \frac{\mathrm{d}s}{s} \right) + \frac{1}{2\pi i} \int_{-1-i\infty}^{1+i\infty} \zeta(s) X^s \frac{\mathrm{d}s}{s} \qquad (*) \\ &= X + \zeta(0) + (\text{error}). \end{aligned}$$

Does the integral in $(*)$ converse? For this, we need some bounds on ζ.

Proposition 17.3.2. *If $\sigma < 0$, we have $\zeta(-\sigma + it) \ll (1 + |t|)^{1/2+\sigma}$.*

Exercise. Use the functional equation and Stirling's approximation to prove the Proposition.

Inside the critical strip, controlling the behavior of ζ is a large problem. But outside $\{0 < \Re s < 1\}$, things are relatively straightforward.

We can use the above Proposition to show that the integral appearing in $(*)$ diverges.

17.4 The Landau method

A big principle in analytic number theory is that it is important to work with "smooth sums."

Proposition 17.4.1. *Let $b : \mathbf{N} \to \mathbf{C}$ and $B(s) = \sum b(n)n^{-s}$. Then*

$$\sum_{n<X} b(n)\left(1-\frac{n}{X}\right) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} B(s)\frac{X^s}{s(s+1)}\,\mathrm{d}s$$
$$\frac{1}{2}\sum_{n<X}\left(1-\frac{n}{X}\right)^2 = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} B(s)\frac{X^s}{s(s+1)(s+2)}\,\mathrm{d}s.$$

Exercise. Prove and generalize the Proposition.

These are all special cases of Mellin inversion. Using the Proposition, we continue our counting of integers:

$$\sum_{n<X}(X-n) = \frac{X^2}{2} - X + \frac{1}{2\pi i}\int_{-1-i\infty}^{1+i\infty}\zeta(s)\frac{X^{1+s}}{s(s+1)}\,\mathrm{d}x.$$

The integral is $O(X^{1/2+\epsilon})$. Moreover,

$$\sum_{n<X+Y}(X+Y-n) = \frac{X+Y}{2} - (X+Y) + O((X+Y)^{1/2+\epsilon}).$$

Subtract the first equation from the second, to conclude that

$$\sum_{n<X} 1 = X + O(X^{1/4+\epsilon}).$$

A more careful analysis of the integrals gets better error terms.

17.5 Why does the zeta function have such good analytic properties?

Definition 17.5.1 (Shintani)**.** The *global zeta function* is, for "nice test functions" $f : V_{\mathbf{R}} \to \mathbf{C}$,

$$Z(f,s) = \int_{\mathrm{GL}_2(\mathbf{R})/\,\mathrm{GL}_2(\mathbf{R})} |\det g|^{2s}\left(\sum_{x\in V(\mathbf{Z})\smallsetminus S} f(gx)\right)\,\mathrm{d}g,$$

where $S = \{x \in V(\mathbf{Z}) : \mathrm{Disc}(x) = 0\}$.

Proposition 17.5.2. *We have the following decomposition:*

$$Z(f,s) = \frac{1}{4\pi}\xi^+(s)\int_{V^+(\mathbf{R})} |\,\mathrm{Disc}(x)|^{s-1}f(x)\,\mathrm{d}x + \frac{\xi_-(s)}{2\pi}\int_{V^-(\mathbf{R})} |\,\mathrm{Disc}(x)|^{s-1}f(x)\,\mathrm{d}x.$$

Recall from Bhargava's lectures that the number of irreducible $\mathrm{GL}_2(\mathbf{Z})$-orbits in $G(\mathbf{Z})$ with $|\,\mathrm{Disc}\,| < X$ is

$$C\frac{\displaystyle\int_{\mathcal{F}} \#\{x \in gB \cap V(\mathbf{Z})^{\mathrm{irr}} : |\,\mathrm{Disc}(x)| < X\}\,\mathrm{d}g}{\displaystyle\int_B |\,\mathrm{Disc}(b)|^{-1}\,\mathrm{d}v}.$$

The sieve gives the estimate

$$N^{\pm}_{\max}(X) = \sum_{q \geqslant 1} \mu(q) N^{\pm}(q, x)$$

where $N^{\pm}_{\max}(X)$ is the number of maximal cubic rings R with $0 < \pm \operatorname{Disc}(R) < X$ and $N^{\pm}(q, X)$ is the number of cubic rings "nonmaximal at q." Define the q-nonmaximal zeta functions:

$$\xi^{\pm}_q(s) = \sum_{x \in \mathrm{GL}_2(\mathbf{Z}) \backslash V^{\pm}(\mathbf{Z})} \frac{1}{\#\operatorname{Stab}(x)} \Phi_q(x) |\operatorname{Disc}(x)|^{-s},$$

where $\Phi_q(x)$ is the characteristic function of the set of cubic forms nonmaximal at q. We get

$$\widehat{\xi}^{\pm}_q(s) = q^{8s-8} \sum_{x \in \mathrm{GL}_2(\mathbf{Z}) \backslash \widehat{V}_{\mathbf{Z}}} \frac{1}{\#\operatorname{Stab}(x)} \widehat{\Phi}_q(x) |\operatorname{Disc}(x)|^{-s},$$

where

$$\widehat{\Phi}_q(x) = \sum_{y \in V(\mathbf{Z}/q^2)} \Phi_q(y) \exp\left(\frac{2\pi i [x, y]}{q^2}\right),$$

and

$$[x, y] = x_r y_1 - \frac{1}{3} x_3 y_2 + \frac{1}{3} x_2 y_3 - x_1 y_4.$$

18 Counting Artin representations and modular form of weight one

We'll start by spending a little bit of time explaining what modular forms are.

18.1 Brief introduction to modular forms

Let $f = \sum_{n \geqslant 1} a_n q^n$ be a primitive cusp form of weight 1, level $N \geqslant 1$, and character $\varepsilon : (\mathbf{Z}/N)^{\times} \to \mathbf{C}^{\times}$. Then f is a holomorphic function $f : \mathfrak{H} \to \mathbf{C}$, where $\mathfrak{H} = \{z \in \mathbf{C} : \Im z > 0\}$ is the upper-half plane. The function f transforms by

$$f(\gamma z) = \varepsilon(d)(cz + d) f(z),$$

for all matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$. Moreover, f is "holomorphic at cusps."

Deligne and Serre proved that to asuch a modular form is attached a continuous Galois representation $\rho_f : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{C})$ such that $\operatorname{tr} \rho_f(\mathrm{fr}_\ell) = a_\ell$ and $\det \rho_f(\mathrm{fr}_\ell) = \varepsilon(\ell)$ for all primes $\ell \nmid N$. Since $G_{\mathbf{Q}}$ is compact and totally disconnected, the image of ρ_f is finite. Once we projectivize, the image of $\widetilde{\rho}_f : G_{\mathbf{Q}} \to \mathrm{PGL}_2(\mathbf{C})$ is one of

D_m	dihedral group of order $2m$
A_4	tetrahedral
S_4	octahedral
A_5	icosahedral

We call the representations with projective image A_4, S_4, A_5 exotic. They are quite rare, with the first one occurring when $N = 800$.

Conjecture 18.1.1. *The number of exotic forms of prime level N is $O(N^\epsilon)$.*

If the level N is prime, then only octahedral or icosahedral images occur. In this talk we will restrict to counting octahedral forms of prime level. Here is recent progress on the conjecture:

person	bound
Duke	$O(N^{7/8+\epsilon})$
Wang	$O(N^{5/6+\epsilon})$
M-V	$O(N^{4/5+\epsilon})$
Ganguly	$O(N^{3/4+\epsilon})$
Ellenberg	$O(N^{2/3+\epsilon})$

Theorem 18.1.2 (Bhargava-Ghate)**.** *Let $N_{\text{oct}}^{\text{prime}}(X)$ be the number of octohedral forms of prime level $< X$. Then*

$$N_{\text{oct}}^{\text{prime}}(X) = O(X/\log X).$$

This is proven in [BG09]. So on average, the number of octahedral forms of prime level is bounded by a constant.

Proof. The idea is to "count forms by counting forms." That is, we use the fact that octahedral modular forms of weight one correspond to Galois representations, which in turn correspond go quartic number fields, which come from quartic forms.

Step 1. It is enough to count (linear) Galois representations. The Artin conjecture says that there is a bijection between octahedral forms and isomorphism classes of $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{C})$ with $\rho(G_{\mathbf{Q}}) \simeq S_4$ and $\det \rho(c) = -1$. The direction $\Rightarrow$ was proved in [DS74], while $\Leftarrow$ is the Langlands-Tunnell theorem proved in [Lan80, Tun81].

One uses Serre's conjecture (proved in full generality in [KW09a, KW09b]) to prove the Artin conjecture. Choose $\mathcal{O} \subset \mathbf{C}$, the ring of integers in a number field such that $\rho(G_{\mathbf{Q}}) \subset \mathrm{GL}_2(\mathbf{C})$. We get a commutative diagram

$$\begin{array}{ccc} G_{\mathbf{Q}} & \xrightarrow{\rho} & \mathrm{GL}_2(\mathcal{O}) \\ & \searrow^{\bar\rho} & \downarrow \\ & & \mathrm{GL}_2(\overline{\mathbf{F}_p}). \end{array}$$

Serre's conjecture predicts that if $\bar\rho$ is odd and irreducible, then $\bar\rho \sim \bar\rho_g$ for a modular form g, where $g \in S_1(\Gamma_0(N), 4)$. Write $g = \sum b_n q^n$. Then $a_\ell \equiv b_\ell \pmod p$ for all $\ell \nmid N$. Since this works for infinitely many p, there exists g such that $a_\ell = b_\ell$. Thus ρ is modular.

Step 2. It is enough to count projective Galois representations. There is a surjection from isomorphism of odd $\rho : G_{\mathbf{Q}} \to \mathrm{GL}_2(\mathbf{C})$ with $\widetilde{\rho}(G_{\mathbf{Q}}) \simeq S_4$, to isomorphism classes of $\widetilde{\rho} : G_{\mathbf{Q}} \to \mathrm{PGL}_2(\mathbf{C})$ with $\widetilde{\rho}(G_{\mathbf{Q}}) \simeq S_4$ and $\widetilde{\rho}(c) \neq 1$. Surjectivity follows

from $\mathrm{H}^2(G_{\mathbf{Q}}, \mathbf{C}^\times) = 0$. The map is not injective: if χ is any character, Tate proved that $\rho \otimes \chi \mapsto \widetilde{\rho}$. But we can control the map when N is square-free. If $p \| N$, then

$$\rho|_{I_p} \sim \begin{pmatrix} \varepsilon_p & \\ & 1 \end{pmatrix}.$$

Choose χ with $\chi^{12} = 1$, and apply this with $\rho \otimes \chi$ instead of ρ. The new character is $(\varepsilon\chi^2)^{12}$. When $N = p$, there are only two such χ so that $\rho \otimes \chi$ map to the same $\widetilde{\rho}$. We get that $\varepsilon = \varepsilon_p$ is odd. Some technical manipulations yield that $\rho \mapsto \widetilde{\rho}$ is 2-to-1 when the level N is prime.

Step 3. It is enough to count quartic fields with Galois closure having group S_4. A projective representation $\widetilde{\rho} : G_{\mathbf{Q}} \to \mathrm{PGL}_2(\mathbf{C})$ with $\widetilde{\rho}(G_{\mathbf{Q}}) \simeq S - 4$ and $\widetilde{\rho}(c) \neq 1$ cuts out a field K over $\mathbf{Q}$ that is not totally real.

There is a key technical problem here. We want to count modular forms by level, but we usually count number fields by discriminant. Let N be the level of f. If K is the field corresponding to f and $D = \mathrm{Disc}(K)$, then we might have $N \neq D_K$. However, a prime $p \mid N_f$ if and only if $p \mid D_K$. Note that if $p \geqslant 5$, then either $p \| N_f$ or $p^2 \| N_f$. However, it is possible for $p^3 \| D_K$. Consider the following ramification table in the octahedral case (for minimal forms):

I_p	G_p	ram. in K	D_K	N_f	$p \equiv$
(12)	I_p	$1^2 11$	p	p	
(12)	(12), (34)	$1^2 2$	p	p^2	
(12)(34)	I_p	$1^2 1^{\underline{2}}$	p^2	p	
(13)(24)	(1234)	2^2	p^2	p	
(12)(34)	V_4	2^2	p^2	p^2	
(12)(34)	(12)(34)	$1^2 1$	p^2	p^2	
(123)	I_p	$1^3 1$	p^2	p	1 (mod 3)
(123)	S_3	$1^3 1$	p^2	p^2	2 (mod 3)
(1234)	I_p	1^4	p^3	p	1 (mod 4)
(1234)	D_4	1^4	p^3	p^2	3 (mod 4)

[... some notation I don't understand...]. Five times, the power of p in D_K is at most the power of p in N_f. The other five possibilities, this fails. For 4/5 of the time we can still control things. The other possibility is $p \equiv 1 \pmod 3$.

We use some facts about quartic fields. Consider field extensions $E- \supset K_6 \supset K_3 \supset \mathbf{Q}$ corresponding to the inclusions [... missed this part...].

The extension K_6/K_3 has Galois group S_4 if and only if

1. $K_3/\mathbf{Q}$ has Galois group S_3
2. $\mathrm{N}_{\mathbf{Q}}(\mathrm{Disc}(K_6/K_3)) = n^2$ for n square-free
3. $\mathrm{Disc}(K_4) = \mathrm{Disc}(k_3)n^2$
4. K_6/K_3 ramifies if and only if $p = 1^4$ or $1^2 1^2$ or 2^2.

We use the following theorem of Serre to simplify the table:

Ram. in K_4	$\lvert\operatorname{Disc}(K_4)\rvert$	level	$\lvert\operatorname{Disc}(K_3)\rvert$	n
$1^2 11$	p	p	p	1
1^4	p^3	p	p	p

Step 4. Use Bhargava's counting results in the quartic case, as well as some sieve methods. Recall, from [Bha04c] that isomorphism classes of maximal S_4-quartic orders are in bijection with $\mathrm{GL}_2(\mathbf{Z}) \times \mathrm{GL}_3(\mathbf{Z})$-classes of pairs of irreducible ternary quadratic forms (A, B). From [Bha05], we the number $N_4(X)$ of number fields of S_4-quartic fields of $|\operatorname{Disc}| < X$ is $O(X)$. A technical modification shows that the number $N_4^{\text{prime}}(X)$ of S_4-quartic fields of prime level $< X$ is $O(X/\log X)$. As a corollary,

$$\sum_{\substack{0<|\operatorname{Disc}(K_3)|<X \\ \text{prime}}} h_2^*(K_3) \leqslant C\frac{X}{\log X}.$$

We can finally count the number of octahedral of prime level $< X$ modular forms. □

Theorem 18.1.3 (Serre)**.** *In prime level, the discriminant of K_4 is either p^3 or $-p$.*

19 Binary quartic forms: bounded average rank of elliptic curves I

19.1 Introduction

Recall that every elliptic curve over $\mathbf{Q}$ can be written as $y^2 = x^3 + Ax + B$ for $A, B \in \mathbf{Q}$.

Theorem 19.1.1 (Mordell)**.** *The abelian group $E(\mathbf{Q})$ is finitely generated.*

So we can write $E(\mathbf{Q}) = T \oplus \mathbf{Z}^r$, where T is a finite abelian group, and $r = \operatorname{rk} E$ is the *rank* of E. We are going to study the average rank of elliptic curves. To do this, we need to order elliptic curves in some way.

The elliptic curve $E_{A<B} : y^2 = x^3 + Ax + B$ is isomorphic to $E_{n^4A,n^6B} : y^2 = x^3 + n^4Ax + n^6B$ for all $n \in \mathbf{Q}^\times$. So we can assume $A, B \in \mathbf{Z}$. If we assume that $p^4 \mid A \Rightarrow p^6 \nmid B$, then the A, B are unique. Thus there is a bijection between isomorphism classes of elliptic curves over $\mathbf{Q}$ and

$$\mathcal{E} = \{E_{A,B} : (A, B) \in \mathbf{Z}^2 \text{ and } p^4 \mid A \Rightarrow p^6 \nmid B\}.$$

We could also look at subfamilies of $\mathcal{E}$ cut out by (possibly infinitely many) congruence conditions. We could also look at "thin" families consisting of all quadratic twists of some elliptic curve.

19.2 Ordering elliptic curves

To talk about averages, we need to order elliptic curves in some way. The most obvious invariants to use are the discriminant and conductor. We have

$$\operatorname{Disc}(E_{A,B}) = \Delta(E_{A,B}) = 4A^3 - 27B^2.$$

The problem with ordering elliptic curves by discriminant is that we don't know how to count the number of elliptic curves with discriminant bounded by X. Essentially, the region $\{(x,y) \in \mathbf{R}^2 : 4x^2 - 27y^3 < X\}$ is noncompact, which makes the counting problem very hard. What is easier is to order elliptic curves by (naive) height:

$$H(E_{A,B}) = \max\{|4A^3|, 27B^2\}.$$

Let f be a function on elliptic curves. The *average* of f is

$$\operatorname{avg}(f) = \lim_{X\to\infty} \frac{\sum_{H(E)<X} f(E)}{\sum_{H(E)<X} 1}.$$

It's not hard to show that $\operatorname{avg}(\#T) = 1$. That is, on average an elliptic curve has no nontrivial torsion. This follows from Hilbert irreducibility.

Our question is: what can we say about the average rank?

Conjecture 19.2.1 (Goldfeld, Katz-Sarnak)**.** $\operatorname{avg}(\mathrm{rk}) = \frac{1}{2}$. *Moreover,* 50% *of elliptic curves have rank* 0 *and* 50% *have rank* 1.

The conjecture was originally made with elliptic curves ordered by conductor, but ordering by conductor, height and discriminant are all expected to yield the same average.

Given an elliptic curve E, we can define an L-function which we denote $L_E(s)$. The *completed L-function* $L_E^*(s)$ satisfies a functional equation

$$L_E^*(s) = \omega(E) L_E^*(1-s),$$

where $\omega(E)$, the *root number* of E, is ± 1. The *analytic rank* of E is the order of vanishing of L_E at $s = 1/2$. The *Birch and Swinnerton-Dyer conjecture* predicts that the analytic rank and algebraic rank of E are the same. But the analytic rank of E is also the analytic rank of L_E^*. So the BSD conjecture implies that $\mathrm{rk}(E)$ is even if and only if $\omega(E) = 1$. It's expected that $\omega(E)$ is equidistributed, i.e. half of elliptic curves have $\omega(E) = 0$ and half have $\omega(E) = 1$. Assuming BSD, it would follow that half of elliptic curves have even rank, and half have odd rank. We also expect the rank of an elliptic curve to be "as small as it can get away with," which would force 100% of elliptic curves with $\omega(E) = 0$ to have rank zero, and 100% of elliptic curves with $\omega(E) = 1$ to have rank one.

In [KS99], Katz and Sarnak studied the family of all L-functions of elliptic curves. Assuming GRH and BSD, the we have the following bounds on avg(rk):

[Bru92]	$\leqslant 2.14$
[HB04]	$\leqslant 2$
[You06]	$\leqslant 1.79$

More recently, we have the following theorem proven in [dJ02].

Theorem 19.2.2 (de Jong)**.** *For the family of all elliptic curves over* $\mathbf{F}_q(t)$*, the average rank is bounded above by* $\frac{7}{6} + \epsilon(q)$*, where* $\epsilon(q) \to 0$ *as* $q \to \infty$*.*

The method is to bound $\operatorname{avg}(\#\operatorname{Sel}_3)$.

19.3 Selmer groups

The fundamental idea is that from the canonical short exact sequence

$$0 \to E(\mathbf{Q})/p \to \mathrm{Sel}_p(E) \to Ш(E)[p] \to 0,$$

we get a bound on $\mathrm{rk}\, E$ in terms of $\mathrm{Sel}_p(E)$. Note that $\#(E(\mathbf{Q})/p) \geqslant p^{\mathrm{rk}\, E}$.

First, we want a parameterization of 2-Selmer elements of elliptic curves. More generally, if $\sigma \in \mathrm{Sel}_p(E)$. then we can think of σ as a locally soluble p-covering of E. Such a covering is a twist of $[p] : E \to E$. It will be a genus-one curve C isomorphic to E over $\overline{\mathbf{Q}}$, along with $C \to E$ such that the following diagram commutes:

$$\begin{array}{ccc} C & & \\ \downarrow & \searrow & \\ E & \xrightarrow{[p]} & E. \end{array}$$

See section 7 for more details. The covering C is *soluble* if $C(\mathbf{Q}) \neq \varnothing$, and it is *locally soluble* if $C(\mathbf{Q}_v) \neq \varnothing$ for all places v of $\mathbf{Q}$. Locally soluble p-coverings of E are in natural bijection with $\mathrm{Sel}_p(E)$, and soluble p-coverings are in bijection with $E(\mathbf{Q})/p$. For the rest of this lecture, we will concentrate on $p = 2$.

It turns out that a locally soluble 2-covering of E yields a binary quartic form over $\mathbf{Q}$. Conversely, a binary quartic form gives a 2-cover.

Let V be the space of binary quartic forms. The group $\mathrm{GL}(2)$ acts on V via

$$(\gamma \cdot f)(x, y) = \frac{1}{(\det \gamma)^2} f\left(\begin{pmatrix} x & y \end{pmatrix} \cdot \gamma\right).$$

The center $Z(\mathrm{GL}_2)$ acts trivially, so the action descends to one of $\mathrm{PGL}(2)$ on V. The ring of invariants is freely generated by two elements I, J, which have degree 2 and 3 respectively in the coefficients of the form.

Theorem 19.3.1 (Birch-Swinnerton-Dyer, Cremona-Fisher-Stoll)**.** *There is a bijection between 2-Selmer elements and the quotient* $\mathrm{PGL}_2(\mathbf{Q})\backslash V(\mathbf{Q})^{\mathrm{ls}}$*, where* $V(\mathbf{Q})^{\mathrm{ls}}$ *is the subset of locally soluble forms, in which* (A, B) *corresponds to* $I = -3 \cdot 2^6 A$ *and* $J = -272 \cdot 2^6 B$*, and* $A(f) = -I(f)/3 \cdot 2^4$ *and* $B(f) = -J(f)/27 \cdot 2^6$*.*

We will write this as $\mathrm{Sel}_2(E_{A,B}) = \mathrm{PGL}_2(\mathbf{Q})\backslash V(\mathbf{Q})^{\mathrm{ls}}_{A,B}$. The identity element of $\mathrm{Sel}_2(E_{A,B})$ corresponds to the orbit of binary quadratic forms with a rational linear factor.

We would like a parameterization of 2-Selmer elements that involves binary quartic forms with integral coefficients instead of just rational coefficients.

Lemma 19.3.2 (Birch, Swinnerton-Dyer)**.** *If* $f \in V(\mathbf{Q}_p)$*, then* $A(f), B(f) \in \mathbf{Z}_p$*. If* f *is* $\mathbf{Q}_p$*-solvable, then* $\mathrm{PGL}_2(\mathbf{Q}_p) \cdot f \cap V(\mathbf{Z}_p) \neq \varnothing$*.*

Theorem 19.3.3. *If* $f \in V(\mathbf{Q})$*, then* $A(f), B(f) \in \mathbf{Z}$*. If* f *is locally soluble, then* $\mathrm{PGL}_2(\mathbf{Q}) \cdot f \cap V(\mathbf{Z}) \neq \varnothing$*.*

Proof. For each prime, find $\gamma_p \in \mathrm{PGL}_2(\mathbf{Q}_p)$ so that $\gamma_p \cdot f \in V(\mathbf{Z}_p)$. Since PGL_2 has class number 1, there exists $\gamma \in \mathrm{PGL}_2(\mathbf{Q})$ so that $\gamma \cdot f \in V(\mathbf{Z}_p)$ for all p, hence $\gamma \cdot f \in V(\mathbf{Z})$. □

Theorem 19.3.4 (Birch, Swinnerton-Dyer and Cremona, Fisher, Stoll)**.** *The set* $\mathrm{Sel}_2(E_{A,B})$ *is naturally in bijection with* $\mathrm{PGL}_2(\mathbf{Q})\backslash V(\mathbf{Z})^{\mathrm{ls}}_{A,B}$.

Define the *height* of a binary quartic form to be

$$H(f) = \max\{4|A(f)|^3, 27B(f)^2\}.$$

So the goal is to count $\mathrm{PGL}_2(\mathbf{Q})$-equivalence classes of integral, locally soluble binary quartic forms with height bounded by X.

19.4 First step

The goal is to count $\mathrm{PGL}_2(\mathbf{Z})$-orbits of $V(\mathbf{Z})^{\mathrm{irr}}_{H<X}$. The method is extremely similar to how Bhargava counted binary cubic forms. First we construct a fundamental domain $\mathcal{F}_X$ for the action of $\mathrm{PGL}_2(\mathbf{Z})$ on $V(\mathbf{R})_{H<X}$. Next we estimate $\#\{\mathcal{F}_X \cap V(\mathbf{Z})\}$ using averaging.

We begin by finding a fundamental set for $\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})$. Over any field k in which 6 is invertible, $\mathrm{PGL}_2(k)\backslash V(k)^{k-\mathrm{sol}}_{A,B}$ is in bijection with $E_{A,B}(k)/2$. Given $A, B \in \mathbf{R}$, the set $\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})^{\mathrm{ls}}_{A,B}$ is in bijection with $E_{A,B}(\mathbf{R})/2$. But the group of $\mathbf{R}$-valued points of an elliptic curves is easy to understand. It is $\mathbf{Z}/2 \times S^1$ or S^1, depending on whether the discriminant is positive or negative. It follows that $E_{A,B}(\mathbf{R})/2$ has either 1 or 2 elements as $\Delta(E_{A,B}) < 0$ or $\Delta(E_{A,B}) > 0$.

If $\Delta(E_{A,B}) < 0$, then $\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})_{A,B}$ is a singleton, and any f in the set will have exactly 2 real roots. If $\Delta(E_{A,B}) > 0$, there are two orbits, one consisting of forms with two real roots, and one consisting of forms with positive-definite binary quartic forms. Define

$$\begin{aligned} V(\mathbf{R})^{(0)} &= \{f \in V(\mathbf{R}) \text{ with 4 real roots}\} \\ V(\mathbf{R})^{(1)} &= \{f \in V(\mathbf{R}) \text{ with 2 real roots}\} \\ V(\mathbf{R})^{(2+)} &= \{f \in V(\mathbf{R}) \text{ positive definite}\}. \end{aligned}$$

A fundamental set for $\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})^{(i)}$ for $i \in \{0, 1, 2+\}$ is $\{f$ having invariants $A, B\}$. We obtain

$$\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})^{(i)}_{H<1} = R_1^{(i)},$$

with, for example,

$$R_1^{(0)} = \{x^3y + Axy^3 + By^4, (A, B) \in \mathbf{R}^2, \Delta(E_{A,B}) > 0, H(E_{A,B}) < 1\}.$$

For general height, we scale:

$$\mathrm{PGL}_2(\mathbf{R})\backslash V(\mathbf{R})^{(i)}_{H<X} = X^{1/6} R_1^{(i)} = R_X^{(i)}.$$

So $R_X^{(i)}$ is a fundamental domain for the action of $\mathrm{PGL}_2(\mathbf{R})$ on $V(\mathbf{R})^{(i)}_{H<X}$. We want a fundamental domain for the action of $\mathrm{PGL}_2(\mathbf{Z})$. Choose $\mathcal{F} = \mathrm{PGL}_2(\mathbf{Z})\backslash \mathrm{PGL}_2(\mathbf{R})$; then $\mathcal{F} \cdot R_X^{(i)}$ is an n_i-fold cover of $\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{R})^{(i)}_X$. It turns out that $n_1 = 2$ and $n_0 = n_{2+} = 4$.

It follows that

$$\#\,\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{(i),\mathrm{irr}}_{H<X} = \frac{1}{n_i}\#(\mathcal{F}\cdot R_X^{(i)}\cap V(\mathbf{Z})^{\mathrm{irr}}).$$

As with binary cubic forms, we average:

$$\begin{aligned}
\#\,\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{(i),\mathrm{irr}}_{H<X} &= \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{G_0}\#\left(\mathcal{F}gR_X^{(i)}\cap V(\mathbf{Z})^{\mathrm{irr}}\right)\,\mathrm{d}g\\
&= \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{\mathcal{F}}\#\left(gG_0R_X^{(i)}\cap V(\mathbf{Z})^{\mathrm{irr}}\right)\,\mathrm{d}g\\
&= \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{\mathcal{F}}\mathrm{Vol}\left(gG_0R_X^{(i)}\right)\,\mathrm{d}g + O(X^{3/4})\\
&= \frac{1}{n_i\,\mathrm{Vol}(G_0)}\int_{G_0}\mathrm{Vol}\left(\mathcal{F}gR_X^{(1)}\right)\,\mathrm{d}g + O(X^{3/4})\\
&= \frac{1}{n_i}\,\mathrm{Vol}\left(\mathcal{F}R_X^{(i)}\right) + O(X^{3/4}).
\end{aligned}$$

We can summarize all of this in the following theorem.

Theorem 19.4.1 (Bhargava, Shankar)**.**

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{(i),\mathrm{irr}}_{H<X}\right) = \frac{1}{n_i}\,\mathrm{Vol}\left(\mathcal{F}R_X^{(i)}\right) + O(X^{3/4}).$$

As an easy corollary, the average rank of elliptic curves is bounded. We can compute this as

$$\frac{1}{n_i}\,\mathrm{Vol}\left(\mathcal{F}R_X^{(i)}\right) = \frac{1}{n_i}|J|\,\mathrm{Vol}(\mathcal{F})\,\mathrm{Vol}\left(R_X^{(i)}\right) + O(X^{5/6})$$

So the number of elliptic curves with height $< X$ is some constant multiple of $X^{5/6}$. Thus $\mathrm{avg}(\#\,\mathrm{Sel}_2)$ is bounded, whence avg(rk) is bounded.

In section 22, we'll derive an explicit bound for rk(avg).

20 Selmer groups and heuristics I

The results of this lecture are joint work with Eric Rains. See the paper [PR12] for details.

20.1 Introduction

For simplicity, we'll work over $\mathbf{Q}$, but most of these results work over any number field, and even over any global field. Let $G = \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, v be a place of $\mathbf{Q}$, and

$$\mathbf{Q}_v = \begin{cases}\mathbf{R} & v=\infty\\ \mathbf{Q}_p & v=p\end{cases}$$

Let $\mathbf{A} = \prod_v'(\mathbf{Q}_v, \mathbf{Z}_v)$ be the restricted direct product of the $\mathbf{Q}_v$. That is,

$$\mathbf{A} = \left\{ (a_v) \in \prod_v \mathbf{Q}_v : a_v \in \mathbf{Z}_v \text{ all but finitely many } v \right\}$$

The ring of adeles $\mathbf{A}$ is locally compact.

Theorem 20.1.1 (Mordell)**.** *Let E be an elliptic curve over $\mathbf{Q}$. Then $E(\mathbf{Q})$ is finitely generated.*

Proof. This is essentially the only known. First, show that $E(\mathbf{Q})/n$ is finite for some $n \geqslant 2$. Then use height functions to conclude that $E(\mathbf{Q})$ is finitely generated. □

The only known proof of finiteness of $E(\mathbf{Q})/n$ passes through the finiteness of Selmer groups. Start with the exact sequence

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

which we think of as an exact sequence of étale sheaves over $\mathrm{Spec}(\mathbf{Q})$, i.e. $G_{\mathbf{Q}}$-modules. Taking global sections (G-invariants), we get a long exact sequence

$$0 \to E[n](\mathbf{Q}) \to E(\mathbf{Q}) \xrightarrow{n} E(\mathbf{Q}) \to \mathrm{H}^1(\mathbf{Q}, E[n]) \to \mathrm{H}^1(\mathbf{Q}, E) \to \cdots$$

We write $\mathrm{H}^1(\mathbf{Q}, E[n])$ for either of the two (naturally isomorphic) groups

$$\mathrm{H}^1(G_{\mathbf{Q}}, E[n](\overline{\mathbf{Q}})) \qquad \mathrm{H}^1(k_{\text{ét}}, E[n]).$$

This gives us an exact sequence

$$0 \to E(\mathbf{Q})/n \to \mathrm{H}^1(\mathbf{Q}, E[n]) \to \mathrm{H}^1(\mathbf{Q}, E) \to 0.$$

Unfortunately, $\mathrm{H}^1(\mathbf{Q}, E[n])$ is (provably) infinite. But we can try to pin down the image of $E(\mathbf{Q})/n$ inside $\mathrm{H}^1(\mathbf{Q}, E[n])$. Working at each place v, we get a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E(\mathbf{Q})/n & \longrightarrow & \mathrm{H}^1(\mathbf{Q}, E[n]) & \longrightarrow & \mathrm{H}^1(\mathbf{Q}, E) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\
0 & \longrightarrow & E(\mathbf{A})/n & \xrightarrow{\alpha} & \mathrm{H}^1(\mathbf{A}, E[n]) & \longrightarrow & \mathrm{H}^1(\mathbf{A}, E) & \longrightarrow & 0.
\end{array}$$

Here we write $\mathrm{H}^1(\mathbf{A}, E[n])$ for either $\mathrm{H}^1(\mathbf{A}_{\text{ét}}, E[n])$ or $\prod_v' \mathrm{H}^1(\mathbf{Q}_v, E[n])$ with respect to the $\mathrm{H}^1(\mathbf{Z}_v, \mathcal{E}[n])$, where $\mathcal{E}$ is the Néron model for E. The two groups are isomorphic – see [uh] for details. Define

$$\mathrm{Sel}_n E = \beta^{-1}(\mathrm{im}\,\alpha)$$
$$Ш(E) = \ker(\gamma).$$

The group $\mathrm{Sel}_n E$ is finite and, in principle, computable. As an exercise, show that there is a natural exact sequence

$$0 \to E(\mathbf{Q})/n \to \mathrm{Sel}_n E \to Ш(E)[n] \to 0.$$

For an elliptic curve $E : y^2 = x^3 + Ax + B$ with $A, B \in \mathbf{Z}$ minimal, put $h(E) = \max(|A|^3, |B|^2)$. Let $\mathcal{E}$ be the set of isomorphism classes of elliptic curves over $\mathbf{Q}$. For $X \in \mathbf{R}$, put $\mathcal{E}_{<X} = \{E \in \mathcal{E} : h(E) < X\}$.

Definition 20.1.2. For $S \subset \mathcal{E}$, define

$$\mathrm{Prob}(S) = \lim_{X \to \infty} \frac{\#(S \cap \mathcal{E}_{<X})}{\#\mathcal{E}_{<X}}.$$

Given a prime p, what is $\mathrm{Prob}(\dim \mathrm{Sel}_p E = s)$? In [HB94], Heath-Brown proved that as E ranges over quadratic twists of $y^2 = x^3 - x$, we have

$$\mathrm{Prob}(\dim \mathrm{Sel}_2(E) - 2 = s) = \prod_{j \geqslant 0} \frac{1}{1 + 2^{-j}} \prod_{j=1}^{s} \frac{2}{2^j - 1}.$$

This is not one of the standard (e.g. Bernoulli or Poisson) distributions. The distribution turns out to model "random maximal isotropic subspaces of a quadratic space."

20.2 Maximal isotropic subspaces

Let $V = \mathbf{F}_p^{2n}$, and let

$$Q(\boldsymbol{x}, \boldsymbol{y}) = Q(x_1, \ldots, x_n, y_1, \ldots, y_n) = x_1 y_1 + \cdots + x_n y_n.$$

The pair (V, Q) is a hyperbolic quadratic space over $\mathbf{F}_p$. To any quadratic space we can associate a symmetric bilinear form $\langle \cdot, \cdot \rangle : V \times V \to \mathbf{F}_p$ by

$$\langle v, w \rangle = Q(v + w) - Q(v) - Q(w).$$

Given a subspace $Z \subset V$, define the "orthogonal complement" $Z^\perp$ as

$$Z^\perp = \{v \in V : \langle v, z \rangle = 0 \text{ for all } z \in Z\}.$$

It's not actually a complement, because it could be that $Z \cap Z^\perp \neq 0$. We do have $V^\perp = 0$. If $\dim Z = m$, then $\dim Z^\perp = 2n - m$.

We call a subspace $Z \subset V$ *isotropic* if $Q|_Z = 0$. We say Z is *maximal isotropic* (or *Lagrangian*) if $Q|_Z = 0$ and $Z = Z^\perp$. The second condition forces $\dim Z = n$.

Example 20.2.1. The subspace $\{(x_1, \ldots, x_n, 0, \ldots, 0)\} \subset V$ is maximal isotropic.

Define the *orthogonal Grassmannian* $\mathrm{OGr}_n \subset \mathrm{Gr}_{n,2n}$ by

$$\mathrm{OGr}_n(\mathbf{F}_p) = \{\text{maximal isotropic } Z \subset V\}.$$

Choose $Z, W \in \mathrm{OGr}_n(\mathbf{F}_p)$ uniformly at random. We get a random variable $\dim_{\mathbf{F}_p}(Z \cap W)$.

Conjecture 20.2.2 (Poonen, Rains)**.** *As E ranges over all elliptic curves over* $\mathbf{Q}$,

$$\mathrm{Prob}(\dim \mathrm{Sel}_p E = s) = \lim_{n \to \infty} \mathrm{Prob}(\dim(Z \cap W) = s).$$

Some basic combinatorics and linear algebra show that the predicted distribution is agrees with the theorem of Heath-Browns when $p = 2$.

20.3 Selmer groups and maximal isotropic subspaces

Why should we expect this conjecture be true? Is there anything in the arithmetic of elliptic curves that would lead us to expect Selmer groups to behave like intersections of maximal isotropic subspaces?

Our goal is to show that $\mathrm{Sel}_p E$ actually is the intersection of two maximal isotropic subspaces of an infinite-dimensional quadratic space.

For the rest of this lecture, for simplicity assume p is an odd prime. Then to choose a quadratic function on an $\mathbf{F}_p$-vector space V is equivalent to choosing a bilinear form $\langle \cdot, \cdot \rangle$ on V. If we have $\langle \cdot, \cdot \rangle$, we set

$$Q(x) = \frac{1}{2}\langle x, x \rangle.$$

20.4 Local fields

Let E be an elliptic curve over $\mathbf{Q}_v$. Let $V_v = \mathrm{H}^1(\mathbf{Q}_v, E[p])$. This is a finite-dimensional $\mathbf{F}_p$-vector space. To prove this, set $L = \mathbf{Q}_v(E[p]) \supset \mathbf{Q}_v(\mu_p)$. The extension $L/\mathbf{Q}_v$ is finite Galois. We get a short exact sequence

$$0 \longrightarrow \mathrm{H}^1(\mathrm{Gal}(L/\mathbf{Q}_v), E[p]) \xrightarrow{\inf} \mathrm{H}^1(\mathbf{Q}_v, E[p]) \xrightarrow{\mathrm{res}} \mathrm{H}^1(L, E[p]).$$

To prove that $\mathrm{H}^1(\mathbf{Q}_v, E[p])$ is finite-dimensional, it is sufficient to prove that the groups on the ends of this sequence are finite. The group on the left is obviously finite. The group on the right is $\hom_{\mathrm{cts}}(G_L, \mathbf{Z}/p)^{\oplus 2}$. Since $\mathbf{Z}/p$ is abelian, local class field theory tells us that $\hom_{\mathrm{cts}}(G_L, \mathbf{Z}/p)$ is $\hom_{\mathrm{cts}}(L^\times/p, \mathbf{Z}/p)$, a finite group. Actually, since $\mu_p \subset L$, we didn't need class field theory – we could have just used Kummer theory.

The Weil pairing $e : E[p] \times E[p] \to \boldsymbol{\mu}_p \subset \mathbf{G}_{\mathrm{m}}$ induces via the cup-product a pairing

$$\langle \cdot, \cdot \rangle : V_v \times V_v \to \mathrm{H}^2(\mathbf{Q}_v, E[p] \otimes E[p]) \to \mathrm{H}^2(\mathbf{Q}_v, \mathbf{G}_{\mathrm{m}}) = \mathrm{Br}(\mathbf{Q}_v) \hookrightarrow \mathbf{Q}/\mathbf{Z}.$$

We will write Q for the induced quadratic map $Q_v : V_v \to \mathbf{R}/\mathbf{Z}$. Let $W_v = \mathrm{im}(E(\mathbf{Q}_v)/p \hookrightarrow \mathrm{H}^1(\mathbf{Q}_v, E[p]))$; it turns out that $W_v = \mathrm{H}^1(\mathbf{Z}_v, \mathcal{E}[p])$, where $\mathcal{E}$ is the Néron model for E over $\mathbf{Z}_v$. The following theorem is proved in [O'N02] using Tate local duality.

Theorem 20.4.1 (O'Neil)**.** *The subspace $W_v \subset V_v$ is maximal isotropic with respect to Q_v.*

20.5 Global fields

Let E be an elliptic curve over $\mathbf{Q}$. For each place v, let $V_v = \mathrm{H}^1(\mathbf{Q}_v, E[p]) \supset W_v$. Define

$$V = \prod\nolimits' (V_v, W_v) = \mathrm{H}^1(\mathbf{A}, E[p])$$
$$Q = \sum Q_v : V \to \mathbf{R}/\mathbf{Z}.$$

Then (V, Q) is a "quadratic locally compact group." Recall our diagram:

$$\begin{array}{ccc} \mathrm{Sel}_p E & \hookrightarrow & \mathrm{H}^1(\mathbf{Q}, E[p]) \\ & & \downarrow \beta \\ E(\mathbf{A})/p & \xrightarrow{\alpha} & \mathrm{H}^1(\mathbf{A}, E[p]) \end{array}$$

in which $\mathrm{Sel}_p E = \beta^{-1}(\mathrm{im}\,\alpha)$.

Theorem 20.5.1. *1.* $\mathrm{im}(\alpha)$ *and* $\mathrm{im}(\beta)$ *are maximal isotropic.*
2. β *is injective.*
3. $\mathrm{im}(\alpha) \cap \mathrm{im}(\beta) = \beta(\mathrm{Sel}_p E) \simeq \mathrm{Sel}_p(E)$

Ingredients of proof. 1. use the fact that $\mathrm{im}(\alpha) = \prod W_v$ and each W_v is isotropic. The fact that $\mathrm{im}(\beta)$ is maximal isotropic follows from the 9-term Poitou-Tate exact sequence in global duality and the Brauer reciprocity law.

2. Use the Čebotarev density theorem, and the fact that the Sylow p-subgroups $\begin{pmatrix} 1 & * \\ & 1 \end{pmatrix} \subset \mathrm{GL}_2(\mathbf{F}_p)$ is cyclic.

3. This follows from 1, 2, and the definition of $\mathrm{Sel}_p E$. The group $\mathrm{H}^1(\mathbf{A}, E[p])$ is self-dual via our pairing $\langle \cdot, \cdot \rangle$. □

So $\mathrm{Sel}_p(E)$ "is" the intersection of the two maximal isotropic subspaces $E(\mathbf{A})/p$ and $\mathrm{H}^1(\mathbf{Q}, E[p])$ of $\mathrm{H}^1(\mathbf{A}, E[p])$. The theorem is false for abelian varieties, and for p^2-torsion.

20.6 Passing to p^∞-torsion

Recall that there is a canonical exact sequence

$$0 \to E(\mathbf{Q})/n \to \mathrm{Sel}_n E \to Ш(E)[n] \to 0$$

for all integers $n \geqslant 1$. Take $n = p^e$ and pass to the direct limit. We get a short exact sequence

$$0 \to E(\mathbf{Q}) \otimes \mathbf{Q}_p/\mathbf{Z}_p \to \mathrm{Sel}_{p^\infty} E \to Ш(E)[p^\infty] \to 0. \qquad (*)$$

These are all $\mathbf{Z}_p$-modules with finitely-generated Pontryagin dual. (That is, this is an exact sequence of co-finitely generated $\mathbf{Z}_p$-modules.) So each is of the form $(\mathbf{Q}_p/\mathbf{Z}_p)^s \oplus T$, where T is a finite abelian p-group. We call s the *corank* of the group. Conjecturally, $Ш(E)$ is finite, so $Ш(E)[p^\infty]$ has corank 0. It would follow that $\mathrm{rk}\, E$ is the p-corank of $\mathrm{Sel}_{p^\infty}(E)$. In section 26, we'll give a probabilistic model for the sequence $(*)$.

21 Rational points on curves

The goal is to discuss how one can show that a hyperelliptic curve has no rational points.

21.1 Hyperelliptic curves

Let k be a field of characteristic not 2.

Definition 21.1.1. A *hyperelliptic curve* over k is the smooth projective curve associated to an affine curve of the form $y^2 = f(x)$ with $f \in k[x]$ squarefree (and $\deg f \geqslant 5$).

We will write $C : y^2 = f(x)$; C means the projective curve associated to the subvariety $V(y^2 - f)$ of the affine plane. We can define the projective curve C as follows: homogenize f as $f(x) = F(x, 1)$ with $F \in k[x, z]$ squarefree and homogeneous of even degree. Then $y^2 = F(x, z)$ describes $C \subset \mathbf{P}^2_{1,g+1,1}$ when $\deg F = 2g + 2$.

There are points at infinity. If $f(x) = f_{2g+2}x^{2g+2} + \cdots + f_1x + f_0$, then $F(x, z) = f_{2g+2}x^{2g+2} + \cdots + f_0z^{2g+2}$. If $\deg f$ is odd, there is just one point $\infty = (1 : 0 : 0)$ at infinity. If $\deg f$ is even, then there are two points at infinity: $\infty_{\pm s} = (1 : \pm s : 0)$, where $f_{2g+2} = s^2$. If f_{2g+2} is not a square in k, these points at infinity will not be defined over k.

It is known that C has genus g. The set of k-rational points of C is

$$C(k) = \{(\xi, \eta) \in k^2 : \eta^2 = f(\xi)\} \cup \begin{cases} \{\infty\} & \deg f \text{ odd} \\ \{\infty_s, \infty_{-s}\} & \deg f \text{ even, } f_{2g+2} \text{ a square} \\ \varnothing & \text{otherwise} \end{cases}$$

The condition $\deg f \geqslant 5$ implies $g \geqslant 2$. Faltings' theorem tells us that $C(\mathbf{Q})$ is finite. Our motivating problem is: determine $C(\mathbf{Q})$ explicitly for given C. This is wide open at the present. Heuristically, we expect 100% of hyperelliptic curves of genus g to have no rational points. This needs some explanation. For a hyperelliptic curve $C : y^2 = f$ defined over $\mathbf{Q}$, we can assume $f \in \mathbf{Z}[x]$. We can then order hyperelliptic curves by the height of f.

21.2 Local solubility

Suppose we have some set A which we want to prove is empty. One way to do this is to construct a map $A \to B$ and show that $B = \varnothing$. If, for example $A = C(\mathbf{Q})$, then for each place v of $\mathbf{Q}$ we have a natural injection $C(\mathbf{Q}) \hookrightarrow C(\mathbf{Q}_v)$. If $C(\mathbf{Q}_v) = \varnothing$ for some v, then $C(\mathbf{Q}) = \varnothing$.

Definition 21.2.1. A curve C is said to be *everywhere locally soluble* if $C(\mathbf{R}) \neq \varnothing$ and $C(\mathbf{Q}_p) \neq \varnothing$ for all primes p.

Equivalently, C is everywhere locally soluble if $C(\mathbf{A}) \neq \varnothing$.

Theorem 21.2.2. *If C is not everywhere locally soluble, then $C(\mathbf{Q}) = \varnothing$.*

Example 21.2.3. Consider $C : y^2 - x^6 - x^2 - 17$. Then $C(\mathbf{R}) = \varnothing$.

Example 21.2.4. The curve $C : y^2 = -x^6 - 3x^5 + 4x^4 + 2x^3 + 4x^2 - 3x - 1$ has $C(\mathbf{Q}_{11}) = \varnothing$, because $C(\mathbf{F}_{11}) = \varnothing$.

As a general principle, local questions are computable, whereas global questions are very hard. In our case, we have the following result.

Proposition 21.2.5. *There is an algorithm that decides if a hyperelliptic curve over* $\mathbf{Q}$ *is everywhere locally soluble or not.*

Sketch of proof. We need to take care of two problems:

- there are infinitely many places of $\mathbf{Q}$
- for each place v, the field $\mathbf{Q}_v$ is uncountable

It is easy to check whether $C(\mathbf{R}) = \varnothing$. This is the case if and only if f has no real roots, which happens exactly when f is strictly positive or strictly negative. This is easy to decide. For fixed p, "$C(\mathbf{Q}_p) = \varnothing$" reduces to a question modulo p^n via Hensel's lemma. So for any completion v of $\mathbf{Q}$, it is a finite problem to check whether $C(\mathbf{Q}_v) = \varnothing$. For p sufficiently large, $p \nmid \mathrm{Disc}(f)$, so we have $C(\mathbf{Q}_p) \neq \varnothing$ via a theorem of Weil, namely $\#C(\mathbf{F}_p) \geqslant p + 1 - 2g\sqrt{p}$. For $p \nmid \mathrm{Disc}(f)$, the curve C is smooth over $\mathbf{F}_p$, so points in $C(\mathbf{F}_p)$ lift to points in $C(\mathbf{Q}_p)$. □

A curve which is everywhere locally soluble does not necessarily have a solution in $\mathbf{Q}$. For fixed $g \geqslant 2$, the everywhere locally soluble curves of genus g have a density δ_g. For example $\delta_2 \approx 0.84$ and $\delta_g \to 1$ as $g \to \infty$. So we expect 100% of hyperelliptic curves to have no rational points, but also for 100% of hyperelliptic curves to be everywhere locally soluble.

21.3 Descent

Consider $C : y^2 = f(x)$ with $f(x) = f_1(x)f_2(x)$, with $\deg f_1, \deg f_2$ even and $f_1, f_2 \in \mathbf{Z}[x]$. Let $P = (\xi, \eta) \in C(\mathbf{Q})$. Then $f_1(\xi) \neq 0$, $f_2(\xi) \neq 0$, or both. There exists a unique squarefree $d \in \mathbf{Z}$, together with $\eta_1, \eta_2 \in \mathbf{Q}$, such that $f_1(\xi) = d\eta_1^2$, $f_2(\xi) = d\eta_2^2$. More geometrically, P lifts to a rational point on $D_d : dy_1^2 = f_1(x), dy2^2 = f_2(x)$ with $\pi_d : D_d \to C$ defined by $(x, y_1, y_2) \mapsto (x, dy_1y_2)$. So we have reduced the problem of finding rational points on C to that of finding rational points on the family $\{D_d : d \in \mathbf{Z}\}$ of curves.

If $p \mid d$, then $\overline{\xi}$ a common root of $\overline{f_1}, \overline{f_2} \in \mathbf{F}_p[x]$ implies $p \mid \mathrm{Res}(f_1, f_2) \neq 0$. This is possible for only finitely many values of d. Let T be the set of possible d. For each $d \in T$, we can use an algorithm to check if D_d is everywhere locally soluble. If none of them are, then $C(\mathbf{Q}) = \varnothing$.

Example 21.3.1. Let $C : y^2 = f_1f_2$, where

$$f_1 = -x^2 - x - 1$$
$$f_2 = x^4 + x^3 + x^2 + x + 2.$$

As an exercise, check that C is everywhere locally soluble. We have $\mathrm{Res}(f_1, f_2) = \pm 19$, so we can set $T = \{\pm 1, \pm 19\}$. If $d < 0$, then $D_d(\mathbf{R}) = \varnothing$. If $d \equiv 1 \pmod 3$, then $D_d(\mathbf{F}_3) = \varnothing$. It follows that $C(\mathbf{Q}) = \varnothing$.

This approach can be generalized to unramified coverings $\pi : D \to C$ that are Galois over $\overline{\mathbf{Q}}$.

21.4 The (fake) 2-Selmer set

Recall our strategy for showing that a set is empty. There is a more sophisticated version. Instead of looking at maps $A \to B$, look at commutative diagrams

$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow{\scriptstyle \beta} \\ C & \xrightarrow{\gamma} & D. \end{array}$$

If $\mathrm{im}(\beta) \cap \mathrm{im}(\gamma) = \varnothing$, then $A = \varnothing$. We will construct such a diagram with $A = C(\mathbf{Q})$.

Let $L = \mathbf{Q}[x]/f$ and $L_v = L \otimes \mathbf{Q}_v$. Write T for the image of x in L and each L_v. Define

$$H = \left\{ \alpha \in L^\times/(L^\times)^2\mathbf{Q}^\times : N_{L/\mathbf{Q}}(\alpha) = \mathrm{lcf}(f) \cdot (\text{square}) \text{ in } \mathbf{Q}^\times \right\}$$

where $\mathrm{lcf}(f)$ is the leading coefficient of f. Similarly define H_v for each place v. There are natural maps $\rho_v : H \to H_v$. Define $\delta : C(\mathbf{Q}) \to H$ by

$$\begin{aligned} &(\xi, \eta) \mapsto (\xi - T)(L^\times)^2\mathbf{Q}^\times && \text{if } \eta \neq 0 \\ &(\xi, 0) \mapsto (\xi - T - f_1(T))(L^\times)^2\mathbf{Q}^\times && \text{if } f(x) = (x - \xi)f_1(x) \\ &\infty_{\pm s} \mapsto (L^\times)^2\mathbf{Q}^\times. \end{aligned}$$

Similarly define $\delta_v : C(\mathbf{Q}_v) \to H_v$. We have a commutative diagram

$$\begin{array}{ccc} C(\mathbf{Q}) & \xrightarrow{\delta} & H \\ \downarrow & & \downarrow{\scriptstyle (\rho_v)_v} \\ C(\mathbf{A}) & \xrightarrow{\prod \delta_v} & \prod H_v. \end{array}$$

Definition 21.4.1. The *(fake) 2-Selmer set* of C is

$$\mathrm{Sel}_2^{\mathrm{fake}}(C) = \{\alpha \in H : \rho_v(\alpha) \in \mathrm{im}(\delta_v) \text{ for all } v\}.$$

If $\mathrm{Sel}_2^{\mathrm{fake}}(C) = \varnothing$, then $C(\mathbf{Q}) = \varnothing$. There is a "2-Selmer set" $\mathrm{Sel}_2(C)$ with a map $\mathrm{Sel}_2(C) \to \mathrm{Sel}_2^{\mathrm{fake}}(C)$ that is surjective. It is bijective for some curves, but is usually 2-to-1.

The set $\mathrm{Sel}_2^{\mathrm{fake}}(C)$ can be computed. Let S be the set of places of L dividing one of $\{2, \infty, \mathrm{Disc}(f), \mathrm{lcf}(f)\}$. Then

$$\mathrm{Sel}_2^{\mathrm{fake}}(C) \subset H_S = \{\alpha(L^\times)^2\mathbf{Q}^\times : v_{\mathfrak{p}}(\alpha) \text{ even for all } \mathfrak{p} \notin S\}.$$

The set H_S is finite by standard arguments.

In [BG13], it is shown that the (upper) density of genus g with $\mathrm{Sel}_2^{\mathrm{fake}}(C) \neq \varnothing$ is $o(2^{-g})$.

Example 21.4.2 (Bruin, Stoll)**.** Of the $\sim$ 200000 isomorphism classes of genus 2 curves of height $\leqslant 3$, all but $\sim$ 1500 either

- have a rational point,
- fail to be everywhere locally soluble, or
- have empty 2-Selmer set .

22 Binary quartic forms: bounded average rank of elliptic curves II

22.1 Review

Recall that if $E_{A,B}$ is the elliptic curve $y^2 = x^3 + Ax + B$, then $\mathrm{Sel}_2(E_{A,B})$ is in bijection with the quotient $\mathrm{PGL}_2(\mathbf{Q})\backslash V(\mathbf{Z})^{\mathrm{ls}}_{(A,B)}$, where $V(\mathbf{Z})^{\mathrm{ls}}_{(A,B)}$ is the set of locally soluble integral binary quartic forms with invariants A, B. Even though the action of $\mathrm{PGL}_2(\mathbf{Q})$ does not preserve $V(\mathbf{Z})$, it still induces an equivalence relation, where $f \sim g$ whenever there is $\gamma \in \mathrm{GL}_2(\mathbf{Z})$ such that $\gamma \cdot f = g$. For example, the forms $p^4x^4 + p^2xy^3 + y^4$ and $x^4 + p^4xy^3p^4y^4$ are $\mathrm{PGL}_2(\mathbf{Q})$-equivalent via the matrix $\begin{pmatrix} p^{-1} & \\ & p \end{pmatrix}$. For asymptotics, we could restrict to irreducible quartic forms. We defined subsets $V(\mathbf{Z})^{(i)}_{(A,B)}$ of $V(\mathbf{Z})$; for definitions, see section 19. We ended up with an estimate

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{\mathrm{irr},(i)}_{H<X}\right) = \frac{1}{n_i}|J|\,\mathrm{Vol}(\mathcal{F})\,\mathrm{Vol}\left(R^{(i)}_X\right) + O(X^{3/4}).$$

In this lecture, we will prove the following theorem.

Theorem 22.1.1 (Bhargava, Shankar)**.** $\mathrm{avg}(\#\,\mathrm{Sel}_2) = 3$.

To do this, we will need to replace $V(\mathbf{Z})^{\mathrm{irr}}$ by $V(\mathbf{Z})^{\mathrm{irr,ls}}$, and replace $\mathrm{PGL}_2(\mathbf{Z})$-orbits by $\mathrm{PGL}_2(\mathbf{Q})$-equivalence classes.

22.2 Local solubility

For each prime p, let $V(\mathbf{Z}_p)^{\mathrm{s}}$ be the subset of $V(\mathbf{Z}_p)$ consisting of soluble binary quartic forms. Let $V(\mathbf{Z})^{\mathrm{s}(p)}$ be the set of forms in $V(\mathbf{Z})$ whose image in $V(\mathbf{Z}_p)$ is soluble. We start by computing

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{\mathrm{irr,s}(p)}_{H<X}\right) = \frac{|J|}{n_i}\,\mathrm{Vol}(\mathcal{F})\,\mathrm{Vol}\left(R^{(i)}_X\right)\,\mathrm{Vol}\left(V(\mathbf{Z}_p)^{\mathrm{s}}\right) + O(X^{3/4}).$$

To do this for locally soluble forms, we will need to look at "soluble at p forms" for all p. For that, we need a sieve.

We want

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{\mathrm{irr},p^2|\Delta}_{H<X}\right) = O(X^{5/6}/p^{1+\delta}),$$

for any $\delta > 0$. In fact, this is stronger than we need. All the proof requires is

$$\sum_{p>M} \#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})^{\mathrm{irr},p^2|\Delta}_{H<X}\right) = O(X^{5/6}/f(M)),$$

where $f(M) \to \infty$ as $M \to \infty$.

Recall the (naive) height is $H(E_{A,B}) = \max\{4|A|^3, 27B^2\}$. We want

$$\#\{E_{A,B} : H(E_{A,B}) < X \text{ and } p^2 \mid \Delta(E_{A,B})\} = O\left(\frac{X^{5/6}}{p^{1+\delta}}\right),$$

where $\delta > 0$. When p is large, map $E_{A,B}$ to the binary cubic form $x^3 + Axy^2 + By^3$, which goes to $\mathrm{GL}_2(\mathbf{Z}) \cdot (x^3 + Axy^2 + By^3)$. We have defined a map

$$\varphi : U_1(\mathbf{Z}) \to U(\mathbf{Z}) \to \mathrm{GL}_2(\mathbf{Z}) \backslash U(\mathbf{Z}),$$

where $U_1(\mathbf{Z})$ is the space of elliptic curves and U is the space of all binary cubic forms. This map is discriminant-preserving.

Theorem 22.2.1 (Delone, Nagell, Siegel, Evertse, Akhtari)**.** *The map φ is at most 7-to-1 for elements with large enough discriminant.*

This is very deep. As one application, a binary cubic form represents one at most seven times. It follows from the theorem that

$$\#\left\{E_{A,B} : H(E_{A,B}) < X \text{ and } p^2 \mid \Delta(E_{A,B})\right\} = O\left(\frac{X}{p^2}\right).$$

This isn't quite what we wanted because the estimate has X instead of $X^{5/6}$. But for p sufficiently large, it works.

Suppose $p^2 \mid \Delta(E_{A,B})$ for "modulo p reasons," i.e. if $E_{A,B}$ has additive reduction. If $p > 3$, then $A \equiv B \equiv 0 \pmod{p}$. The bound in this case is

$$O\left(\left(\frac{X^{1/3}}{p^2} + 1\right)\left(\frac{X^{1/2}}{p} + 1\right)\right) = O\left(\frac{X^{5/6}}{p^2} + \frac{X^{1/2}}{p} + 1\right).$$

If $p \mid \Delta(E_{A,B})$ for "modulo p^2 reasons," then fixing A determines B modulo p^2. In this case, the bound is

$$O\left(X^{1/3} \cdot \left(\frac{X^{1/2}}{p} + 1\right)\right) = O\left(X^{5/6}/p^2 + X^{1/3}\right).$$

Combining these estimates yields the uniform bound

$$\#\{E : H(E) < X \text{ and } p^2 \mid \Delta(E)\} = O\left(\frac{X^{5/6}}{p^{3/2}}\right).$$

Recall that there is a bijection between $\mathrm{PGL}_2(\mathbf{Z}) \backslash V(\mathbf{Z})$ and the set of (Q, C, x), where Q is a quartic ring, C is a cubic resolvent ring, and x generates C (?) The map $V(\mathbf{Z}) \to \mathbf{Z}^2 \otimes \mathrm{Sym}^2(\mathbf{Z}^3)$ induces the map sending (Q, C, x) to (Q, C). On the side of forms, the map sends $ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$ to

$$\begin{pmatrix} & & 1/2 \\ & -1 & \\ 1/2 & & \end{pmatrix}, \begin{pmatrix} a & b/2 & \\ b/2 & c & d/2 \\ & d/2 & e \end{pmatrix}.$$

Our map yields the bound

$$\#\left(\mathrm{PGL}_2(\mathbf{Z}) \backslash V(\mathbf{Z})_{H<X}^{p^2 \mid \Delta}\right) = O\left(\frac{X}{p^2}\right).$$

Combining everything, we get

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})_{H<X}^{p^2|\Delta,\ \bmod p^2}\right) = O\left(\frac{X^{5/6}}{p^2} + X^{2/3}\right)$$

$$\sum_{p<M} \#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})_{H<X}^{p^2|\Delta}\right) = O\left(\frac{X^{5/6}}{\log M}\right).$$

22.3 Weights

The problem is that a single $\mathrm{PGL}_2(\mathbf{Q})$-class in $V(\mathbf{Z})$ could break up into seven different $\mathrm{PGL}_2(\mathbf{Z})$-orbits. Given a form f, let

$$B_f = \mathrm{PGL}_2(\mathbf{Z})\backslash\left(\mathrm{PGL}_2(\mathbf{Q})\cdot f \cap V(\mathbf{Z})\right).$$

For $f \in V(\mathbf{Z})$, we define

$$W(f) = \begin{cases} 0 & \text{if } f \text{ is not locally soluble} \\ \left(\displaystyle\sum_{g\in B_f} \frac{\#\operatorname{Aut}_{\mathbf{Q}}(g)}{\operatorname{Aut}_{\mathbf{Z}}(g)}\right)^{-1} & \text{otherwise} \end{cases}$$

The weight of f is a product of local weights. That is, define for each p

$$W_p(f) = \begin{cases} 0 & \text{if } f \text{ is not locally soluble} \\ \left(\displaystyle\sum_{g\in B_p(f)} \frac{\#\operatorname{Aut}_{\mathbf{Q}_p}(g)}{\operatorname{Aut}_{\mathbf{Z}_p}(g)}\right)^{-1} & \text{otherwise} \end{cases}$$

Then we have the following proposition (3.3 in my paper).

Proposition 22.3.1. *For $f \in V(\mathbf{Z})$, we have $W(f) = \prod_p W_p(f)$.*

We get the following formula:

$$\#\left(\mathrm{PGL}_2(\mathbf{Z})\backslash V(\mathbf{Z})_{H<X}^{\mathrm{irr},W,(i)}\right) = \frac{|J|}{n_i}\operatorname{Vol}(\mathcal{F})\operatorname{Vol}\left(R_X^{(i)}\right)\prod_p \int_{V(\mathbf{Z}_p)} W_p(f)\,\mathrm{d}f.$$

Proposition 22.3.2.

$$\int_{V(\mathbf{Z}_p)} W_p(f)\,\mathrm{d}f = |J|_p \operatorname{Vol}(\mathrm{PGL}_2(\mathbf{Z}_p)) \int_{\mathbf{Z}_p^2} \frac{\#\left(E_{A,B}(\mathbf{Q}_p)/2\right)}{\#E_{A,B}[2](\mathbf{Q}_p)}\,\mathrm{d}(A,B).$$

We also have

$$|J|\operatorname{Vol}\left(\mathrm{PGL}_2(\mathbf{Z})\backslash \mathrm{PGL}_2(\mathbf{R})\right) \int_{\{(A,B)\in\mathbf{R}^2 : H(E_{A,B})<X\}} \frac{\#(E_{A,B}(\mathbf{R})/2)}{\#E_{A,B}[2](\mathbf{R})}\,\mathrm{d}(A,B).$$

Proposition 22.3.3 (Brummer and Kramer)**.**

$$\frac{\#(E(\mathbf{Q}_v)/2)}{\#E[2](\mathbf{Q}_v)} = \begin{cases} 1/2 & v = \infty \\ 2 & v = 2 \\ 1 & \textit{otherwise} \end{cases}$$

Thus the average of $\#\operatorname{Sel}_2) - 1$ is the following limit:

$$\lim_{X\to\infty} \frac{\operatorname{Vol}(\operatorname{PGL}_2(\mathbf{Z})\backslash \operatorname{PGL}_2(\mathbf{R})) \prod_p \operatorname{Vol}(\operatorname{PGL}_2(\mathbf{Z}_p) + O(X^{3/2})) \int_{H(A,B)<X} \mathrm{d}(A,B)}{\int_{H<X} \mathrm{d}(A,B) + O(X^{1/2})}.$$

The integrals cancel, and the product over all places in the numerator is the Tamagawa number $\tau(\operatorname{PGL}_2) = 2$. It follows that

$$\operatorname{avg}(\#\operatorname{Sel}_2) = 2 + 1 = 3.$$

When we generalize to Sel_n for $n \geqslant 3$, things get a bit more complicated, as in the following table:

n	group	space	$\tau(G)$	$\operatorname{avg}(\#\operatorname{Sel}_n)$
2	$\operatorname{PGL}_2(\mathbf{Z})$	$\operatorname{Sym}^4(\mathbf{Z}^2)$	2	3
3	$\operatorname{PGL}_3(\mathbf{Z})$	$\operatorname{Sym}^3(\mathbf{Z}^3)$	3	4
4	qt. of $\operatorname{GL}_2 \times \operatorname{GL}_4$	$\mathbf{Z}^2 \otimes \operatorname{Sym}^2(\mathbf{Z}^4)$	4	7
5	qt. of $\operatorname{GL}_5 \times \operatorname{GL}_5$	$\mathbf{Z}^5 \otimes \bigwedge^2 \mathbf{Z}^5$	5	6

23 Coregular spaces and genus one curves

23.1 Introduction and motivation

Something we have done many times is take a representation V of a group G and study the orbits V/G. The stabilizers of points in V are also important. We have tried to describe the orbits in terms of "arithmetically interesting" objects, e.g. elliptic curves with extra data. If we impose local conditions on V/G, we get elliptic curves with Selmer elements. One final note: we want stabilizers in G to match up with automorphism groups of the "arithmetically interesting objects." This is a subtle but important point.

In section 19 and section 22, the "extra data" attached to an elliptic curve E was an n-covering of E. This consists of a E-torsor C with a degree n line bundle on C.

Example 23.1.1. In the binary quartic case, the form $f(x,y) = ax^4 + \cdots$ corresponds to the curve $C : z^2 = ax^4 + bx^3y + cx^2y^2 + dxy^3 + ey^4$; this is a double cover of $\mathbf{P}^1$ ramified at four points. The invariants of f determine an elliptic curve. The map $C \to \mathbf{P}^1$ determines the line bundle on C.

Example 23.1.2. Recall that we can describe 3-Selmer elements with ternary cubics (homogeneous degree three polynomials in three variables). The representation

involved is $\mathrm{Sym}^3(3) = \mathrm{Sym}^3 \mathbf{A}^3$. Any ternary cubic f gives a genus one curve in $\mathbf{P}^2$. It's Jacobian is canonically the elliptic curve with invariants those of f. The degree three line bundle on C comes from the embeddigng $C \hookrightarrow \mathbf{P}^2$. The rings of invariants of the action of $\mathrm{SL}(3)$ on $\mathrm{Sym}^3(3)$ is polynomial in two generators S, T, of degree 4 and 6 respectively. The Jacobian of C_f is $E_{S(f),T(f)}$.

To generalize these ideas, we need to find other representations that parameterize interesting data. Let k be an algebraically closed field. If (G, V) is a prehomogeneous vector space over k and $U \subset V$ is open and G-stable, then $U(k)/G(k)$ might be "zero-dimensional," e.g. a single Zariski-open orbits. But there are lots of non-isomorphic elliptic curves, even over an algebraically closed field. so we would want $U(k)/G(k)$ to be "bigger," e.g. the affine line.

More geometrically, we are trying to find prehomogeneous vector spaces (G, V) such that the coarse moduli space V/G is a moduli space we already are familiar with. In all our examples, the invariant rings are polynomial rings. We call such representations *coregular*.

The moduli space of elliptic curves is birational to $\mathbf{P}(2, 3)$. So we should look for coregular representations with ring of invariants free in two variables. Elliptic curves with one marked point are of the form $y^2 + d_3 y = x^3 + d_2 x^2 + d_4 x^2$. If we take our scaling, we get $\mathbf{P}(2, 3, 4)$, or $\mathbf{A}^3$ if we also keep track of the differential.

To summarize: many (but not all!) families of elliptic curves with extra data have coarse moduli space (look over an algebraically closed field) birational to a weighted projective space. This means the invariant ring of any possible (G, V) is a polynomial ring.

[. . . couldn't follow. . .]

24 Arithmetic invariant theory and hyperelliptic curv I

Let k be a field, G a reductive group over k, and $G \to \mathrm{GL}(V)$ a representation of G. The ring $\mathrm{Sym}^\bullet(V^\vee)$ contains a subring $\mathrm{Sym}^\bullet(V^\vee)^G$ of invariant polynomials. An important theorem is that $\mathrm{Sym}^\bullet(V^\vee)^G$ is a finitely generated k-algebra. Write $V/\!\!/G$ for the variety $\mathrm{Spec}\left(\mathrm{Sym}^\bullet(V^\vee)^G\right)$; this comes with a canonical "projection" $\pi : V \to V/\!\!/G$.

24.1 First examples and results

Example 24.1.1. Consider $G = \mathrm{GL}(W)$ and $V = \mathfrak{gl}(W) = \mathrm{End}\, W$, with the adjoint action of G on V. It is known that $\mathfrak{gl}(W)/\!\!/\mathrm{GL}(W)$ is an affine space with coordinates the "coefficients of the symmetric polynomial."

More generally, if G is a reductive group and $\mathfrak{g} = \mathrm{Lie}\, G$ under the adjoint representation, then $\mathfrak{g}/\!\!/G$ is affine. That is, Chevalley proved that the adjoint representation of a reductive group is coregular. Winberg generalized this even further. If $\theta : G \to G$ is an automorphism of order m and $\mathfrak{g} = \bigoplus_a \mathfrak{g}(a)$, then the action of G^θ on each $\mathfrak{g}(a)$ is coregular.

Suppose $f \in (V/\!\!/G)(k)$. Let V_f be the fiber in V of π over f. Then $V_f(k)$ is a (possibly empty) union of $G(k)$-orbits.

Example 24.1.2. When $G = \mathrm{GL}(n)$ and $V = \mathfrak{gl}(n)$, then V_f is all linear T with fixed characteristic polynomial f. The set $V_f(k)$ is always nonempty. Indeed, let $L = k[x]/f$ and $\theta : L \to L$ be "multiplication by x." Then θ is a k-linear transformation with characteristic polynomial f. Choosing an isomorphism $L \simeq k^n$ gives an element in $V_f(k)$. Roughly, "every polynomial is the characteristic polynomial of a map defined over the base field."

If the discriminant $\Delta(f) \neq 0$, there is a single orbit of $G(k)$ on $V_f(k)$. For any $T \in V_f(k)$, the stabilizer G_T is isomorphic to the Weil restriction $\Pi_{L/k}\mathbf{G}_{\mathrm{m}}$; a maximal torus in $\mathrm{GL}(V)$ if L is étale. If $f(x) = x^k$, then V_f is known as the *nilpotent cone*; the orbit we constructed is the *regular nilpotent*.

Example 24.1.3. Consider the action of $\mathrm{SL}(W)$ on $\mathfrak{sl}(W) = \mathrm{End}(W)^{\mathrm{tr}=0}$. The ring of invariants is freely generated by all but the constant term of the characteristic polynomial, so $\mathfrak{sl}(n)/\!\!/\mathrm{SL}(n) \simeq \mathbf{A}^{n-1}$. If $\Delta(f) \neq 0$, then orits in $V_f(k)$ are in bijection with $k^\times / \mathrm{N}(L^\times)$.

If for example $\dim W = 2$ and $f(x) = x^2 + 1$, then the orbit space $V_f(k)/G_f(k)$ is in bijection with $\mathbf{Q}^\times / \mathrm{N}(\mathbf{Q}(i)^\times)$; a (huge) abelian 2-group.

For $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$, we obtained that $V_f(k)$ was a torsor over $\mathrm{H}^1(k, G_f)$. But this used $\mathrm{H}^1(k, \mathrm{SL}_n) = \mathrm{H}^1(k, \mathrm{GL}_n) = 0$.

24.2 Principles of arithmetic invariant theory

Principle 24.2.1. *Assume $V_f(k) \ni v$, and that $G(k^s)$ acts transitively on $V_f(k^s)$. Then the set of $G(k)$-orbits on $V_f(k)$ is in bijection with the kernel of the map of pointed sets $\mathrm{H}^1(k, G_v) \to \mathrm{H}^1(k, G)$.*

Proof. Say $v' \in V_f(k) \subset V_f(k^s)$, write $v' = g(v)$ for some $g \in G(k^s)$. Send the orbit of v' to the class in $\mathrm{H}^1(k, G_v)$ of the cocycle $\sigma \mapsto c_\sigma = g^{-1} \circ g^\sigma$. Checking that this is a bijection is easy. □

If G is one of $\mathrm{GL}(n)$, $\mathrm{SL}(n)$, $\mathrm{Sp}(n)$, then $\mathrm{H}^1(k, G) = 0$, so $V_f(k)/G(k) = \mathrm{H}^1(k, G_v)$.

Example 24.2.2. Let W be a split orthogonal space over k of dimension $n = 2g+1$. So W is a direct sum of g hyperbolic planes and a single copy of k. Let $G = \mathrm{SO}(W)$. For example, if $g = 1$, then $G \xrightarrow{\sim} \mathrm{PGL}(2) \xrightarrow{\sim} \mathrm{SO}_3$. Let $V = \mathfrak{so}(W)$; the space of trace-zero self-adjoint operators on W. Then $\mathrm{Sym}^\bullet(W^\vee)^G = k[c_2, \ldots, c_{2g+1}]$, freely generated on the coefficients of the characteristic polynomial. It's a bit more difficult to show that the fibers of $\mathfrak{so}(W) \to \mathfrak{so}(W)/\!\!/\mathrm{SO}(W)$ are all nonempty. Given f in the quotient, define as before $L = k[x]/f$. This is a k-algebra of rank $2g+1$. Let $\langle \lambda, \mu \rangle$ be the coefficient x^{2g} in $\lambda\mu$; also $\mathrm{tr}_{L/k}(\lambda\mu/f'(x))$. The operator "multiplication by x" on L is self-adjoint with characteristic polynomial f. A bit of work shows that this gives an element of $V_f(k)$.

In fact, for $V = \mathfrak{so}(n)$, $G = \mathrm{SO}(n)$, the map $V(k) \to V/\!\!/G$ has a standard section known as the Kostant section.

Let's compute stabilizers. For f with $\Delta(f) \neq 0$, it is easy to see that $G_v = \mathrm{SO}(W) \cap L^\times$. This is $L^\times[2]^{\mathrm{N}=1}$. As a group scheme, this is $\ker(\Pi_{L/k}\boldsymbol{\mu}_2 \xrightarrow{\mathrm{N}} \boldsymbol{\mu}_2)$. An easy computation of Galois cohomology shows that $\mathrm{H}^1(k, G_v) = (L^\times/2)^{\mathrm{N}=0}$. In this case, $\mathrm{H}^1(k, G_v)$ is also $\mathrm{H}^1(k, J[2])$, where J is the Jacobian of $y^2 = f(x)$.

In general, our first principle is not very useful, because the map $\gamma : \mathrm{H}^1(k, G_v) \to \mathrm{H}^1(k, G)$ can be pretty complicated, and is not easy to pin down explicitly.

Principle 24.2.3. *For any $c \in \mathrm{H}^1(k, G)$, there is a twisted group G^c over k and twisted representation V^c over k. The fiber of γ over c is the set of orbits of $G^c(k)$ in $V_f^c(k)$.*

Recall our map $\mathrm{H}^1(k, G_v) = \mathrm{H}^1(k, J[2]) \xrightarrow{\gamma} \mathrm{H}^1(k, \mathrm{SO}(W))$, where J is the Jacobian of $y^2 = f$. The 2-Selmer group $\mathrm{Sel}_2(J)$ is a subset of $\mathrm{H}^1(k, J[2])$, and in [BG13], Bhargava and I showed that $\mathrm{Sel}\,2(J) = \ker(\gamma)$. In general, when is V_f empty for all G^c?

Principle 24.2.4. *Assume $G(k^s)$ acts transitively on $V_f(k^s)$ and $G_v(k^s)$ is abelian.*

1. *If the class of d_f is non-trivial in $\mathrm{H}^2(k, G_f)$, there is no k-rational point in any fiber.*
2. *If $d_f = 0$, then the fiber is nontrivial for some pure inner form G^c.*

Proof. Take $v \in V_f(k^s)$. Then $\sigma_v = {}^\sigma f \circ f$ is also in $V_f(k^s)$. Define $\theta_\sigma : G_{c_v} \to G_v$ by $\alpha \mapsto g_\sigma \alpha g_\sigma^{-1}$. This map is an isomorphism that does not depend on v. Since $\theta_{\sigma\tau} = \theta_\sigma \circ {}^\sigma\theta_\tau$, this descends G_v to a group G_f defined over k. □

So if $G_v(k^s)$ is abelian and $f \in (V/\!\!/G)(k)$, there is a "stabilizer" G_f of f even if $V_f(k) = \varnothing$.

Example 24.2.5. Let $G = \mathrm{SL}(W)$, where $\dim W = 2g + 2$. Let $V = \mathrm{Sym}^2(W^\vee) \oplus \mathrm{Sym}^2(W^\vee)$. We can think of V as the space of pairs $v = (A, B)$ of symmetric matrices. The ring of invariants is freely generated by the coefficients of the bilinear form $f(x, y) = (-1)^{g+1} \det(xA - yB)$. Assume $\Delta(f) \neq 0$. Then put $G_f = (\Pi_{L/k}\boldsymbol{\mu}_2)^{\mathrm{N}=1}$, where L/k is the extension constructed earlier. If we put $f(x, 1) = f_0 \cdot g(x)$, then $L = k[x]/g$. What is the class $d_f \in \mathrm{H}^2(k, G_f)$? This cohomology group has a subgroup $k^\times/k^\times\,\mathrm{N}(L^\times)$. The class d_f is just the class of f_0 in $k^\times/k^\times\,\mathrm{N}(L^\times) \subset \mathrm{H}^2(k, G_f)$.

For example, if $k = \mathbf{R}$, $g = 0$, $f_0 = -1$, $f = -x^2 - y^2$, and $g = x^2 + 1$, then there are no orbits.

25 Most hyperelliptic curves have no rational points

25.1 Summary of results

According to Don Zagier, the title should be "most hyperelliptic curves are pointless." Recall that a *hyperelliptic curve* is a smooth projective geometrically irreducible curve

with a degree-two map to $\mathbf{P}^1$. More concretely, any hyperelliptic curve over $\mathbf{Q}$ can be expressed in the form

$$C : z^2 = f_0 x^n + f_1 x^{n-1} y + \cdots + f_{n-1} x y^{n-1} + f_n y^n, \tag{$*$}$$

where $n = 2g + 2$ and g is the genus of C, at least if $\mathrm{Disc}(f) \neq 0$. By scaling, we may assume the $f_i \in \mathbf{Z}$. Define a height on hyperelliptic curves by $H(C) = \max\{|f_i|\}$. We will order hyperelliptic curves by height. The results of this section are mostly from [BG13].

Theorem 25.1.1. *Order all hyperelliptic curves ($*$) over $\mathbf{Q}$ of genus g by height. Then as $g \to \infty$, a density approaching 100% of hyperelliptic curves of genus g have no rational points.*

More precisely, the upper density of hyperelliptic curves of genus g having a rational point is $o(2^{-g})$. Since most ($> 75\%$) hyperelliptic curves of genus $g \geqslant 1$ have points over $\mathbf{Q}_v$ for all places v (everywhere locally soluble), we obtain the following.

Corollary 25.1.2. *As $g \to \infty$, a density approaching 100% of everywhere locally soluble hyperelliptic curves ($*$) of genus g fail the Hasse principle.*

For $g = 1$, the density is $> 20\%$, and for $g = 2$, the proportion of $> 50\%$. For $g = 10$, the density is $> 99\%$. In the 1940's, Lind and Reichardt independently gave examples of equations of the form $z^2 = f(x, y)$, where f is a quartic, failing the Hasse principle. Later on, Selmer gave an example of an elliptic curve (minus the origin) failing the Hasse principle. A more elementary reformulation of these results is that binary forms rarely take square values.

25.2 Key construction

The main idea is: use the representation $V(\mathbf{Z}) = \mathbf{Z}^2 \otimes \mathrm{Sym}^2(\mathbf{Z}^n)$ of $G(\mathbf{Z}) = \mathrm{GL}_n(\mathbf{Z})$. We can view elements of $V(\mathbf{Z})$ as pairs (A, B) of $n \times n$ symmetric matrices with integer entries. The group G acts by $\gamma \cdot (A, B) = (\gamma A\, {}^{\mathrm{t}}\gamma, \gamma B\, {}^{\mathrm{t}}\gamma)$. Given such a $v = (A, B) \in V(\mathbf{Z})$, define the *invariant binary n-form* $f_v(x, y) = -1^{n/2} \det(Ax - By)$. The coefficients of $f_v(x, y)$ give invariants for the action of $G(\mathbf{Z})$ on $V(\mathbf{Z})$. In fact, these freely generate the ring of invariants over $\mathbf{C}$.

Given a binary n-ic form f over $\mathbf{Z}$, when does it arise as f_v for some $v = (A, B) \in V(\mathbf{Z})$? Unfortunately not always.

Proposition 25.2.1. *Let f be a binary n-ic form over $\mathbf{Z}$. If $z^2 = f(x, y)$ has a rational point, then $f = f_v$ for some $v \in V(\mathbf{Z})$.*

Proof. We use a classification of the orbit space $G(\mathbf{Z})\backslash V(\mathbf{Z})$. Given a rational point, we'll produce an "algebraic object," which will give our pair $v = (A, B)$ of $n \times n$ symmetric matrices. Given a binary n-ic form f over $\mathbf{Z}$, assume $\mathrm{Disc}(f) \neq 0$ and $f_0 \neq 0$, where $f = f_0 x^n + \cdots + f_n y^n$. Let $K_f = \mathbf{Q}[x]/f(x, 1) = \mathbf{Q}[\theta]$; this is an n-dimensional $\mathbf{Q}$-algebra. Inside K_f, there is a lattice R_f with basis $\{1, \zeta_1, \zeta_2, \ldots, \zeta_{n-1}\}$, where

$$\zeta_k = f_0 \theta^k + f_1 \theta^{k-1} + \cdots + f_{k-1} \theta.$$

See section 12 for details of this construction. The ζ_k are integral over $\mathbf{Z}$. In [BM72], Birch and Merriman proved that $\mathrm{Disc}(R_f) = \mathrm{Disc}(f)$. Much more recently, Nakagawa proved that R_f is a ring. Define further lattices in K_f:

$$I_f(k) = \langle 1, \theta, \theta^2, \ldots, \theta^k, \zeta_{k+1}, \ldots, \zeta_{n-1} \rangle,$$

for any $0 \leqslant k < n$. Then $I_f(k)$ is an R_f-module and $I_f(k) = I_f^k$. Note that $I_f(0) = R_f$; we define $I_f = I_f(1)$. Checking this is a good exercise. The $I_f(k)$ come equipped with bases. Given (I, α) as in the following theorem, take coefficients of ζ_{n-1} and ζ_{n-2} in $\frac{1}{\alpha} : I \times I \to I_f(n-3)$. These coefficients are the (A, B) we wanted to construct. □

Warning: the converse to the Proposition is false. Recall that the ring R_f is constructed in section 12 as the ring of global functions on the subscheme of $\mathbf{P}^1_{\mathbf{Z}}$ cut out by f. The module $I_f(k)$ is global sections of the pullback of $\mathscr{O}(k)$.

Theorem 25.2.2 (Wood)**.** *The set $(G(\mathbf{Z})\backslash V(\mathbf{Z}))_f$ is naturally in bijection with the set of equivalences classes of pairs (I, f), where I is a fractional ideal of R_f and $\alpha \in K_f^\times$ such that $I^2 \subset \alpha I_f(n-3)$ and $\mathrm{N}(I)^2 = \mathrm{N}(\alpha)\,\mathrm{N}(I_f^{n-3})$. Here the equivalence relation is $(I, \alpha) \sim (\kappa I, \kappa^2 \alpha)$ for $\kappa \in K_f^\times$.*

Theorem 25.2.3. *Let n be an odd integer. Then there always exists $v = (A, B) \in V(\mathbf{Z})$ such that $f_v = f$.*

Proof. Take $\alpha = 1$, $I = I_f^{(n-3)/2}$. □

Theorem 25.2.4. *Let n be an even integer. Assume $z^2 = f(x, y)$ has a rational point. Then there exists $v \in V(\mathbf{Z})$ such that $f = f_v$.*

Proof. We can further assume that $f(0, 1)$ is square, i.e. $f_n = c^2$ for some $c \in \mathbf{Z}$. Take $\alpha = \theta$ and

$$I = \langle c, \theta, \theta^2, \ldots, \theta^{(n-2)/2}, \zeta_{n/2}, \ldots, \zeta_{n-1} \rangle.$$

Check that $I^2 \subset \theta I_f^{n-3}$ and $\mathrm{N}(I^2) = \mathrm{N}(\theta)\,\mathrm{N}(I_f^{n-3})$. By the theorem of Wood, this gives rise to a pair $v = (A, B)$. □

If for example $n = 6$, there is an explicit formula for v:

$$(A, B) = \left(\begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & f_0 & f_1 & f_2 \\ 0 & 0 & 1 & f_1 & f_2 & f_3 \\ 0 & 1 & 0 & f_2 & f_3 & f_4 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & f_1 & f_2 & 0 \\ 0 & 1 & 0 & f_2 & 0 & 0 \\ c & 0 & 0 & 0 & 0 & -f_5 \end{pmatrix} \right).$$

Now count the number of orbits of $G(\mathbf{Z})$ on $V(\mathbf{Z})$ having bounded height.

Theorem 25.2.5. *The number of orbits of $G(\mathbf{Z})$ on $V(\mathbf{Z})$ having height $< X$ is*

$$C \cdot X^{n+1} = C(X^{1/n})^{n(n+1)} + O(X^{x+1-1/n}).$$

So the number of orbits per f is bounded by a constant C' on average.

This is not good enough for our purposes. We need the number of orbits per f to be strictly less than 1 on average. But we are only interested in some orbits. The number of orbits per f locally looking like f_v is $C'' < C'$, where $C'' = o(2^{-g})$.

Corollary 25.2.6. *As C ranges over hyperelliptic curves of genus g,* $\operatorname{avg}(\#\operatorname{Sel}_2^{\text{fake}}(C)) = o(2^{-g})$.

26 Selmer groups and heuristics II

The main references for this lecture are [BKL+13] and *Boundedness of rank*, by Derek Garton, Jennifer Park, Jon Voight, Melanie Matchett Wood.

In section 20, we gave a conjecture predicting the average size of $\operatorname{Sel}_n E$ as E ranges over elliptic curves over $\mathbf{Q}$. At the end, we passed to the p^∞-component: there is an exact sequence

$$0 \longrightarrow E(\mathbf{Q}) \otimes (\mathbf{Q}_p/\mathbf{Z}_p) \longrightarrow \operatorname{Sel}_{p^\infty}(E) \longrightarrow Ш(E)[p^\infty] \longrightarrow 0. \qquad (\text{Seq}_E)$$

We are going to try to predict the distribution of this sequence.

26.1 The orthogonal Grassmannian

Recall that for a ring A, A-valued points $\operatorname{Gr}_{n,2n}(A)$ of the $(n,2n)$-Grassmannian consists of locally free rank-n A-submodules $Z \subset A^{2n}$ such that Z is a direct summand. There is a subfunctor $\operatorname{OGr}_n \subset \operatorname{Gr}_{n,2n}$, for which $\operatorname{OGr}_n(A)$ is the set of $Z \in \operatorname{Gr}_{n,2n}(A)$ such that $Q|_Z = 0$.

Both the functors $\operatorname{Gr}_{n,2n}$ and OGr_n are represented by smooth projective schemes over $\mathbf{Z}$. The orthogonal Grassmannian OGr_n has two connected components. For any field k, $Z, Z' \in \operatorname{OGr}_n(k)$ are in the same component if and only if $\dim(Z \cap Z') \equiv n \pmod 2$.

The fact that OGr_n is smooth tells us that the fibers in each level of the inverse system

$$\cdots \to \operatorname{OGr}_n(\mathbf{Z}/p^{e+1}) \to \operatorname{OGr}_n(\mathbf{Z}/p^e) \to \operatorname{OGr}_n(\mathbf{Z}/p^{e-1}) \to \cdots$$

have constant cardinality. We get a "uniform" probability measure on $\operatorname{OGr}_n(\mathbf{Z}_p) = \varprojlim \operatorname{OGr}_n(\mathbf{Z}/p^e)$ by taking the inverse limit of the uniform measure on each $\operatorname{OGr}_n(\mathbf{Z}/p^e)$.

26.2 The model

Start with $V = \mathbf{Z}_p^{2n}$. Fix $W = \mathbf{Z}_p^n \times 0 \subset V$; this is maximal isotropic. Choose a random $Z \in \operatorname{OGr}_n(\mathbf{Z}_p)$. Form the sequence

$$0 \to (Z \cap W) \otimes (\mathbf{Q}_p/\mathbf{Z}_p) \to (Z \otimes \mathbf{Q}_p/\mathbf{Z}_p) \cap (W \otimes \mathbf{Q}_p/\mathbf{Z}_p) \to \text{cokernel} \to 0. \qquad (\text{Seq}_Z)$$

So from a random $Z \in \operatorname{OGr}_n(\mathbf{Z}_p)$ we get a random sequence of co-finitely generated $\mathbf{Z}_p$-modules. The following theorem is proved in [BKL+13].

Theorem 26.2.1. *The limit as $n \to \infty$ of the distribution of* Seq_V *exists.*

Conjecture 26.2.2. *The limit as $n \to \infty$ of the distribution of the Seq_V is the distribution of Seq_E for $E \in \mathcal{E}$. That is, if $\mathcal{S}$ is a short exact sequence of $\mathbf{Z}_p$–modules, then*

$$\lim_{n\to\infty} \mathrm{Prob}_{Z\in \mathrm{OGr}_n(\mathbf{Z}_p)}(\mathrm{Seq}_Z \simeq \mathcal{S}) = \mathrm{Prob}_{E\in\mathcal{E}}(\mathrm{Seq}_E \simeq \mathcal{S}).$$

26.3 Consequences for rank

We have $(Z\cap W)\otimes \mathbf{Q}_p/\mathbf{Z}_p = (\mathbf{Q}_p/\mathbf{Z}_p)^r$, where $r = \dim_{\mathbf{Q}_p}(Z\otimes\mathbf{Q}_p \cap W\otimes\mathbf{Q}_p)$. Outside a (measure zero) lower-dimensional locus, the rank is 0 for Z in one component of OGr, and 1 for Z in the other component.

Corollary 26.3.1. *The conjecture implies that 50% of elliptic curves have rank 0, 50% have rank 1, and 0% have rank $\geqslant 2$.*

26.4 Consequences for Ш

Write Seq_Z as $0 \to R \to S \to T \to 0$.

Proposition 26.4.1. *In Seq_Z, R is the maximal divisible subgroup of S and T is finite (for each n).*

Corollary 26.4.2. *The conjecture implies that $Ш(E)[p^\infty]$ is finite for 100% of elliptic curves over $\mathbf{Q}$.*

There are three distributions on the set of isomorphism classes of finite abelian p-groups, each conjectured to be the distribution of $Ш(E)[p^\infty]$ for $E \in \mathcal{E}$ of rank r.

Prediction 26.4.3. *This is due to Delaunay [Del01]. As E ranges over elliptic curves over $\mathbf{Q}$ with rank r, the distribution of $Ш(E)[p^\infty]$ is*

$$\mathrm{Prob}(G) = \frac{\#G^{1-r}}{\#\,\mathrm{Aut}(G,[\cdot])} \prod_{i=r+1}^{\infty}(1-p^{1-2i}).$$

Prediction 26.4.4. *This is due to [BKL+13]. Choose $Z \in \mathrm{OGr}_n(\mathbf{Z}_p)$ such that $\mathrm{rk}(Z\cap W) = r$. Form $0 \to R \to S \to T \to 0$, and take*

$$\lim_{n\to\infty} (\text{distribution of } T).$$

Then $Ш(E)[p^\infty]$ is distributed according to this limit.

Prediction 26.4.5. *This is also due to [BKL+13]. Choose random $A \in M_{2n+r}(\mathbf{Z}_p)$ such that $\mathrm{rk}\, A = 2n$ and ${}^{\mathrm{t}}A = -A$. We have an exact sequence*

$$\mathbf{Z}_p^{2n+r} \xrightarrow{A} \mathbf{Z}_p^{2n+r} \longrightarrow \mathrm{cok}\, A \longrightarrow 0.$$

Then $Ш(E)[p^\infty]$ is distributed as

$$\lim_{n\to\infty} (\text{distribution of } (\mathrm{cok}\, A)_{\mathrm{tors}}).$$

This is similar to the Friedman-Washington heuristic for class groups, which chooses a random $A \in M_n(\mathbf{Z}_p)$, and takes

$$\lim_{n\to\infty} (\text{distribution of } \operatorname{cok} A).$$

They prove that this distribution is identical to the Cohen-Lenstra distribution for $\mathrm{Cl}(k)[p^\infty]$, as k ranges over imaginary quadratic fields.

The following is one of the main theorems of [BKL+13].

Theorem 26.4.6. *For each $r \geqslant 0$, the three distributions above coincide.*

26.5 Heuristics for boundedness of $\operatorname{rk} E$

Consider an elliptic curve E with conductor / discriminant / height N. Choose a random $A \in M_n(\mathbf{Z})$ such that ${}^tA = -A$, with entries of size $X = X(N)$. The function X will be determined later. Then

- $(\operatorname{cok} A)_{\text{tors}}$ models $Ш(E)$
- $\operatorname{rk}_{\mathbf{Z}}(\operatorname{cok} A)$ models $\operatorname{rk} E$

The following has not been previously written up. For each $r \geqslant 1$, $\operatorname{Prob}(\operatorname{rk} E \geqslant r)$ should be modeled by $\operatorname{Prob}(\operatorname{rk} A \leqslant n - r)$. It is a "probable theorem" that

$$\operatorname{Prob}(\operatorname{rk} A \leqslant n-r) \sim \frac{X^{n(n-r)/2}}{X^{n(n-1)/2}} \sim \frac{1}{(X^{n/2})^{r-1}}.$$

where $n \equiv r \pmod 2$. (The analogous problem for symmetric integer matrices is an honest theorem proved in [EK95].)

There is a heuristic to suggest $X^{n/2} \sim N^{1/24}$. When $r = 2$, random matrix theory already suggests an answer. Consider E with the sign in the functional equation being $+1$ (so E should have even rank). Define

$$\#Ш_0 = \begin{cases} \#Ш & \operatorname{rk} E = 0 \\ 0 & \text{otherwise} \end{cases}$$

Part of the Birch and Swinnerton-Dyer conjecture says that

$$L(E,1) = \frac{\#Ш_0 \Omega \prod c_p}{\#E(\mathbf{Q})^2_{\text{tors}}}.$$

(If $\operatorname{rk} E > 0$, then both sides should be zero.) Solving for $\sqrt{\#Ш_0}$ (and throwing out $\#E(\mathbf{Q})_{\text{tors}}$, the Tamagawa numbers, and $L(E,1)$) we get

$$\sqrt{\#Ш_0} \sim O(\Omega^{-1/2}),$$

and $\Omega \sim N^{-1/12}$. Thus

$$X^{-n/2} \sim \operatorname{Prob}(\operatorname{rk} E \geqslant 2) = \operatorname{Prob}(\sqrt{\#Ш_0} = 0) \sim N^{-1/24}$$

so we conclude that

$$\mathrm{Prob}(\mathrm{rk}\, E \geqslant r) \sim N^{(1-r)/24}.$$

There are $N^{5/6}$ elliptic curves of height $\leqslant N$. If $r - 1 > 20$, we would expect finitely many elliptic curves of rank $\geqslant r$.

Prediction 26.5.1. $\mathrm{rk}\, E \leqslant 21$, *with finitely many exceptions.*

In particular, there is a global bound for the rank of elliptic curves over $\mathbf{Q}$.

27 Pencils of quadrics: the geometry

Pencils of quadrics have shown up many times (though not under that name) in this summer school. I will explain some of the geometry of quadrics.

27.1 Notation

Let k be a perfect field of characteristic not 2. Let $\mathscr{L}$ be a rational generic pencil of quadrics in $\mathbf{P}^{2n+1}$. Such an $\mathscr{L}$ will be of the form

$$\{xQ - yQ_2 : (x : y) \in \mathbf{P}^1\}$$

where $Q_1, Q_2 \subset \mathbf{P}^{2n+1}$ are quadrics. "Rational" means the Q_i are defined over k. "Generic" means the binary form $f(x, y) = (-1)^{n+1} \det(xQ_1 - yQ_2)$ has no repeated factors, i.e. $\mathrm{Disc}(f) \neq 0$. Alternatively, $C : z^2 = f(x, y)$ should be a smooth hyperelliptic curve of genus n. The *base locus* $B = Q_1 \cap Q_2$ will be smooth of dimension $2n - 1$.

Let F be the variety of maximal linear subspaces of B. That is:

$$F = \{X \simeq \mathbf{P}^{n-1} : X \subset B\}.$$

For example, when $n = 1$, we have two quadrics $Q_1, Q_2 \subset \mathbf{P}^3$. We have $F = B = Q_1 \cap Q_2$, a genus one curve. So over $\bar{k}$, F is isomorphic to an elliptic curve. This is basically the construction Bhargava used to study Selmer elements of elliptic curves.

If $n = 2$, $B = Q_1 \cap Q_2$ is a degree 4 three-fold. The variety F turns out to be an abelian surface (over $\bar{k}$).

This apparent pattern holds.

Theorem 27.1.1 (Reid, Donagi, Desale-Ramanan)**.** *Over $\bar{k}$, the variety $F \simeq \mathrm{Jac}\, C$; an abelian variety of dimension n.*

To obtain arithmetic information about C, we need a result that works over an arbitrary (possibly non algebraically closed) base field.

Theorem 27.1.2. *The variety F is a $J = \mathrm{Jac}(C)$-torsor. Moreover, there is an algebraic group structure on $G = J \sqcup F \sqcup J^1 \sqcup F$ compatible with that of J, and for which $G/J \simeq \mathbf{Z}/4$.*

Here $J^1 = \mathrm{Pic}^1(C)$, the moduli space of degree-1 line bundles on C.

For $n = 1$, $Q_1, Q_2 \subset \mathbf{P}^3$, we had $F = Q_1 \cap Q_2$. The curve C is defined by $z^2 = \det(xQ_1 - yQ_2)$. The curve $\det(xQ_1 - yQ_2)$ is cut out by a binary quartic form. There is a canonical isomorphism $H^1 = C$, so we have a group structure on $G = E \sqcup F \sqcup F \sqcup F$. Multiplication by 2 on G gives a map $2 : C \to E$; this is a 2-cover of E. This is the cover used in the study of 2-Selmer groups of elliptic curves. Multiplication by 4 gives a map $4 : F \to E$; this is the 4-cover used in the study of 4-Selmer groups of elliptic curves.

For general $n \geqslant 2$, there are a couple cases.

Case 1: $C(k) \neq \varnothing$. Choose $\infty \in C(k)$. Put $F[2]_\infty = \{X \in F : X + X = (\infty)\}$; this is a $J[2]$-torsor. So we get an element of $\mathrm{H}^1(k, J[2])$. There are two subcases corresponding to whether or not ∞ is a Weierstrass point. If C has a Weierstrass point, we get all torsors of $J[2]$ in this way. When ∞ is a non-Weierstrass point, we don't get all of $\mathrm{H}^1(k, J[2])$, but we do get "enough" points in $\mathrm{H}^1(k, J[2])$, namely the entire kernel of $\gamma : \mathrm{H}^1(k, J[2]) \to \mathrm{H}^1(k, ?)$ as in section 24.

Case 2: the map $2 : F \to J^1$ is a 2-cover of J^1. If k is a global field and C is everywhere locally soluble, we get all locally soluble 2-covers of J^1 using this method.

28 Arithmetic invariant theory and hyperelliptic curv II

28.1 Redefining hyperelliptic curves

We'll start with a slightly more general definition of a hyperelliptic curve. If C is a curve of genus $g \geqslant 1$ over a field k, then $\mathrm{H}^0(C, \Omega^1)$ is a g-dimensional k-vector space. It is known that this has no base points. So we get a map $\pi : C \to \mathbf{P}(\mathrm{H}^0(\Omega^1)) \simeq \mathbf{P}^{g-1}$; this is called the *canonical map*. If $g \geqslant 2$, you can use the Riemann-Roch theorem to prove that this map is either an embedding with image a smooth curve of degree $2g-2$, or it's 2-to-1 onto a rational normal curve X of degree $g-1$. Say C is *hyperelliptic* if we're in the second case.

If $X(k) \neq \varnothing$, then we get a map $C \to \mathbf{P}^1$ and recover the standard definition of a hyperelliptic curve. This always happens if g is even. But if g is odd, we might not have a rational point. The image X of $\pi : C \to \mathbf{P}^{g-1}$ will be of the form $\{Q(x, y, z) = 0\}$.

[...stuff I didn't catch...]

We define $C \subset \mathbf{P}(1, 1, 1, \frac{g+1}{2})$.

We don't really need this, because we will be studying C with local points everywhere. Via $C \to X$, the curve X has local points everywhere. By the Hasse principle, $X(k) \neq \varnothing$, so $X \simeq \mathbf{P}^1$. Thus $C \to X$ realizes C as a curve of the form $z^2 = F(x, y)$ inside $\mathbf{P}(1, 1, g+1)$. For the rest of this lecture, C will be a hyperelliptic curve defined by an equation of this form.

28.2 Main result

The following is a strenthening of the theorem Bhargava proved in section 25.

Proposition 28.2.1. *A positive proportion of such C have no rational points over any extension of odd degree over $\mathbf{Q}$.*

For heuristic reasons, we suspect the proportion is 3/4.

Let J be the Jacobian of C; this is an abelian variety over k of dimension g. For each n, there is a variety $J^n = \mathrm{Pic}^n(C)$ which classifies line bundles of degree n over C. Each J^n is a J-torsor. For hyperelliptic curves, there is a canonical element $h = \pi^*\mathscr{O}(1) \in J^2(k)$ coming from the degree-two map $C \to \mathbf{P}^1$. (Just pull back any point in $\mathbf{P}^1$.) Thus $J^r \simeq J^{r+2n}$ for all n. The degree-one part J^1 is especially important because the Abel map $C \to J^1$ given by $x \mapsto [x]$ is an embedding defined over k.

Proposition 28.2.2. *The following are equivalent:*

- *C has no rational points over any extension of odd degree*
- *$J^1(\mathbf{Q}) = \varnothing$*

Proof. Indeed, if $x \in C(L)$ and $[L : \mathbf{Q}] = 2n+1$, then $\sum[x^\sigma]$ will be a divisor of degree $2n+1$, hence an element of $J^1(\mathbf{Q})$. For the converse, we only have to prove that J^1 is *not* isomorphic to J over $\mathbf{Q}$, because any element of $J^1(\mathbf{Q})$ gives an isomorphism $J \xrightarrow{\sim} J^1$. □

We're implicitly using the fact that the curve has local points everywhere. There is a "Brauer obstruction class" measuring whether a rational divisor class comes from a rational divisor. When our curve is everywhere locally soluble, the Brauer class vanishes locally, so in this case it vanishes.

28.3 Fundamental groups

There is a beautiful idea, going back to Serre, Grothendieck, ... that studies varieties via their unramified covers. We will distinguish J from J^1 by studying their *arithmetic fundamental groups*. In particular, we will study their unramified 2-covers.

Recall that a *2-covering* of J is a J-torsor F with an étale covering $\pi : F \to J$ such that $\pi(f + a) = \pi(f) + 2$. So $\pi^{-1}(0)$ is a $J[2]$-torsor. It follows that $\pi : F \to J$ has degree 2^{2g}. There is an obvious notion of equivalence of 2-covers: via commutative diagrams

$$\begin{array}{ccc} F' & \xrightarrow{\pi} & J \\ \downarrow \wr & & \| \\ F & \xrightarrow{\pi} & J. \end{array}$$

For a 2-covering F, put $F[2] = \pi^{-1}(0)$; the 2-covering F is completely determined by the class of $F[2]$ in $\mathrm{H}^1(\mathbf{Q}, J[2])$. Solvable 2-coverings correspond to the image of $J(\mathbf{Q})/2 \hookrightarrow \mathrm{H}^1(\mathbf{Q}, J[2])$. Put $\mathrm{Pic}(C) = \coprod_{n\in\mathbf{Z}} J^n$; the group $\mathrm{Pic}(C)/\mathbf{Z}h = J \sqcup J^1$. So multiplication by 2 induces a 2-cover $J^1 \to J$. For this cover, $\pi^{-1}(0) = \{f \in J^1 : 2f = h\}$, which we call $W[2]$. For 100% of hyperelliptic curves, $W[2]$ is nontrivial. But $W[2]$ is locally soluble, so it gives a class in $\mathrm{Sel}_2(J)$. Since $W[2]$ is nontrivial, we see that 100% of the time, $\mathrm{Sel}_2(J) \neq 0$.

We can even define 2-covers of J^1. They are maps $\pi : F \to J^1$ such that $\pi(f+a) = \pi(f) + 2a$. Composing with $J^1 \to J$, we get a 4-cover $F \to J$. So $\pi^{-1}(0_J)$ is a $J[4]$-torsor. Let $\mathrm{Sel}_2(J^1)$ be the set of locally soluble 2-covers of J^1. We will distinguish between J and J^1 by showing that on average, $\#\,\mathrm{Sel}_2(J^1) < \#\,\mathrm{Sel}_2(J)$. Thus there will be many curves with $J \not\simeq J^1$. Note that

$$\mathrm{Sel}_2(J^1) = \{\alpha \in \mathrm{Sel}_4(J) : 2\alpha = W[2]\}.$$

So if $W[2]$ is not divisible by 2, $J^1 \not\simeq J$. We have a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & J(\mathbf{Q})/4 & \longrightarrow & \mathrm{Sel}_4\, J & \longrightarrow & Ш(J)[4] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & J(\mathbf{Q})/2 & \longrightarrow & \mathrm{Sel}_2(J) & \longrightarrow & Ш(J)[2] & \longrightarrow & 0. \end{array}$$

We will try to count the Selmer set $\mathrm{Sel}_2(J^1)$. Pencils $xA - yB$ of quadrics in $\mathbf{P}^{2g+1}(\mathbf{Q})$ give 2-coverings $\pi : F \to J^1$. This we saw in section 27.

We study the action of $\mathrm{SL}(2g+2)$ on $\mathrm{Sym}^2(2g+2) \oplus \mathrm{Sym}^2(2g+2)$. Using Bhargava's methods, we show that $\mathrm{avg}(\#\,\mathrm{Sel}_2\, J^1) \leqslant 2 = \tau(\mathrm{SL}_{2g+2})$. This is (just barely) enough. We know that $\mathrm{avg}(\#\,\mathrm{Sel}_2\, J) \geqslant 2$. So we need to show that a positive proportion of the time, $\mathrm{avg}(\#\,\mathrm{Sel}_2\, J) > 2$. The Dokchitser brothers have shown that the parity of $\mathrm{Sel}_2\, J$ is equidistributed. So there have to be curves with $J \not\simeq J^3$. A better result would be $\mathrm{avg}(\#\,\mathrm{Sel}_2\, J) = 3$, but even this is not good enough.

Note: one usually creates abelian covers of C through the fundamental group of its Jacobian. But when C has no rational points, C embeds into J^1, not J, so we would have to look at covers of J^1. But these don't always exist!

Proposition 28.3.1. *Let $f(x,y)$ be a binary form of degree $2g+2$ with $\Delta \neq 0$ and $f_0 \neq 0$, defined over $\mathbf{Q}$. The following are equivalent:*

1. *There exists an orbit (A,B) with $\mathrm{Disc}(xA - yB) = f(x,y)$.*
2. *$f_0 \in (\mathbf{Q}^\times)^2\, \mathrm{N}(L^\times)$.*
3. *$J^1_{\mathfrak{m}}$ is divisible by 2 in $\mathrm{H}^1(\mathbf{Q}, J_{\mathfrak{m}})$, where $\mathfrak{m} = [p_\infty] + [p'_\infty]$ and $C_{\mathfrak{m}} : z^2 = y^2 f(x,y)$ has genus $g+1$.*
4. *The maximal abelian 2-cover of C ramified only at $\{p_\infty, p'_\infty\}$ descends to a cover $D \to C$ over $\mathbf{Q}$.*

29 Chabauty methods and hyperelliptic curves

29.1 Introduction

For an elementary introduction to Chabauty's method, see [MP12].

Theorem 29.1.1 (Faltings, 1983)**.** *If C is a curve of genus $\geqslant 2$ over $\mathbf{Q}$, then $C(\mathbf{Q})$ is finite.*

The proof of this is "bad" in the sense that it is highly ineffective. It does produce an upper bound on the number of rational points, but neither Faltings' or Vojta's later proof give horrible upper bounds for $\#C(\mathbf{Q})$, and give no upper bound on the height of those points.

Much earlier Chabauty created a more effective method for attacking this problem for certain families of curves.

Theorem 29.1.2 (Poonen, Stoll, 2013). *Let $C : y^2 = f(x)$, where $\deg f = 2g + 1$.*

1. *For each $g \geqslant 3$, the fraction of such C satisfying $C(\mathbf{Q}) = \{\infty\}$ is positive.*
2. *This fraction tends to 1 as $g \to \infty$. More precisely, it is $\geqslant 1 - (12g + 20)g^{-g}$.*
3. *Chabauty's method at the prime 2 effectively determines $C(\mathbf{Q})$ for such a fraction of curves.*

Proof. See [PS13]. □

Conjecturally, 100% of such C have $C(\mathbf{Q}) = \{\infty\}$. This theorem is essentially the first case in which an effective version of Faltings' theorem was proven for a large class of curves.

29.2 Genus one

Let E be an elliptic curve over $\mathbf{C}$. Then $E(\mathbf{C}) \simeq \mathbf{C}/\Lambda$ for some discrete subgroup $\Lambda \subset \mathbf{Z}^2$ in $\mathbf{C}$. The differential dz is a well-defined holomorphic one-form on $\mathbf{C}/\Lambda$, corresponding to an algebraic differential $\omega \in \Omega^1(E)$. If we had started with ω, we can define $E(\mathbf{C}) \to \mathbf{C}/\Lambda$ by

$$x \mapsto \int_0^x \omega.$$

Changing the path $0 \to x$ changes the integral by an element in the discrete lattice $\Lambda = \{\int_\gamma \omega : \gamma \in \pi_1 E(\mathbf{C})\}$.

29.3 Genus g

Let C be a curve of genus g over $\mathbf{C}$. Then

$$\{\text{holomorphic 1-forms on } C\} = \Gamma(C, \Omega^1).$$

One definition of the genus of C is that $h^0(\Omega^1) = g$, i.e. $\Gamma(C, \Omega^1)$ is g-dimensional. Let $\omega_1, \ldots, \omega_g$ be a basis. Fix $0 \in C(\mathbf{C})$. Define a holomorphic map $i : C(\mathbf{C} \to \mathbf{C}^g/\Lambda$ by

$$x \mapsto \left(\int_0^x \omega_1, \ldots, \int_0^x \omega_g \right).$$

This is called the Abel-Jacobi map. If z_i are the coordinates on $\mathbf{C}^g/\Lambda$, we have $i^* \mathrm{d}z_j = \omega_j$.

As mentioned in section 28, Weil discovered an algebraic analogue of this. Let C be a curve of genus g over any field k, with chosen $0 \in C(k)$. Then there exists a g-dimensional abelian variety J, the *Jacobian* of C such that

- For any field $L \supset k$,

$$J(L) \simeq \mathrm{Pic}^0(C_L) = \mathrm{Div}^0(C_L)/\text{linear equivalence}.$$

- There is a morphism $i : C \to J$, $x \mapsto [x] - [0]$, which is an embedding if $g \geqslant 1$.
- The pullback map $i^* : \Gamma(J, \Omega^1) \to \Gamma(C, \Omega^1)$ is an isomorphism.
- If $k = \mathbf{C}$, the analytic Abel-Jacobi map factors as

$$C(\mathbf{C}) \xrightarrow{i} J(\mathbf{C}) \xrightarrow{\sim} \mathbf{C}^g/\Lambda.$$

One of the difficulties of dealing with curves of higher genus is that $C(k)$ is not naturally a group. But at least $C \hookrightarrow J$, and $J(k)$ is a group.

29.4 (Bad) real-analytic approach

Let C be a curve of genus $g \geqslant 2$ with $0 \in C(\mathbf{Q})$. Let $i : C \hookrightarrow J$. First let's try a (bad) real-analytic approach to proving finiteness of $C(\mathbf{Q})$. We have a commutative diagram

$$\begin{array}{ccc} C(\mathbf{Q}) & \longrightarrow & J(\mathbf{Q}) \\ \downarrow & & \downarrow \\ C(\mathbf{R}) & \longrightarrow & J(\mathbf{R}). \end{array}$$

The group $J(\mathbf{Q})$ is finitely generated by Mordell-Weil, let $r = \mathrm{rk}\, J(\mathbf{Q})$. Since J is projective, $J(\mathbf{R})$ is a compact real Lie group, so $J(\mathbf{R}) \simeq \mathbf{R}^g/\mathbf{Z}^g \oplus \text{finite}$. Note that $C(\mathbf{Q}) = J(\mathbf{Q}) \cap C(\mathbf{R}) \subset J(\mathbf{R})$. But typically the group generated by some $x \in C(\mathbf{R}) \subset J(\mathbf{R})$ is dense inside $J(\mathbf{R})$. Alternatively, the closure of $J(\mathbf{Q})$ in the classical topology is often open in $J(\mathbf{R})$.

29.5 p-adic approach

Chabauty suggested replacing $\mathbf{R}$ with $\mathbf{Q}_p$. Again we have a commutative diagram

$$\begin{array}{ccc} C(\mathbf{Q}) & \longrightarrow & J(\mathbf{Q}) \\ \downarrow & & \downarrow \\ C(\mathbf{Q}_p) & \longrightarrow & J(\mathbf{Q}_p). \end{array}$$

We have $C(\mathbf{Q}) \subset C(\mathbf{Q}_p) \cap \overline{J(\mathbf{Q})}$, where here the closure is taken in the p-adic analytic topology.

Theorem 29.5.1 (Chabauty 1941)**.** *If $r < g$, then $C(\mathbf{Q}_p) \cap \overline{J(\mathbf{Q})}$ is finite. In particular, $C(\mathbf{Q})$ is finite.*

Often, the points in $C(\mathbf{Q}_p) \cap \overline{J(\mathbf{Q})}$ can be approximated p-adically.

29.6 Structure of $J(\mathbf{Q}_p)$

For simplicity, assume C has good reduction at p. So C extends to a smooth proper curve over $\mathrm{Spec}(\mathbf{Z}_p)$. Weil's construction of the Jacobian works in great generality, so $J = \mathrm{Jac}\, C$ also has good reduction, i.e. extends to an abelian scheme over $\mathrm{Spec}(\mathbf{Z}_p)$.

We can understand $J(\mathbf{Q}_p)$ by looking at a reduction map. First note that by the "valuative criterion for properness," $J(\mathbf{Q}_p) = J(\mathbf{Q}_p)$, so we have a homomorphism $J(\mathbf{Z}_p) \to J(\mathbf{F}_p)$, where $J(\mathbf{F}_p)$ is a finite abelian group. Let U be the kernel of this map. By smoothness, the map $J(\mathbf{Z}_p) \to J(\mathbf{F}_p)$ is surjective, so $J(\mathbf{Z}_p)$ is a disjoint union of $J(\mathbf{F}_p)$ copies of U. For suitable local coordinates $t_1, \ldots, t_g$ at 0, we get a p-adic analytic isomorphism $\boldsymbol{t} = (t_1, \ldots, t_g) : U \xrightarrow{\sim} (p\mathbf{Z}_p)^g$. We conclude that $J(\mathbf{Z}_p) \xrightarrow{\sim} \coprod_{J(\mathbf{F}_p)} (p\mathbf{Z}_p)^g$.

We would like a p-adic analogue of the Abel-Jacobi map. For $\omega \in \Gamma(J, \Omega^1)$, there is a canonical homomorphism $\eta_\omega : J(\mathbf{Q}_p) \to \mathbf{Q}_p$ which we write

$$x \mapsto \int_0^x \omega.$$

If $\omega = \sum w_j(\boldsymbol{t})\mathrm{d}t_j$ with $w_j \in \mathbf{Z}_p[\![t_1, \ldots, t_g]\!]$, just "integrate formally" and evaluate on $(p\mathbf{Z}_p)^g$. This definition works for $x \in U$. If we require η_ω to be a homomorphism, there is a unique extension of η_ω from U to $J(\mathbf{Q}_p)$. If $x \notin U$, there exists some $n \geqslant 1$ such that $nx \in U$; then define

$$\int_0^x \omega = \frac{1}{n} \int_0^{n \cdot x} \omega.$$

Putting everything together, we get a p-adic analytic homomorphism $\log : J(\mathbf{Q}_p) \to \mathbf{Q}_p^g$ defined by

$$\log x = \left(\int_0^x \omega_1, \ldots, \int_0^x \omega_g \right).$$

This is a local diffeomorphism.

29.7 Consequences

We know that $J(\mathbf{Q}_p) \simeq \mathbf{Z}_p^g \oplus \text{finite}$. If $r < g$, $\log J(\mathbf{Q}) \subset \mathbf{Q}_p^g$ will be contained in some hyperplane. Therefore there is some $0 \neq \omega \in \Gamma(J_{\mathbf{Q}_p}, \Omega^1)$ such that $\eta_\omega|_{J(\mathbf{Q})} = 0$. This gives an explicit way of finding a "smaller box" inside $J(\mathbf{Q}_p)$ in which $C(\mathbf{Q})$ fits.

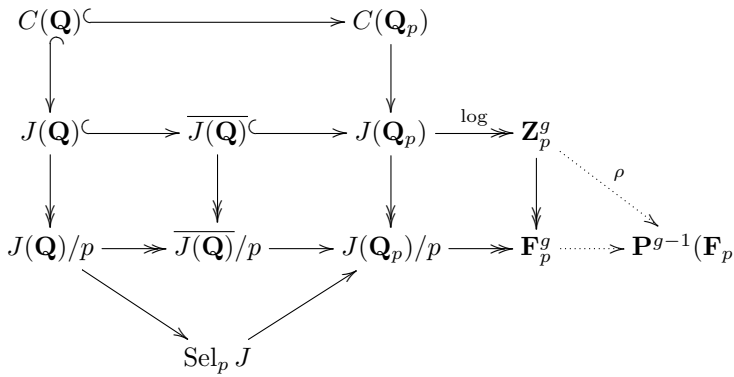
Just to recap, $C(\mathbf{Q}_p) \cap \overline{J(\mathbf{Q})}$ is a subset of $C9\mathbf{Q}_p) \cap \{\eta = 0\}$, the set of zeros of $\int_0^x \omega$ on $C(\mathbf{Q}_p)$. We can write $\omega = w(t)\,\mathrm{d}t$ with $w \in \mathbf{Z}_p[\![t]\!]$. Then

$$\eta = \int \omega \in \mathbf{Q}_p[\![t]\!].$$

We have some control on the Newton polygon of η. If we write $\eta = \sum a_i t^i$, plot the points $(i, v_p(i))$. The lower-convex hull of these points is the Newton polygon, and from this we can understand the valuations of the zeros of η. In particular, the number of zeros can be controlled in terms of g.

29.8 Main result

From [BG13], we know that $\operatorname{avg}(\#\operatorname{Sel}_2 J) = 3$. This wouldn't be sufficient, except that $\operatorname{Sel}_2 J$ carries "one 2-adic digit" of information about $J(\mathbf{Q})$. We have a commutative diagram



We know that $J(\mathbf{Q})/p$ lives inside $\operatorname{Sel}_p J$ in $J(\mathbf{Q}_p)/p$. We will concentrate on the case $p = 2$. We can compare the image of $\operatorname{Sel}_p J$ and $\rho \log C(\mathbf{Q}_p)$ inside $\mathbf{P}^{g-1}(\mathbf{F}_p)$. Bhargava and Gross proved an equidistribution result for Selmer elements.

30 Topological and algebro-geometric methods over function fields I

I will give a "sales pitch" for thinking about these problems in the context of global function fields. The idea is that the main problems can be approached more geometrically. Some problems in arithmetic statistics are much easier in the function field context.

30.1 Motivating examples

Question 30.1.1. *How many integers are there between N and $2N$?*

See section 6 for an interesting (and sophisticated) approach to this question.

Question 30.1.2. *How many squarefree integers are there between N and $2N$?*

Call this number $\operatorname{sf}(N)$. To be squarefree is to be indivisible by $4, 9, 25, 49, \ldots$, i.e. not divisible by p^2 for any prime p. One might expect "being indivisible by p^2" to be independent for distinct p, so

$$\begin{aligned}\operatorname{sf}(N) &\sim N \cdot \left(1 - \frac{1}{4}\right)\left(1 - \frac{1}{9}\right)\cdots \\ &= N \cdot \prod_p \left(1 - p^{-2}\right)^{-1} \\ &= \zeta(2)^{-1} N.\end{aligned}$$

So $\lim_{N\to\infty} \frac{\text{sf } N}{N} = \zeta(2)^{-1}$. This is a common phenomenon in arithmetic statistics – some kind of behavior asymptotically occurs an L-value percent of the time.

First, let's understand how this problem looks over more general global fields.

Definition 30.1.3. A *global field* is either

- A number field, i.e. a finite extension of $\mathbf{Q}$.
- The function field of a curve over a finite field $\mathbf{F}_q$. (Equivalently, a field isomorphic to a finite extension of $\mathbf{F}_q(t)$.)

We will be quite loose in identifying a curve over $\mathbf{F}_q$ and its function field, because there is an (anti-)equivalence of categories between smooth proper geometrically irreducible (aka "nice") curves over $\mathbf{F}_q$ and field extensions of $\mathbf{F}_q$ of transcendence degree 1.

30.2 The analogy between number fields and function fields

For a number field K, there is a unique embedding $\mathbf{Q} \hookrightarrow K$. But there might be *many* ways to embed $\mathbf{F}_q(t)$ into a global field K. For example, $\mathbf{F}_q(t^{17}) \subset \mathbf{F}_q(t)$ is a "non-standard" embedding of $\mathbf{F}_q(t)$ into $\mathbf{F}_q(t)$. So global function fields do not "come with" the structure of an extension of $\mathbf{F}_q(t)$. This phenomenon lies behind the fact that Mordell-Weil ranks are unbounded over function fields. (See examples of Ulmer.) In this lecture we'll mainly talk about $\mathbf{F}_q(t)$.

What is the function-field analogue of counting squarefree integers in a box? One problem is that $\mathbf{Q}$ has only one "nice" subring, whereas $\mathbf{F}_q(t)$ has lot of "nice" subrings. We'll use the following analogy:

number fields	function fields
$\mathbf{Q}$	$\mathbf{F}_q(t)$
$\mathbf{Z}$	$\mathbf{F}_q[t]$
$\lvert\cdot\rvert : \mathbf{Z} \to \mathbf{R}$	$\lvert f\rvert_\infty = q^{\deg f}$
$[N, 2N] = \{n \in \mathbf{N} : \lvert n\rvert \sim N\}$	set of monic polys with $\lvert f\rvert = N = q^n$
$\#(\mathbf{N} \cap [N, 2N]) \sim N$	$\#(\text{monic polys with } \lvert f\rvert = N) = N$
of these, $\sim \zeta_{\mathbf{Z}}(2)^{-1}N$ are squarefree	$\sim \zeta_{\mathbf{F}_q[t]}(2)^{-1}$ are squarefree

That is, the limiting proportion of squarefree monic polynomials in $\mathbf{F}_q[t]$ is

$$\prod_p \left(1 - |p|^{-2}\right) = 1 - q^{-1}.$$

as p ranges over monic irreducible polynomials in $\mathbf{F}_q[t]$. In fact, the number of squarefree monic polynomials of degree n in $\mathbf{F}_q[t]$ is exactly $q^n - q^{n-1}$ for all $n \geqslant 2$, and q for $n = 1$. So we have a power-saving result with *much* better error term over function fields. So we shouldn't think of there being an analogy between any particular number field and any particular function field. Rather, there is an analogy between the *class* of number fields and the *class* of function fields.

30.3 Geometric picture

What is geometric about what we've done? We introduce yet another function field, $\mathbf{C}(t)$. We can once again think about the set of monic squarefree polynomials of degree n in $\mathbf{C}[t]$. This set is not just a set – it is a *space* (namely an algebraic variety). The space of monic squarefree polynomials of degree n is called the (unordered) *configuration space* of $\mathbf{C}$, denoted $\mathrm{Conf}^n \mathbf{C}$. It parameterizes n-tuples of distinct points in $\mathbf{C}$, up to permutation. This isomorphism is given by $f \mapsto \{\text{roots of } f\}$. The inverse sends an n-tuple $(z_1, \ldots, z_n)$ to the polynomial $f(t) = (t - z_1) \cdots (t - z_n)$.

We are morally constrained to think of this configuration space not just as a complex manifold, but as a scheme over $\operatorname{Spec} \mathbf{Z}$. Namely, there is a scheme $\mathrm{Conf}^n \mathbf{A}^1$ over $\operatorname{Spec} \mathbf{Z}$ such that

$$(\mathrm{Conf}^n \mathbf{A}^1)(K) = \{\text{monic squarefree polynomials of degree } n \text{ in } K[t]\}$$

for any field K. In fact, this has a simple description. Namely, the moduli space of *all* monic polynomials of degree n is $\mathbf{A}^n$. A polynomial f is squarefree if and only if the discriminant $\Delta(f)$ is nonzero, where Δ is a polynomial in the coefficients of f. For example,

$$\Delta(t^2 + a_1 t + a_2) = a_1^2 - 4a_2.$$

So $\mathrm{Conf}^n \mathbf{A}^1$ is $\mathbf{A}^n \smallsetminus V(\Delta)$. Note: we would get a different space if we parameterized ordered n-tuples numbers, where we care about ordering. We'll call that PConf^n, the *pure configuration space*. The group S_n acts on PConf^n by permuting the n-tuples, and the quotient PConf^n / S_n is Conf_n. Note that $\mathrm{PConf}^n \mathbf{A}^1 = \mathbf{A}^n \smallsetminus \bigcup_{i \neq j} V(z_i - z_j)$, where $z_1, \ldots, z_n$ are the the coordinates of $\mathbf{A}^n$.

The set of monic squarefree polynomials in $\mathbf{F}_q[t]$ of degree n is just $\mathrm{Conf}^n \mathbf{A}^1(\mathbf{F}_q)$. So our counting problem is: what is $\#\,\mathrm{Conf}^n \mathbf{A}^1(\mathbf{F}_q)$? We saw that the answer is $q^n - q^{n-1}$.

What if we only cared about what happens as $q \to \infty$? For example, what is the probability that a degree-n polynomial over $\mathbf{F}_q[t]$ is squarefree? We had been fixing q and letting $n \to \infty$. A simpler question is fixing n and letting $q \to \infty$. As $q \to \infty$,

$$\lim_{q \to \infty} \frac{\#\,\mathrm{Conf}^n \mathbf{A}^1(\mathbf{F}_q)}{q^n} = \lim_{q \to \infty} \frac{\#(\mathbf{A}^n \smallsetminus V(\Delta))(\mathbf{F}_q)}{q^n} = 1.$$

30.4 Möbius functions

Question 30.4.1. *What is the average of the Möbius function?*

Recall the *Möbius function* μ is the arithmetic function defined by

$$\mu(n) = \begin{cases} 0 & n \text{ is squarefree} \\ 1 & n \text{ the product of an even number of distinct primes} \\ -1 & n \text{ the product of an odd number of distinct primes} \end{cases}$$

Earlier, we computed that the expected value of μ^2 is $\mathbf{E}(\mu^2) = \zeta(2)^{-1}$. How does this look for function fields? We can define the Möbius function of a polynomial in exactly the same way. But over function fields, we have the beautiful

Fact 30.4.2.

$$\mu(f) = (-1)^{\deg f} \left(\frac{\Delta(f)}{q} \right)$$

where $\left(\frac{\cdot}{q}\right)$ *is the Legendre symbol.*

Note that $(-1)^n \mu(f) + 1$ is the number of square roots of $\Delta(f)$ in $\mathbf{F}_q$, where $n = \deg f$. Let's make a variety geometrizing this problem. Define Y_n to be the space parameterizing pairs (f, y), where f is a monic squarefree degree n polynomial, and y is a square root of $\Delta(f)$. Then

$$\#Y_n(\mathbf{F}_q) = \sum_{f : \deg f = n} ((-1)^n \mu(f) + 1)$$

So $\#Y_n(\mathbf{F}_q) - q^n = (-1)^n \sum_f \mu(f)$. We expect

$$\#Y_n(\mathbf{F}_q) = q^n + o(q^n).$$

Why? There is a map $Y_n \to \mathbf{A}^n$ given by $(f, y) \mapsto f$. This is a double branched at the vanishing locus of Δ. In general, we expect an n-dimensional variety to have approximately q^n points over $\mathbf{F}_q$. But this expectation only is valid when the variety is irreducible. So we think $\#Y_n(\mathbf{F}_q) \sim q^n$ because we think Y_n is irreducible. Indeed, the Weil conjectures guarantee that if Y_n is geometrically irreducible, then

$$\lim_{q \to \infty} \frac{\#Y_n(\mathbf{F}_q)}{q^n} = 1,$$

so the limit as $q \to \infty$ of the average of the Möbius function is zero.

But how do we *actually know* that Y_n is irreducible? What if $\Delta(a_1, \ldots, a_n)$ were actually G^2 for some other polynomial G? Then $\mu(f)$ would be $(-1)^n$ for *all* f of degree n. I argue that the underlying idea here is a computation of monodromy.

30.5 Monodromy

Recall the S_n-Galois cover $\mathrm{PConf}^n \twoheadrightarrow \mathrm{Conf}^n$. The normal subgroup $A_n \subset S_n$ corresponds to an intermediate (degree-2) Galois cover $U_n \to \mathrm{Conf}^n$. In fact, the following diagram is Cartesian with the left arrow being an étale cover with group $\mathbf{Z}/2$:

$$\begin{array}{ccc} U_n & \longrightarrow & Y_n \\ \downarrow & & \downarrow \\ \mathrm{Conf}^n & \hookrightarrow & \mathbf{A}^n. \end{array}$$

It is sufficient to show that U_n is irreducible. A $\mathbf{Z}/2$-cover of Conf^n is a map $\pi_1(\mathrm{Conf}^n) \to \mathbf{Z}/2$, and the cover is irreducible if and only if this map is surjective. Whenever we have a "cover of moduli spaces" $Y \to X$ of degree n, we have a map $\pi_1(X) \to S_n$. The image of this map is called the *monodromy group* of the cover and Y is irreducible if and only if the monodromy group is transitive.

In section 33, we'll look at the idea that big monodromy implies "averages are what you expect" in the large q regime. Sometimes, monodromy is not big, and its "smallness" can sometimes explain the failure of of heuristics.

30.6 Computational question

This is related to the discussion of variation of Mordell-Weil ranks. As discussed above, there are $q^n - q^{n-1}$ squarefree onic polynomials of degree n in $\mathbf{F}_q[t]$. For each such $f(t)$, let

$$C_f : y = f(t)$$

be the corresponding hyperelliptic curve. Its zeta function has the form

$$\zeta(C_f, s) = \frac{P_f(q^{-s})}{(1-q^{-s})(1-q^{1-s})},$$

where $P_f \in \mathbf{Z}[X]$ has degree $2g$, and all its roots have absolute value $q^{1/2}$.

The question is: for how many f does P_f have $q^{1/2}$ as a root. Does this proportion look like $q^{\alpha n}$ for some $0 < \alpha < 1$?

31 Counting methods over global fields

In the previous lectures, we have seen how to parameterize objects of arithmetic interest by looking at orbits of group actions, and we have counted these orbits using analytic methods. In this lecture, we'll generalize the techniques to global fields.

31.1 Terminology

A *global field* K is one of the following:

- number field (finite extension of $\mathbf{Q}$)
- finite separable extension of $\mathbf{F}_q(T)$

For global fields of finite characteristic, we are implicitly choosing a map from the corresponding curve to $\mathbf{P}^1$. With that map, we can always define a ring of integers $\mathcal{O}$, and a set M_∞ of infinite places. Our underlying example will be the average size of $\mathrm{Sel}_2(E/K)$.

Throughout, K is a global field, not of characteristic 2 or 3.

31.2 Heights for global fields

Let K be a global field. An elliptic curve E over K can be written as $E : y^2 = x^3 + Ax + B$ with $A, B \in K$. We think of $(A, B) \in \mathbf{P}(4, 6)$, where $\mathbf{P}(4, 6) = \mathbf{G}_\mathrm{m} \backslash \mathbf{A}^2 \smallsetminus 0$ via the action $\alpha \cdot (A, B) = (\alpha^4 A, \alpha^6 B)$. If $(A, B) \in \mathbf{A}^2(K)$, define a fractional ideal

$$I = \{\alpha \in K : \alpha \cdot (A, B) \in \mathbf{A}^2(\mathcal{O})\}.$$

Set

$$H(A, B) = \mathrm{N}(I) \prod_{v \in M_\infty} \max\left\{|A|_v^{1/4}, |B|_v^{1/6}\right\}.$$

A simple application of the product formula shows that this height is invariant under the action of $\mathbf{G}_\mathrm{m}$. Unfortunately, the set $\mathbf{A}^2(K)_{<X} = \{x \in \mathbf{A}^2(K) : H(x) < X\}$

might not be "bounded." The solution is to construct a nice fundamental domain for $\mathbf{G}_{\mathrm{m}}(K)\backslash S(K)$ (Here and elsewhere $S = \mathbf{A}^2 \smallsetminus 0$) so that $(\mathbf{G}_{\mathrm{m}}(K)\backslash S(K)) \cap S(K)_{<X}$ is bounded.

31.3 Orbit parameterization over K

There is a bijection between $\mathrm{Sel}_2(E)$ and the set of locally soluble orbits for the action of $G(K)$ on $V(K)$. Here $G = \mathrm{PGL}(2)$ and $V = \mathrm{Sym}^4(2)$. If K has characteristic not 2 or 3, everything works fine.

31.4 Locally soluble K-orbits to integral orbits

The key input over $\mathbf{Q}$ is that if $v \in V(\mathbf{Q}_p)^{\mathrm{sol}}$ with integral invariants, then there exists $g \in G(\mathbf{Q}_p)$ such that $gv \in V(\mathbf{Z}_p)$. Unlike our definition of heights, which used $h_{\mathbf{Q}} = 1$ and needed to be modified for general K, things here translate easily.

Lemma 31.4.1. *If $v \in V(K_{\mathfrak{p}})^{\mathrm{sol}}$ with invariants in $\mathcal{O}_{\mathfrak{p}}$, $\mathfrak{p} \notin M_\infty$, then there exists $g \in G(K_{\mathfrak{p}})$ such that $gv \in V(\mathcal{O}_{\mathfrak{p}})$.*

Morally, replace $\mathbf{Q}$ with K, $\mathbf{Q}_p$ with $K_{\mathfrak{p}}$, and $\mathbf{Z}_p$ with $\mathcal{O}_{\mathfrak{p}}$. But we need to be careful: there is a fundamental difference in behavior between $\mathbf{Z}$ and general $\mathcal{O}_K$.

Suppose $v \in V(K)^{\mathrm{ls}}$ with invariants in $\mathcal{O}$. Then for all $\mathfrak{p} \notin M_\infty$, there exists $g_{\mathfrak{p}} \in G(K_{\mathfrak{p}})$ such that $g_{\mathfrak{p}} v \in V(\mathcal{O}_{\mathfrak{p}})$. Put $g = (g_{\mathfrak{p}})_{\mathfrak{p} \notin M_\infty} \in G(\mathbf{A}_f)$, where $\mathbf{A}_f$ is the ring of finite adeles. Inside $G(\mathbf{A}_f)$ are two subgroups. One is $U = \prod_{\mathfrak{p} \notin M_{\mathfrak{p}}} G(\mathcal{O}_{\mathfrak{p}})$, the other is $G(K)$. If K has trivial class group, the double quotient $U\backslash G(\mathbf{A}_f)/G(K)$ will be trivial, but in general it is only finite. For number fields, this due to Borel [Bor63], and for function fields this is due to Conrad [Con12]. Put

$$G(\mathbf{A}_f) = \coprod_{\beta \in \mathrm{Cl}(G)} U \beta G(K).$$

There exists $g'_{\mathfrak{p}} \in G(\mathcal{O}_{\mathfrak{p}})$, $\beta \in \mathrm{Cl}(G)$, $h \in G(K)$ such that for all $\mathfrak{p} \notin M_\infty$ we have $G_{\mathfrak{p}} = g'_{\mathfrak{p}} \beta h$. Define

$$V_\beta = V(K) \cap \beta^{-1}\Big(\prod_{\mathfrak{p} \notin M_\infty} V(\mathcal{O}_{\mathfrak{p}}) \Big)$$
$$G_\beta = G(K) \cap \beta^{-1} U.$$

Then the groups V_β and G_β are commensurable with $V(\mathcal{O})$ and $G(\mathcal{O})$. Since $\beta h v \in V(\mathcal{O}_{\mathfrak{p}})$ for all $\mathfrak{p} \notin M_\infty$, we have $hv \in V_\beta$.

For any subgroup $G_0 \subset G(K)$ and any G_0-invariant subset $V_0 \subset V(K)$, $X >$), let $N(V_0, G_0, X)$ be the number of irreducible G_0-orbits in V_0 of height $\leqslant X$, where each $G_0 v$ is weighted by

$$\frac{1}{\#\,\mathrm{Stab}_{G_0}(v)}.$$

Let $m : V(K) \to [0,1]$ be a G_0-invariant map defined by some congruence conditions (i.e. $m = \prod_{\mathfrak{p}} m_{\mathfrak{p}}$). Let $N_m(V_0, G_0, X)$ be defined as $N(V_0, G_0, X)$, but weighted by

$$\frac{m(v)}{\#\,\mathrm{Stab}_{G_0}(v)}.$$

Theorem 31.4.2. *Define a weight m' by*

$$m'(v) = \chi_{V(K)^{\mathrm{ls}}}(v) \frac{1}{\#\operatorname{Stab}_{G(K)}(v)} \left(\sum_{\beta \in \mathrm{Cl}(G)} \sum_{v_\beta \in G_\beta \backslash V_\beta \cap V(K)v} \frac{1}{\#\operatorname{Stab}_{G_\beta}(v_\beta)} \right)^{-1}.$$

Then

$$N(V(K)^{\mathrm{ls}}, G(K), X) = \sum_{\beta \in \mathrm{Cl}(G)} N_{m'}(V_\beta, G_\beta, X).$$

It is nontrivial (but true) that $m' = \prod m'_{\mathfrak{p}}$.

31.5 Count integral orbits soluble at infinity

Define $N_{m'_\infty}(V_\beta, G_\beta, X)$. Put $K_\infty = \prod_{v \in M_\infty} K_v$. Construct $G_\beta \backslash V(K_\infty)_{<X}$. Set

$$R(X) = G(K_\infty) \backslash V(K_\infty)_{<X}$$
$$\mathcal{F}_X = G_\beta \backslash G(K_\infty)$$

Then $\mathcal{F}_\beta R(X) \twoheadrightarrow G_\beta \backslash V(K_\infty)_{<X}$. The fiber over $G_\beta v$ has size

$$\frac{\#\operatorname{Stab}_{G(K_\infty)}(v)}{\#\operatorname{Stab}_{G_\beta}(v)}.$$

We want

$$N_{m_\infty}(V_\beta, G_\beta, X) = \int_{\mathcal{F}_\beta R(X)} \frac{m_\infty(v)}{\#\operatorname{Stab}_{G(K_\infty)}(v)} \, \mathrm{d}v_{\infty,\beta} + \text{error},$$

were we normalize our Haar measure by $v_{\infty,\beta}(V(K_\infty)/V_\beta) = 1$. To continue, we need a version of Davenport's lemma for function fields. It is proved using Poisson summation. A more serious problem is the cusps. Without loss of generality assume $G_\beta - G(\mathcal{O})$ and $V_\beta = V(\mathcal{O})$. Reduction theory (worked out by Springer) tells us what $G(\mathcal{O}) \backslash G(K_\infty)$ looks like for all local fields. There is still a "NAK decomposition."

When $G = \mathrm{PGL}(2)$, we have

$$N = \begin{pmatrix} 1 & \\ * & 1 \end{pmatrix}$$
$$A = \left\{ \begin{pmatrix} t^{-1} & \\ & t \end{pmatrix} : t > \frac{\sqrt{3}}{2} \right\}$$

Just as when $K = \mathbf{Q}$, we cut off the cusps. Restrict the representation V to the torus A: it decomposes as $V = \bigoplus_{\chi \in U_0} \chi$. For example,

$$\mathrm{Sym}^4(2) = \underbrace{\chi_{x^4}}_{t^{-4}} \oplus \underbrace{\chi_{x^3y}}_{t^{-2}} \oplus \underbrace{\chi_{x^2y^2}}_{1} \oplus \underbrace{\chi_{xy^3}}_{t^2} \oplus \underbrace{\chi_{y^4}}_{t^4}.$$

Describe reducibility using subsets of U_0. The cusp set $V(K_\infty)^{\mathrm{cusp}} \subset V(K_\infty)$ is defined by $|v(\chi)| < c_1$ for some $\chi \in U_0$, where c_1 is chosen so that if $v \in V(\mathcal{O})$, $|v(\chi)| < c_1$, then $v(\chi) = 0$.

There is a combinatorial condition on the characters of A that implies

1. The number of irreducible elements of the cusp is small.

2. The volume of the cusp is small.

This condition only depends on the field K through the torus T. For example, if the group is split over $\mathbf{Q}$, the condition does not depend on the field at all. This has been worked out for all the representations we've seen so far.

Finally, we need an estimation of reducibility. This is a purely local computation.

31.6 Solubility at finite primes

We do this through the weight function $m'_{\mathfrak{p}}$. Again, this is purely local.

31.7 Uniformity estimate

This is more-or-less local. The methods work over any global field. It has been worked out for $\mathrm{Sel}_n(E)$ with $n \in \{2, 3, 4, 5\}$, and to count field extensions.

31.8 Compute local integrals

This is (obviously) a local computation.

The final result is:

$$N(V(K)^{\mathrm{ls}}, G(K), X) = \tau(G/K)\mu_\infty(X) \prod_{\mathfrak{p} \notin M_\infty} \mu_{\mathfrak{p}}.$$

Theorem 31.8.1. *When all E/K are ordered by height, then for $n \in \{2, 3, 4, 5\}$, we have*

$$\mathrm{avg}(\#\,\mathrm{Sel}_n\, E) = \sum_{d|n} d.$$

32 The Chabauty method and symmetric powers of curves

32.1 Introduction

Following Poonen, we say a curve is *nice* if it is smooth, projective, and geometrically irreducible.

Question 32.1.1. *Let X be a nice curve over $\mathbf{Q}$ of genus $g \geqslant 2$. Find all degree-d points on X.*

If $x \in X(\overline{\mathbf{Q}})$, we say x has *degree* d if $[\kappa(x) : \mathbf{Q}] \leqslant d$, where $\kappa(x) = \mathscr{O}_{X,x}/\mathfrak{m}_x$ is the residue field at x. We could rephrase the problem as: find

$$\bigcup_{[K:\mathbf{Q}] \leqslant d} X(K).$$

The problem is: the compositum of all degree-$\leqslant d$ extensions of $\mathbf{Q}$ is not a number field, so we can't apply Faltings' theorem to conclude this set is finite.

Throughout, we assume X has an effective divisor of degree d. This is not harmful, because if X had no such divisor, it would have no points of degree d. We also assume X has a rational point $0 \in X(\mathbf{Q})$.

Question 32.1.2. *Let X be a nice curve over $\mathbf{Q}$ of genus $g \geqslant 2$. Find all $\mathbf{Q}$-points on $\operatorname{Sym}^d X = \overbrace{X \times \cdots X}^{d}/S^d$.*

A point on $\operatorname{Sym}^d X$ will be a multiset $\{x_1, \ldots, x_d\}$; this lies in $(\operatorname{Sym}^d X)(\mathbf{Q})$ if and only if the x_i are σ-conjugates, where $\sigma \in \operatorname{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ has order $\leqslant d$. We can apply a generalized theorem of Faltings, proved in [Fal94], to study $(\operatorname{Sym}^d X)(\mathbf{Q})$.

Theorem 32.1.3 (Faltings)**.** *Let A be an abelian variety over $\mathbf{Q}$, $Y \subset A$ a closed subvariety. Then there exists finitely many subvarieties $Y_i \subset Y$ such that each Y_i is the translate of an abelian subvariety of A, and*

$$Y(\mathbf{Q}) = \bigcup Y_i(\mathbf{Q}).$$

There is a map $j : \operatorname{Sym}^d X \to J = \operatorname{Jac} X$ defined by $\{x_1, \ldots, x_d\} \mapsto [x_1 + \cdots + x_d - d(0)]$. The fibers are $\mathbf{P}^n$ for varying n. Applying Faltings' theorem to the image of the map j, we get $(j(\operatorname{Sym}^d X))(\mathbf{Q}) = \bigcup Y_i(\mathbf{Q})$, whence

$$\begin{aligned}(\operatorname{Sym}^d X)(\mathbf{Q}) &= \bigcup_{n_i} \mathbf{P}^{n_i}(\mathbf{Q}) \cup \bigcup j^{-1}(Y_i(\mathbf{Q})) \\ &= \bigcup_{n_i} \mathbf{P}^{n_i}(\mathbf{Q}) \cup \bigcup_{\dim Y_i \geqslant 1} j^{-1}(Y_i(\mathbf{Q})) \cup \bigcup_{\dim Y_i = 0} j^{-1}(Y_i(\mathbf{Q})).\end{aligned}$$

Since $\bigcup \mathbf{P}^{n_1}(\mathbf{Q})$ is definitely infinite and $\bigcup_{\dim Y_i \geqslant 1} j^{-1}(Y_i(\mathbf{Q}))$ could be infinite, we will count the quantity $\bigcup_{\dim Y_0 = 0} j^{-1}(Y_i(\mathbf{Q}))$. The following theorem proved in [HS91] is useful.

Theorem 32.1.4 (Harris-Silverman)**.** *Let X be a nice curve over $\mathbf{C}$. If $\operatorname{Sym}^2 X$ contains an elliptic curve, then X is either hyperelliptic or bielliptic.*

Here, a *bielliptic curve* is a double cover of an elliptic curve. If X is the hyperelliptic $y^2 = f$, we get $\mathbf{P}^1 \subset \operatorname{Sym}^2 X$, and $\{(x, \sqrt{f(x)}), (x, -\sqrt{f(x)}) : x \in \mathbf{Q}\} \subset (\operatorname{Sym}^2 X)(\mathbf{Q})$.

Definition 32.1.5. The *special set* $\mathcal{S}(V)$ of a $\mathbf{Q}$-variety V is the Zariski closure of the union of the images of all nonconstant maps $f : G \to V$, where G ranges over group varieties defined over $\overline{\mathbf{Q}}$.

We will try to count $(\operatorname{Sym}^d X)(\mathbf{Q}) \smallsetminus \mathcal{S}(\operatorname{Sym}^d X)$. When $d = 1$, we have the following theorem proved in [Col85].

Theorem 32.1.6 (Coleman)**.** *Fix $g \geqslant 2$ and a prime p. There is an effectively computable bound $N(g,p)$ such that if X is a nice curve over $\mathbf{Q}$ of genus g with good redution at p and with $g > \operatorname{rk} J(\mathbf{Q})$, then $\#X(\mathbf{Q}) \leqslant N(g,p)$.*

If $p > 2g$, then $\#X(\mathbf{Q}) \leqslant \#X(\mathbf{F}_p) + (2g-2)$. In that case, Stoll improved the bound to $\#X(\mathbf{F}_p) + 2r$. If $p > 2$, Stoll improved it further to $\#X(\mathbf{F}_p) + \lfloor \frac{2r}{p-2} \rfloor$.

When $d \geqslant 2$, some things are known. In 1993, Klassen counted points on $\mathrm{Sym}^d X$ away from a divisor of dimension $d-1$, where X has gonality $> d$. Here, the *gonality* of a curve X is the minimal degree of a map $X \to \mathbf{P}^1$.

Baker-Bhargava-Wetherell explicitely found all points on $(\mathrm{Sym}^2 X)(\mathbf{Q})$ for X hyperelliptic.

In 2009, Siksek removed the gonality hypothesis from Klassen's result.

Theorem 32.1.7 (Park). *Let $d \geqslant 1$, p be a prime, and $g \geqslant 2$. Then there exists an effectively computable bound $N(g,d,p)$ such that for every nice curve X over $\mathbf{Q}$ of genus g with good reduction at p with $\operatorname{rk} J \leqslant g-d$, satisfying (*), then*

$$\#\left(\mathrm{Sym}^d(\mathbf{Q}) \smallsetminus \mathcal{S}(\mathrm{Sym}^d X)\right) \leqslant N(g,d,p).$$

If $\operatorname{rk} J \leqslant 1$, the hypothesis (*) is unnecessary. It is a rather technical hypothesis that we won't go into here.

32.2 Some applications

Proposition 32.2.1 (Park). *We can take $N(3,3,2) = 1539$ for any degree-7 odd hyperelliptic curve X such that $\operatorname{rk} J \leqslant 1$, with good reduction at 2.*

This bound, while not fantastic, is far better than the one from Faltings' theorem.

Proposition 32.2.2 (Park). *A positive proportion of hyperelliptic curves with X genus $g \geqslant 3$ have no points of $\deg \leqslant 2$-points in $\mathrm{Sym}^2 X$ outside of the special set of $\mathrm{Sym}^2 X$.*

For these curves, $(\mathrm{Sym}^2 X)(\mathbf{Q})$ is parameterized by $\mathbf{P}^1$.

32.3 Chabauty's method

For a more in-depth introduction, see section 29. Let $X/\mathbf{Q}$ be a nice curve with rank $r < g$ and good reduction at p. For $\omega \in \Gamma(J_{\mathbf{Q}_p}, \Omega^1)$, there is a unique group homomorphism $\eta_\omega : J(\mathbf{Q}_p) \to \mathbf{Q}_p$ sending x to $\int_0^x \omega$ when this integral is defined. It is known that there exists ω such that $\eta_\omega(J(\mathbf{Q})) = 0$.

On each residue disk U, there is a coordinate system in which

$$\omega|_U = \sum w_i(t_1, \ldots, t_g)\, \mathrm{d}t_i.$$

Restricting to the curve $X|_U$, we get $\omega|_{X\cap U} = w(t)\,\mathrm{d}t$. Then

$$\#\{x \in U : \eta_\omega(x) = 0\} \geqslant \#(X(\mathbf{Q}) \cap U).$$

Consider the restriction $\eta_\omega : (\mathrm{Sym}^d X)(\mathbf{Q}_p) \to \mathbf{Q}_p$, given by

$$\begin{aligned}\{x_1, \dots, x_d\} &\mapsto \int_0^{[x_1+\cdots+x_d-d(0)]} \omega \\ &= \int_0^{[x_1-0]} \omega + \int_0^{[x_2+\cdots+x_d-(d-1)(0)]} \omega \\ &= \int_0^{[x_1-0]} \omega + \cdots + \int_0^{[x_d-0]} \omega \\ &= I(t_1) + \cdots + I(t_d).\end{aligned}$$

We can estimate the zeros of this map in terms of the power series I. We have to assume $r + d \leqslant g$. From this we get d independent power series in $K_{x_i}[\![t_i]\!]$, where $[K_{x_i} : \mathbf{Q}_p] \leqslant d$, vanishing on $(\mathrm{Sym}^d X)(\mathbf{Q})$.

There is a theory of generalized Newton polygons, partially done by Rabinoff and Bernstein.

33 Topological and algebra-geometric methods over function fields II

In section 30, we finished by looking at the equidistribution of the Möbius function. In this lecture, we'll show how the underlying geometry gives some hints as to why this is the case.

33.1 Möbius function and big monodromy

Recall: we found that "average of $\mu(f) = 0$" comes down to $\#Y_n(\mathbf{F}_q) = q^n + o(q^n)$, where Y_n parameterizes pairs (f, y) with y a square root of $\Delta(f)$. The scheme Y_n is a branched double cover of $\mathbf{A}^n = \mathrm{Sym}^n \mathbf{A}^1$, the space of monic integral polynomials of degree n. For this to work, we need Y_n to be irreducible.

On the level of function fields (that is, generic points) the extension corresponding to $Y_n \to \mathbf{A}^n$ is the quadratic extension

$$K = k(a_1, \dots, a_n) \hookrightarrow K(\sqrt{\Delta}).$$

This quadratic extension is given by a map $G_K = \mathrm{Gal}(\bar{K}/K) \to \mathbf{Z}/2$. The extension is a field precisely when $G_K \to \mathbf{Z}/2$ is nontrivial. That is, $G_K \to \mathbf{Z}/2$ surjectivs if and only if $K(\sqrt{\Delta})$ is a fied, if and only if Δ is *not* a square in $K = k(a_1, \dots, a_n)$. The image of $G_K \to \mathbf{Z}/2$ is the monodromy group, so this is a "large monodromy" result.

Conjecture 33.1.1 (Cohen-Lenstra)**.** *Let ℓ be an odd prime and $E_{r,\ell,N}$ be the expected value of*

$$\mathrm{Surj}\left(\mathrm{Cl}\left(\mathbf{Q}(\sqrt{-d})\right), (\mathbf{Z}/\ell)^{\oplus r}\right)$$

for d random in $[N, 2N]$. Then $\lim_{N\to\infty} E_{r,\ell,N} = 1$.

For example, when $r = 1$, we have $\mathrm{Surj}(A, \mathbf{Z}/\ell) = \#A[\ell] - 1$. We will try to give a function-field analogue of the Cohen-Lenstra heuristic.

33.2 Cohen-Lenstra over function fields

The analogue of $\mathbf{Q}(\sqrt{-d})$ for $d \in [N, 2N]$ is $\mathbf{F}_q(t, \sqrt{f})$ for $\deg f = n$. The analogue of $-d < 0$ is to require the extension $\mathbf{F}_q(t)(\sqrt{f})$ to be ramified at infinity. This happens exactly when n is odd. So we're looking at "hyperelliptic curves of odd degree." The analogue of an ideal on a curve is a divisor, and the analogue of the class group is the Jacobian, namely

$$\mathrm{Cl}\left(\mathbf{F}_q(t)(\sqrt{f})\right) \simeq \mathrm{Jac}(C_f)(\mathbf{F}_q).$$

where C_f is the curve $y^2 = f(t)$. For simplicity, write J_f instead of $\mathrm{Jac}(C_f)$. Note that J_f is a g-dimensional abelian variety, and $J_f[\ell](\overline{\mathbf{F}_q}) \simeq (\mathbf{Z}/\ell)^{2g}$. Here $g = (n-1)/2$.

Define a new space $\mathrm{Conf}_1^n(\ell)$, which parameterizes pairs (f, P), where f is squarefree of degree n and $P \in J_f[\ell] \smallsetminus 0$. So $\pi : \mathrm{Conf}_1^n(\ell) \to \mathrm{Conf}^n$ is an étale cover of degree $\ell^{2g} - 1$. Define $E_{q,\ell,r,n}$ to be the expected value of $\#\,\mathrm{Surj}\,(J_f(\mathbf{F}_q), (\mathbf{Z}/\ell)^{\oplus r})$. Cohen-Lenstra would suggest that $\lim_{n\to\infty} E_{q,\ell,r,n} = 1$.

When $r = 1$, this is

$$\begin{aligned}\mathbf{E}_f\left(\#J_f(\mathbf{F}_q)[\ell] - 1\right) &= \mathbf{E}_f\left(\pi^{-1}(f)(\mathbf{F}_q)\right)\\ &= \frac{\#\,\mathrm{Conf}_1^n(\ell)(\mathbf{F}_q))}{\#\,\mathrm{Conf}^n(\mathbf{F}_q)}\\ &= \frac{\#\,\mathrm{Conf}_1^n(\ell)(\mathbf{F}_q)}{q^n - q^{n-1}}.\end{aligned}$$

So Cohen-Lenstra suggests that $\#\,\mathrm{Conf}_1^n(\ell)(\mathbf{F}_q) \sim q^n$ as $n \to \infty$. Note that to get

$$\lim_{q\to\infty} \frac{\#\,\mathrm{Conf}_1^n(\ell)(\mathbf{F}_q)}{\#\,\mathrm{Conf}^n(\mathbf{F}_q)} = 1,$$

we would need $\mathrm{Conf}_1^n(\ell)$ to be irreducible. This is known to be true, and is proven via a monodromy computation.

Once again, let $K = \mathbf{F}_q(a_1, \ldots, a_n)$, the function field of Conf^n over $\mathbf{F}_q$. Let L be the function field of $\mathrm{Conf}_1^n(\ell)$ over $\mathbf{F}_q$. Then L is K adjoined an ℓ-torsion point of J_f, where $f = t^n + a_1 t^{n-1} + \cdots + a_n$.

A curve C_f over K gives rise to a Galois representation

$$\rho : G_K = \mathrm{Gal}(\bar{K}/K) \to \mathrm{GSp}_{2g}(\mathbf{Z}/\ell) = \mathrm{GSp}(J_f[\ell](K^s), e),$$

where e denotes the Weil pairing. Our question is: is L a field? Equivalently, does the monodromy group $\rho(G_K)$ act transitively on $(\mathbf{Z}/\ell)^{2g} \smallsetminus 0$? It's a good exercise to show that $\mathrm{GSp}_n(\mathbf{F}_q)$ acts transitively on $\mathbf{F}_q^n \smallsetminus 0$ for any finite field $\mathbf{F}_q$.

Theorem 33.2.1 (J-K Yu, 1995)**.** *For K as above, $\rho(G_K) = \mathrm{GSp}_{2g}(\mathbf{F}_\ell)$.*

This is a nice "big monodromy theorem." That is, the monodromy group is as large as it can be.

33.3 What happens when $r = 2$

We could define $\mathrm{Conf}_2^n(\ell)$, which parameterizes triples (f, P, Q), where f is above and P, Q are linearly independent points on $J_f[\ell]$. Then

$$\frac{\#\,\mathrm{Conf}_2^n(\ell)(\mathbf{F}_q)}{\#\,\mathrm{Conf}^n(\mathbf{F}_q)} = E_{q,2,\ell,n},$$

because $\#\,\mathrm{Surj}(J_f(\mathbf{F}_q), (\mathbf{Z}/\ell)^2)$ is (by duality for finite abelian groups) the number of injections $(\mathbf{Z}/\ell)^2 \to J_f(\mathbf{F}_q)$. Unfortunately, the space $\mathrm{Conf}_2^n(\ell)$ is not irrereducible. The components of $\mathrm{Conf}_2^n(\ell)$ are naturally identified with the orbits of $\mathrm{Gal}(\bar{K}/K\overline{\mathbf{F}_q})$ on the set of injections $(\mathbf{Z}/\ell)^2 \hookrightarrow (\mathbf{Z}/\ell)^{2g}$. Even when $g = 1$, this action is not transitive. Let $V_0 = \mathbf{F}_\ell^2$ and $\iota : V_0 \hookrightarrow \mathbf{F}_\ell^{2g}$. The Weil pairing ω pulls back to a pairing on V_0. This pairing is preserved by the monodromy group. The quantity $\omega(P, Q)$ is an invariant.

So in fact, $\mathrm{Conf}_2^n(\ell)$ has ℓ components, parameterized by $\langle P, Q\rangle$ via the map $(f, P, Q) \mapsto \langle P, Q\rangle \in \mu_\ell$. How does $G_{\mathbf{F}_q} = \mathrm{Gal}(\overline{\mathbf{F}_q}/\mathbf{F}_q)$ act on the components? It (the Frobenius at q) just multiplies $\langle P, Q\rangle$ by q. So the number of $\mathbf{F}_q$-rational components of $\mathrm{Conf}_2^n(\ell)$ is the number of elements $x \in \mathbf{Z}/\ell$ such that $qx = x$. This depends strongly on q. It is 1 if $q \not\equiv 1 \pmod{\ell}$, but ℓ when $q \equiv 1 \pmod{\ell}$. That is, when $\mu_\ell \cap \mathbf{F}_q = 1$ there is one component, and ℓ components if $\mu_\ell \subset \mathbf{F}_q$.

The following is work of Derek Garton. You can compute the "modified Cohen-Lenstra distribution," whose moments match those given by

$$E_{q,r,\ell,n} = \#\{\mathbf{F}_q\text{-rational components of } \mathrm{Conf}_r^n(\ell)\},$$

a geometrically motivated repair of Cohen-Lenstra in the presence of extra roots of unity.

33.4 Selmer groups

The inspiration here is the beautiful paper [dJ02]. What does the 3-Selmer group of a random $E/\mathbf{F}_q(t)$ look like? If $E : y^2 = f_t(x)$, let $\mathcal{E}$ be the elliptic surface $y^2 = f(t, x)$. It turns out that $\mathrm{Sel}_3(E/\mathbf{F}_q(t))$ "is" $\mathrm{H}^2(\mathcal{E}_{\overline{\mathbf{F}_q}}, \mathbf{Z}/3)(\mathbf{F}_q)$. You can play the same game, expressing the average number of nontrivial elements of $\mathrm{Sel}_p(E)$ as $\frac{\#Y_n(\mathbf{F}_q)}{\#X_n(\mathbf{F}_q)}$, where $Y_n \to X_n$ is a cover corresponding to the action of G_K on $\mathrm{H}^2(\mathcal{E}_{\overline{\mathbf{F}_q}}, \mathbf{Z}/\ell)$. This cohomology group has a canonical symmetric pairing, so "big monodromy" would mean $\rho(G_K)$ is the entire orthogonal group. Given a big monodromy theorem, the number of components would be the number of orbits of $\mathrm{O}_d(\mathbf{Z}/\ell)$ acting on $(\mathbf{Z}/\ell)^d$. It is $\ell + 1$.

It seems like a proof that "in the large q limit," $\mathrm{avg}(\#\,\mathrm{Sel}_\ell) = \ell + 1$.

33.5 From $q \to \infty$ to the fixed q regime

The reason we know that $\#X_n(\mathbf{F}_q) = q^n + o(q^n)$ when X_n irreducible is the Weil bounds. What's really going on is the Grothendieck-Lefschetz trace formula

$$\#X(\mathbf{F}_q) = \sum_{i=0}^{2\dim X} (-1)^i \,\mathrm{tr}\left(\mathrm{fr}_q, \mathrm{H}_c^i(X_{\overline{\mathbf{F}}_q}, \mathbf{Q}_\ell)\right).$$

The Weil bounds tell you that as $q \to \infty$, the $i = 0$ term contributes 100% of this sum. But H^0 is just the vector space generated by connected components.

If we want to let q stay fixed, we have to show that the H^i contribute nothing to start with. The easiest way for this to be true is for the higher H^i to vanish. This fits into the general idea that "families of moduli spaces have vanishing higher cohomology," i.e. cohomological stability results. I have recent joint work with Vanketesh and Westerland on this.

34 Future perspectives

34.1 Future directions / things to think about after the workshop

We now have a general technique to count (nondegenerate, a.k.a. irreducible) orbits having bounded invariants in a representation $G(\mathbf{Z})$ acting on $V(\mathbf{Z})$, for a representation (G, V) defined over $\mathbf{Z}$. Our first goal is thus to find interesting representations! Of particular interest would be representations where the orbits correspond to objects or arithmetic / geometric / topological interest.

In the future, it would be nice to parameterize rings of rank $n > 5$ (i.e. parameterize n points in $\mathbf{P}^{n-2}$). For example, three points in $\mathbf{P}^1$ are parameterized by binary cubic forms. Four points in $\mathbf{P}^2$ are parameterized by pencils of conics, i.e. pairs of ternary quadratic forms, which we know correspond to quartic rings. Five points in $\mathbf{P}^3$ arise as quadruple of 5×5 symmetric matrices via the Pfaffians of their linear combinations. This representation of $\mathrm{SL}(4) \times \mathrm{SL}(5)$ parameterizes quintic rings. In general, we would like to understand n points in $\mathbf{P}^{n-2}$ in terms of forms. Even for $n = 6$, we do not know how to do this (in terms of prehomogeneous vector spaces or something similar).

It would also be nice to parameterize n-Selmer elements of elliptic curves for $n > 5$. Geometrically, we want to parameterize maps of genus one curves C to $\mathbf{P}^{n-1}$ of degree n. For example, maps $C \to \mathbf{P}^1$ of degree 2 ramify at four points, so they correspond to binary quartic forms. Maps $C \to \mathbf{P}^2$ are plane cubics, parameterized by ternary cubic forms. Maps $C \to \mathbf{P}^3$ are the intersection of two quadrics, hence parameterized by pairs $2 \otimes \mathrm{Sym}^2(4)$. Finally, maps $C \to \mathbf{P}^4$ are the intersection of five quadrics, which come from $5 \otimes \bigwedge^2 5$ (quintuples of 5×5 skew-symmetric matrices).

All these constructions were (in some form) known in the 19th century, though not over $\mathbf{Z}$. If we added yet another variable, we would go from points, to genus one curves, to K3 surfaces. See [BHK13] for work along these lines.

We could look for analogous maps for objects involving surfaces or higher-dimensional varieties. Also, we could try to count objects parameterized by non-coregular surfaces, as in [BY13]. Aside from a few "baby steps," essentially nothing is known.

Parameterizing rings of rank $\leqslant 5$ allowed us to parameterize number fields. We could add more structure and try to parameterize fields with special Galois groups. For example, we don't know how to count A_4-quartic, A_5-quintic, or D_5-quintic fields with bounded discriminant. Some work along these lines is in [BS13], in which the action of $\mathrm{SO}(x^2 + xy + y^2)$ on a 2-dimensional space is used to parameterize C_3-cubic

rings. The subspace of pairs of ternary quadratic forms

$$\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & * & * \\ 0 & * & * \end{pmatrix}, *\right)$$

is preserved by a parabolic in $\mathrm{SL}(2) \times \mathrm{SL}(3)$. This group parameterizes D_4-quartic orders.

We could try to parameterize n-Selmer elements in families of elliptic curves with extra structures (e.g. marked rational points). We know how to parameterize 2 rational points, but have no idea how to do $k > 2$ rational points.

Similarly, we could try to parameterize n-Selmer group / set elements of higher genus curves, or k-tuples of elements of the this type with given invariants. (This is needed to get the k-th moments.) For example, $2 \times 2 \times 2$ cubes up to $\mathrm{SL}_2(\mathbf{Z})$-action parameterize pairs of ideal classes in a cubic ring. Kevin Wilson has done some work along these lines.

There are lists of coregular representations. One way of proving new theorems is to look through these lists and try to find interesting geometric interpretations of these representations.

Families of modular forms could also be attacked via orbit-spaces of representations.

Is there a systematic way of constructing "good" representations for a given arithmetic object? Coregular representations always seem to work. There is great work of Jack Thorne [Tho13] in this direction, using Vinberg theory. Also there is upcoming work of Ho / Sam.

Currently, we have no way of showing that a given group action on a unirational variety can't parameterize something interesting. Prehomogeneous vector spaces and coregular spaces have *not* all been classified. The classification in [SK77] is only of irreducible reductive reduced prehomogeneous vector spaces. Even their list is not finite – it contains some infinite families. There are already known interesting (a.k.a. parameterize something interesting) prehomogeneous vector spaces that are non-reductive or non-reduced, for examples. It would be very nice if there were a complete classification of prehomogeneous vector spaces.

It's important to further develop counting techniques. Right now, we have two methods: geometry of number and zeta functions. So far, zeta-function methods have not worked on higher-dimensional representations. Hopefully, there is a way of unifying these (or just using both) to strengthen results. The adelic counting method described in [Poo13] can at least explain some of the cancellations that we saw in section 22.

We could try to connect with other problems in other areas. For example, Miller has maps from families of knots with given invariants into certain orbit spaces. Also, we could to connect counting techniques as in section 31 with topological techniques as in section 33. Chabauty methods as in section 29 and section 32 could also be strengthened. Finally, we should try to develop better heuristics for all of the above, along with arithmetic justifications for the heuristics. Ellenberg and Venjakob have results in this direction for imaginary quadratic fields. Finally, Skinner and Urban's work [SU14] on p-adic L-functions vis-à-vis the Iwasawa main conjecture for GL(2) is also relevant.

34.2 Applications to the Birch and Swinnerton-Dyer conjecture

This is partially an advertisement for the November workshop "Counting arithmetic objects (ranks of elliptic curves)" (http://www.crm.umontreal.ca/act/theme/theme_2014_2_en/counting_e.php) at the Centre de Recherches Mathématiques. Here are some examples of the sort of results that will be covered.

Theorem 34.2.1 (Bhargava, Shankar)**.** *A positive proportion of elliptic curves over* $\mathbf{Q}$ *have rank* 0*.*

Proof. Use Dokchitser-Dokchitser to show existence of curves with odd (resp. even) p-Selmer rank. From $\operatorname{avg}(\#\operatorname{Sel}_5) = 6$, we conclude that the average rank of an elliptic curve is $\leqslant 1.05$. This could be achieved if 95% have rank 1 and 5% have rank 2. This situation cannot happen if p-Selmer rank parities are well-distributed, as they are (often enough to rule out this scenario). This reduces the average rank bound to $\leqslant 0.885$. □

Theorem 34.2.2 (Bhargava, Skinner)**.** *A positive proportion of elliptic curves have rank* 1*.*

Proof. Dokchitser-Dokchitser implies a lot of elliptic curves have 5-Selmer rank 5. We want to show that these actually have rank 1. Use work of Skinner [Ski14] on Heegner points. □

Theorem 34.2.3 (Bhargava, Skinner, Zhang)**.** *At least* $> 66.48\%$ *elliptic curves over* $\mathbf{Q}$ *satisfy the rank part of the BSD conjecture. They also satisfy the p-part of BSD for all* $p \neq 2$*.*

Corollary 34.2.4. *Most elliptic curves over* $\mathbf{Q}$ *have finite Tate-Shafarevich group.*

Proof. This follows from work of Kolyvagin. □

Clearly, the standard assumption "$p \neq 2$" at the start of most papers needs to be removed.

35 Exercises

Those directing the problem sessions were (in alphabetical order): Jordan Ellenberg, Wei Ho, Jennifer Park, Arul Shankar, Frank Thorne, and Jerry Wang.

35.1 The parameterization of cubic, quartic, and quintic rings

35.1.1 Directly from Wood's cubic rings lecture

a) Prove that the inverse maps $R \mapsto \operatorname{Disc}(R)$ and $D \mapsto \mathbf{Z}[\tau]/\left(\tau^2 - D\tau + \frac{D^2-D}{4}\right)$ induce a bijection between the set of quadratic rings (up to isomorphism) and

$$\{D \in \mathbf{Z} : D \equiv 0, 1 \pmod 4\}.$$

b) In the Delone-Faddeev equations

$$\begin{aligned}\omega\theta &= n\\ \omega^2 &= m - b\omega + a\theta\\ \theta^2 &= \ell - d\omega + c\theta\end{aligned}$$

prove that associativity is equivalent to the equations

$$\begin{aligned}n &= -ad\\ \ell &= -bd\\ m &= -ac\end{aligned}$$

c) Wood mentioned that if you write $+b$ and $+d$ in place of $-b$ and $-d$ above, the correspondence comes out slightly wrong. Try it and see what happens.

d) Orders in cubic number fields correspond to irreducible cubic forms $f(x, y)$ and the number field can be recovered as $\mathbf{Q}[\theta]/f(\theta, 1)$. What happens if you substitute $f(1, \theta)$ for $f(\theta, 1)$. (What *must* happen?)

e) For a cubic form f, prove that the functions on its vanishing set V_f determine a cubic ring, which is the same ring obtained by the Delone-Faddeev correspondence. (Describe any special conditions, e.g. $f \neq 0$, which are necessary in your proof.)

35.1.2 Other exercises concerning cubic rings

We give more exercises for cubic than for quartic or quintic rings. Note that most or all of these exercises are interesting for all three parameterizations being discussed. You are *strongly encouraged* to extrapolate problems from one section to another! What is the same, and what is different?

a) A good way to get started is to compute lots of examples of the Delone-Faddeev correspondence. (If you don't do any of the other exercises, you should probably do at least this, and the quartic and quintic analogues!) What binary cubic form f corresponds to the cubic ring $\mathbf{Z}^3$? To $\mathbf{Z}[\sqrt[3]{n}]$? Conversely, what cubic rings corresponds to the cubic form $u^3 - uv^2 + v^3$? To $u(u-v)(u+v)$? To u^3? To 0? Work out these, as well as other examples of your own invention, and compute all of their discriminants.

b) Another good way to get started is to work out the details of the Delone-Faddeev and Davenport-Heilbronn correspondences. The exposition given in [BST13, §2] leaves many small details to be checked by the reader. Pick your favorite lemma or proposition and work out the proof in more detail than given in the paper.

c) The Delone-Faddeev correspondence is very interesting over $\mathbf{F}_p$. Assuming for simplicity that $p \neq 2, 3$, determine all of the cubic rings over $\mathbf{F}_p$ as well as

the $\mathrm{GL}_2(\mathbf{F}_p)$-equivalence classes of cubic forms over $\mathbf{F}_p$. How many equivalence classes are there? On the cubic forms side, how large is each $\mathrm{GL}_2(\mathbf{F}_p)$-equivalence class, and how big is each of the corresponding stabilizer groups? If you reduce an integral binary cubic form modulo p, what is the relationship between the cubic ring over $\mathbf{Z}$ and the cubic ring over $\mathbf{F}_p$?

d) Work out what the Delone-Faddeev correspondence says over $\mathbf{C}$: The fact that $\mathrm{GL}_2(\mathbf{C})$ acts prehomogeneously on binary cubic forms over $\mathbf{C}$ can be restated by saying that all nonsingular binary cubic forms form a single $\mathrm{GL}_2(\mathbf{C})$-orbit, and therefore (by Delone-Faddeev) that there is exactly nondegenerate cubic ring over $\mathbf{C}$ up to isomorphism. Work out this case of the Delone-Faddeev correspondence, describe this cubic ring, and prove its uniqueness directly.

e) Classify the set of those $\mathrm{GL}_2(\mathbf{Z}_p)$-orbits on $V(\mathbf{Z}_p)$ whose discriminants are not divisible by p. What about those whose discriminants are exactly divisible by p or p^2? Which of those extensions are maximal?

35.2 Some cohomological computations for the representation $V = \mathrm{Sym}_2(n) \oplus \mathrm{Sym}_2(n)$ of $G = \mathrm{SL}_n$ and $H = \mathrm{SL}_n / \boldsymbol{\mu}_2$

Suppose G is a reductive group with a representation V over a field k. et $V/\!\!/G = \operatorname{Spec} k[V]^G$ denote the canonicl quotient. Let $f \in (V/\!\!/G)(k)$ be a rational invariant and suppose $G(k^s)$ acts transitively on $V_f(k^s)$ with abelian stabilizers. In Gross' talk, we learned two obstructions for the existence of a rational element $v \in V(k)$ with invariant f. In this worksheet, we will make some computations regarding these obstructions when the representation V is the space $\mathrm{Sym}_2(n) \oplus \mathrm{Sym}_2(n)$ of pairs of symmetric bilinear forms and when the reductive group is either SL_n or $\mathrm{SL}_n / \boldsymbol{\mu}_2$ with the action given by $g \cdot (A, B) = ({}^t gAg, {}^t gBg)$.

35.2.1 Warm up

Consider the conjugation action of $G = \mathrm{GL}(W)$ on $V = \mathrm{End}(W)$. One can obtain invariants by taking the coefficients $c_1, \ldots, c_n$ of the characteristic polynomial.

Exercise. Show that the ring of polynomial invariants $k[V]^G = k[c_1, \cdots, c_n]$ via the following steps (or however you want to).

1. Show that for any $c_1, \ldots, c_n$ in k^s, there exists some $T \in V(k^s)$ with characteristic polynomial

$$\det(x \cdot 1 - T) = x^{2n+1} + c_1 x^{2n} + \cdots + c_{2n+1}.$$

   This shows that there is no relation among the invariants.

2. Show that for any $c_1, \ldots, c_n$ in k^s such that $f(x) = x^{2n+1} + c_1 x^{2n} + \cdots + c_{2n+1}$ has no repeated roots and for any $T, T' \in V_f(k^s)$, there exists some $g \in G(k^s)$ such that $gTg^{-1} = T'$. This shows that there are no other invariants.

Next we consider the conjugation action of the subgroup $H = \mathrm{SL}(W)$ on $V = \mathrm{End}(W)$. Let $T \in V(k)$ be a regular semisimple operator, that is its characteristic polynomial $f(x)$ has no repeated factors. Let $L = k[x]/f(x)$ be the associated k-vector space of dimension n.

Exercise. 1. Show that the stabilizer H_T of T is isomorphic to $(\mathrm{R}_{L/k}\,\mathbf{G}_\mathrm{m})^{\mathrm{N}=1}$, the kernel of the norm map $\mathrm{R}_{L/k}\,\mathbf{G}_\mathrm{m} \to \mathbf{G}_\mathrm{m}$.

2. Show that $\mathrm{H}^1(k, H_T) \simeq k^\times/\,\mathrm{N}(L^\times)$ by taking Galois cohomology of the short exact sequence
$$1 \to H_f \to \mathrm{R}_{L/k}\,\mathbf{G}_\mathrm{m} \to \mathbf{G}_\mathrm{m} \to 1.$$
Note Shapiro's lemma implies that $\mathrm{H}^1(k, \mathrm{R}_{L/K}\,M) = \mathrm{H}^1(L, M)$ for any $\mathrm{Gal}(L^s/L)$-module M.

3. Using the same idea, show that $\mathrm{H}^1\left(k, (\mathrm{R}_{L/k}\,\boldsymbol{\mu}_2)^{\mathrm{N}=1}\right) \simeq (L^\times/2)^{\mathrm{N}=\square}$.

35.2.2 Cohomological obstructions

The reference for his section is [BGW13, §2]. As shown in section 28, the abelian groups G_v for $v \in V_f(k^s)$ descend to a commutative group scheme G_f over k unique up to unique isomorphism. In other words, there are canonical isomorphisms $\iota_v : G_f(k^s) \xrightarrow{\sim} G_v(k^s)$ for any $v \in V_f(k^s)$ such that for any $h \in G(k^s)$, $b \in G_f(k^s)$, $\sigma \in \mathrm{Gal}(k^s/k)$,

$$\iota_{hv}(b) = h\iota_v(b)h^{-1}$$
$$^\sigma\iota_v(b) = \iota_{^\sigma v}(^\sigma b).$$

For any $v \in V_f(k^s)$ and any $\sigma \in \mathrm{Gal}(k^s/k)$, choose g_σ such that $g_\sigma\,{}^\sigma v = v$. Define

$$d_{\sigma,\tau} = \iota_v^{-1}(g_\sigma g_\tau g_{\sigma\tau}^{-1}) \in G_f(k^s).$$

Exercise. 1. Show that $(d_{\sigma,\tau})$ is a 2-cocycle whose image d_f in $\mathrm{H}^2(k, G_f)$ does not depend on the choice of g_σ. That is, show that for any $\sigma, \tau, \mu \in \mathrm{Gal}(k^s/k)$,

$$^\sigma d_{\tau,\mu} d_{\sigma,\tau\mu} = d_{\sigma\tau,\mu} d_{\sigma,\tau}.$$

and that if each g_σ is changed to $h_\sigma g_\sigma$ for some h_σ in G_v, then

$$d'_{\sigma,\tau} = d_{\sigma,\tau}\iota_v^{-1}(h_\sigma)(^\sigma\iota_v{}^{-1}(h_\tau))(\iota_v^{-1}(h_{\sigma\tau})^{-1}.$$

2. Show that the 2-cochain $(d_{\sigma,\tau})$ does not depend on the choice of $v \in V_f(k^s)$.

Given a class $c \in \mathrm{H}^1(k, G)$, one can obtain a pure inner form G^c of G and a representation V^c of G^c as follows. Suppose c is given by the 1-cocycle (c_σ) with values in $G(k^s)$. Then $G^c(k^s) = G(k^s)$ with action

$$\sigma(h) = c_\sigma\,{}^\sigma h c_\sigma^{-1}. \tag{1}$$

If we compose the cocycle c with values in $G(k^s)$ with the homomorphism $\rho : G \to \mathrm{GL}(V)$, we obtain a cocycle $\rho(c)$ with values in $\mathrm{GL}(V)(k^s)$. By the generalization

of Hilbert's Theorem 90, we have $\mathrm{H}^1(k, \mathrm{GL}(V)) = 1$. Hence there is an element g in $\mathrm{GL}(V)(k^s)$, well-defined up to left multiplication by $\mathrm{GL}(V)(k)$, such that $\rho(c_\sigma) = g^{-1}\,{}^\sigma g$ for all σ in $\mathrm{Gal}(k^s/k)$. We use the element g to define a twisted representation of the group G^c on the vector space V over k. The homomorphism

$$\rho_g : G^c(k^s) \to \mathrm{GL}(V)(k^s)$$

defined by $\rho_g(h) = g\rho(h)g^{-1}$ commutes with the respective Galois actions, so defines a representation over k. We emphasize that the Galois action on $G^c(k^s)$ is defined as in (1), whereas the Galois action on $\mathrm{GL}(V)(k^s)$ is the usual action. We write V^c for the representation ρ_g of G^c. For any $f \in (V/\!\!/G)(k^s)$, we write

$$V_f^c(k) = (k) \cap gV_f(k^s).$$

Exercise. 1. Show that the isomorphism class of G^c does not depend on the choice of the representative (c_τ).

2. Show that the isomorphism class of the representation V^c of G^c does not depend on the choice of the element g.

3. Observe that the group scheme G_f over k does not depend on the twist $c \in \mathrm{H}^1(k, G)$. Show that the class $d_f \in \mathrm{H}^2(k, G_f)$ does not depend on the twist $c \in \mathrm{H}^1(k, G)$.

Theorem. *Let G be a reductive group with representation V. Suppose there exists $v \in V(k)$ with invariant $f \in (V/\!\!/G)(k)$ and stabilizer G_v such that $G(k^s)$ acts transitively on $V_f(k^s)$. Then there is a bijection between the set of $G^c(k)$-orbits on $V_f^c(k)$ and the fiber $\gamma^{-1}(c)$ of the map*

$$\gamma : \mathrm{H}^1(k, G_v) \to \mathrm{H}^1(k, G)$$

above the class $c \in \mathrm{H}^1(k, G)$. In particular, the image of $\mathrm{H}^1(k, G_v)$ in $\mathrm{H}^1(k, G)$ determines the set of pure inner forms of G for which the k-rational invariant f lifts to a k-rational orbit of G^c on V^c.

Exercise. Prove the theorem via the following steps (or however you want to).

1. Fix some $c \in \mathrm{H}^1(k, G)$. Show that $V_f^c(k)$ is nonempty if and only if c is in the image of γ.

2. Suppose $V_f^c(k)$ is nonempty and take $w \in V_f^c(k)$. Then there is a bijection between $G^c(k)\backslash V_f^c(k)$ and $\ker(\gamma_c)$, where γ_c is the natural map of sets $\mathrm{H}^1(G_w^c) \to \mathrm{H}^1(k, G^c)$. Show that there is a bijection between $\gamma^{-1}(c)$ and $\ker(\gamma_c)$.

Theorem. *Suppose that f is a rational invariant, and that $G(k^s)$ acts transitively on $V_f(k^s)$ with abelian stabilizers. Then $d_f = 0$ in $\mathrm{H}^2(k, G_f)$ if and only if there exists a pure inner form G^c of G such that $V_f^c(k)$ is nonempty. That is, the condition $d_f = 0$ is necessary and sufficient for the existence of rational orbits for some pure inner twist of G. In particular, when $\mathrm{H}^1(k, G) = 1$, the condition $d_f = 0$ in $\mathrm{H}^2(k, G_f)$ is necessary and sufficient for the existence of rational orbits of $G(k)$ on $V_f(k)$.*

Exercise. $\Leftarrow$ is trivial. Prove $\Rightarrow$ via the following steps (or however you want to).

1. Show that there exists a 1-cochain (e_σ) with values in $G_v(k^s)$ such that $c = (e_\sigma g_\sigma)$ is a 1-cocycle.

2. Show that $V_f^c(k)$ is nonempty.

35.2.3 The representation $V = \mathrm{Sym}_2(n) \oplus \mathrm{Sym}_2(n)$ of $G = \mathrm{SL}_n$ and $H = \mathrm{SL}_n / \boldsymbol{\mu}_2$

The ring $k[V]^G$ of polynomial invariants is freely generated by the coefficients of the invariant binary form $f(x,y) = (-1)^{n(n-1)/2} \det(Ax - By)$ for $(A,B) \in V$. Fix some binary n-ic form $f(x,y) = f_0 x^n + \cdots + f_n y^n$ with coefficients in k and suppose that f_0 and $\Delta(f)$ are nonzero. Let L denote the étale extension $k[x]/f(x,1)$. Then $G(k^s)$ acts transitively on $V_f(k^s)$ with abelian stabilizers. Moreover, we have

$$\begin{aligned} G_f &\simeq (\mathrm{R}_{L/k}\, \boldsymbol{\mu}_2)^{\mathrm{N}=1}, \\ H_f &\simeq (\mathrm{R}_{L/k}\, \boldsymbol{\mu}_2)^{\mathrm{N}=1} / \boldsymbol{\mu}_2. \end{aligned}$$

The groups G_f and H_f fit inside short exact sequences

$$\begin{aligned} 1 \to (\mathrm{R}_{L/k}\, \boldsymbol{\mu}_2)^{\mathrm{N}=1} \to \mathrm{R}_{L/k}\, \boldsymbol{\mu}_2 \xrightarrow{\mathrm{N}} \boldsymbol{\mu}_2 \to 1 \\ 1 \to \boldsymbol{\mu}_2 \to (\mathrm{R}_{L/k}\, \boldsymbol{\mu}_2)^{\mathrm{N}=1} \to (\mathrm{R}_{L/k}\, \boldsymbol{\mu}_2)^{\mathrm{N}=1} / \boldsymbol{\mu}_2 \to 1. \end{aligned}$$

Taking Galois cohomology gives long exact sequences

$$\begin{aligned} L^\times/2 \xrightarrow{\mathrm{N}} k^\times/2 \xrightarrow{\delta_0} \mathrm{H}^2(k, G_f), \\ \mathrm{H}^1(k, H_f) \xrightarrow{\delta} \mathrm{H}^2(k, \boldsymbol{\mu}_2) \xrightarrow{\alpha} h^2(k, G_f). \end{aligned}$$

Let d_f^G (resp. d_f^H) denote the corresponding classes in $\mathrm{H}^2(k, G_f)$ (resp. $\mathrm{H}^2(k, H_F)$) that obstruct the existence of a rational lift for some pure inner form of G (resp. H). For the following exercise, see [BGW13, §4.5].

Exercise. 1. Show that $d_f^G = \delta_0(f_0)$. Since $\mathrm{H}^1(k, G) = 1$, we see that $V_f(k)$ is nonempty if and only if $f_0 \in (k^\times)^2\, \mathrm{N}(L^2)$.

2. Show that d_f^H is the image of d_f^G under the natural map $\mathrm{H}^2(k, G_f) \to \mathrm{H}^2(k, H_f)$.

3. Suppose now that $d_f^H = 0$. We would like to know for which pure inner forms of H do there exist rational orbits with invariant f.

   (a) Show that $d_f^G \in \mathrm{H}^2(k, G_f)$ is the image of some $d \in \mathrm{H}^2(k, \boldsymbol{\mu}_2)$ under α where d lies in the image of the map $\delta_2 : \mathrm{H}^1(k, H) \hookrightarrow \mathrm{H}^2(k, \boldsymbol{\mu}_2)$ obtained from the short exact sequence $1 \to \boldsymbol{\mu}_2 \to G \to H \to 1$.

   (b) Let c be an element of $\mathrm{H}^1(k, H)$ such that $\delta_2(c) = d$. Show that $v_f^c(k)$ is nonempty.

   (c) Show that the pure inner forms of H for which rational orbits exist with invariant f correspond to classes $c \in \mathrm{H}^1(k, H)$ such that $\alpha(\delta_2(c)) = d_f^G$.

References

[BE77] David A. Buchsbaum and David Eisenbud. Algebra structures for finite free resolutions, and some structure theorems for ideals of codimension 3. *Amer. J. Math.*, 99(3):447–485, 1977.

[BG09] Manjul Bhargava and Eknath Ghate. On the average number of octahedral newforms of prime level. *Math. Ann.*, 344(4):749–768, 2009.

[BG13] Manjul Bhargava and Benedict Gross. The average size of the 2-Selmer group of Jacobians of hyperelliptic curves having a rational Weierstrass point. In *Automorphic representations and L-functions*, volume 22 of *Tata Inst. Fundam. Res. Stud. Math.*, pages 23–91. 2013.

[BGW13] Manjul Bhargava, Benedict Gross, and Xiaoheng Wang. Arithmetic invariant theory II. 2013. Preprint, `arXiv:1310.769`.

[Bha01] Manjul Bhargava. *Higher composition laws.* ProQuest LCC, Ann Arbor, MI, 2001. Thesis (Ph.D.) – Princeton University.

[Bha04a] Manjul Bhargava. Higher composition laws. I. A new view on Gauss composition, and quadratic generalizations. *Ann. of Math. (2)*, 159(1):217–250, 2004.

[Bha04b] Manjul Bhargava. Higher composition laws. II. On cubic analogues of Gauss composition. *Ann. of Math. (2)*, (2):865–886, 2004.

[Bha04c] Manjul Bhargava. Higher composition laws. III. The parametrization of quartic rings. *Ann. of Math. (2)*, 159(3), 2004.

[Bha05] Manjul Bhargava. The density of discriminants of quartic rings and fields. *Ann. of Math. (2)*, 162(2):1031–1063, 2005.

[Bha07] Manjul Bhargava. Mass formulae for extensions of local fields, and conjectures on the density of number field discriminants. *Int. Math. Res. Not. IMRN*, (17), 2007.

[Bha08] Manjul Bhargava. Higher composition laws. IV. The parameterization of quintic rings. *Ann. of Math. (2)*, 167(1):53–94, 2008.

[BHC62] Armand Borel and Harish-Chandra. Arithmetic subgroups of algebraic groups. *Ann. of Math. (2)*, 75:485–535, 1962.

[BHK13] Manjul Bhargava, Wei Ho, and Abhinav Kumar. Orbit parameterizations for K3 surfaces. 2013. Preprint, `arXiv:1312.0898`.

[BKL+13] Manjul Bhargava, Daniel Kane, Hendrik Lenstra, Bjorn Poonen, and Eric Rains. Modeling the distribution of ranks, selmer groups, and shafarevich-tate groups of elliptic curves. 2013. Preprint, `arXiv:1304.3971v2`.

[BM72] B. J. Birch and J. R. Merriman. Finiteness theorems for binary forms wth given discriminant. *Proc. London Math. Soc. (3)*, 24:385–394, 1972.

[Bor63] Armand Borel. Some finiteness properties of adele groups over number fields. *Inst. Hautes Études Sci. Publ. Math.*, (16):5–30, 1963.

[Bor91] Armand Borel. *Linear algebraic groups*, volume 126 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1991.

[Bru92] Armand Brumer. The average rank of elliptic curves. I. *Invent. Math.*, 109(3):445–472, 1992.

[BS13] Manjul Bhargava and Ariel Shnidman. On the number of cubic orders of bounded discriminant having automorphism group C_3, and related problems. 2013. Preprint, `arXiv:1206.4746`.

[BST13] Manjul Bhargava, Arul Shankar, and Jacob Tsimerman. On the Davenport-Heilbronn theorems and second order terms. *Invent. Math.*, 193(2):439–499, 2013.

[BY13] Manjul Bhargava and Andrew Yang. On the number of binary n-ic forms having bounded julia invariant. 2013. Preprint, `arXiv:1312.7339`.

[CE96] G. Casnati and T. Ekedahl. Covers of algebraic varieties. I. A general structure theorem, covers of degree 3, 4, and Enriques surfaces. *J. Algebraic Geom.*, 5(3):439–460, 1996.

[CEVV09] Dustin Cartwright, Daniel Erman, Mauricio Velasco, and Bianca Viray. Hilbert schemes of 8 points. *Algebra Number Theory*, 3(7):763–795, 2009.

[Cha54] B. Charles. Sur l'algèbre des opérateurs linéaires. *J. Maths. Pures. Appl.*, 33:81–145, 1954.

[Col85] Robert Coleman. Effective Chabauty. *Duke Math. J.*, 52(3):765–770, 1985.

[Con12] Brian Conrad. Finiteness theorems for algebraic groups over function fields. *Compos. Math.*, 148(2):555–639, 2012.

[Dav51] H. Davenport. On a principle of Lipschitz. *J. London Math. Soc.*, 26:179–183, 1951.

[Dav64] H. Davenport. Corrigendum: "On a principle of Lipschitz". *J. London Math. Soc.*, 39:580, 1964.

[DC70] Jean Dieudonné and James Carrell. Invariant theory, old and new. *Advances in Math.*, 4:1–80, 1970.

[Del01] Christophe Delaunay. Heuristics on Tate-Safarevitch groups of elliptic curves defined over $\mathbf{Q}$. *Experiment. Math.*, 10(2):191–196, 2001.

[DF64] B. Delone and D Faddeev. *The theory of irrationalities of the third degree*, volume 10 of *Translations of Mathematical Monographs*. American Mathematical Society, 1964.

[DH69] H. Davenport and H. Heilbronn. On the density of discriminants of cubic fields. *Bull. London Math. Soc.*, 1:345–348, 1969.

[dJ02] A. J. de Jong. Counting elliptic curves over finite fields. *Mosc. Math. J.*, (2):281–311, 2002.

[DS74] Pierre Deligne and Jean-Pierre Serre. Formes modulaires de poids 1. *Ann. Sci. École Norm. Sup. (4)*, 7:507–530, 1974.

[Dym66] Z. Dyment. Maximal commutative nilpotent subalgebras of a matrix algebra of the sixth degree. *Vesci Akad. Navuk BSSR Ser. Firz.-Mat. Navuk*, (3):55–68, 1966.

[EK95] Alex Eskin and Yonatan Katznelson. Singular symmetric matrices. *Duke Math. J.*, 79(2):515–547, 1995.

[Elk91] Noam Elkies. *ABC* implies Mordell. *Internat. Math. Res. Notices*, (7):99–109, 1991.

[Fal94] Gerd Faltings. The general case of S. Lang's conjecture. In *Barsotti Symposium in Algebraic Geometry (Abano Terme, 1991)*, volume 15 of *Perspect. Math.*, pages 175–182. Academic Press, 1994.

[FH91] William Fulton and Joe Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, 1991.

[GGS02] Wee Teck Gan, Benedict Gross, and Gordan Savin. Fourier coefficients of modular forms on G_2. *Duke Math. J.*, 115(1):105–169, 2002.

[GW09] Roe Goodman and Nolan Wallach. *Symmetry, representations, and invariants*, volume 255 of *Graduate Texts in Mathematics*. Springer, 2009.

[HB94] D. R. Heath-Brown. The size of Selmer groups for the congruent number problem. II. *Invent. Math.*, 118(2):331–370, 1994. With an appendix by P. Monsky.

[HB04] D. R. Heath-Brown. The average analytic rank of elliptic curves. *Duke Math. J.*, 122(3):591–623, 2004.

[HS91] Joe Harris and Joe Silverman. Bielliptic curves and symmetric products. *Proc. Amer. Math. Soc.*, 112(2):347–356, 1991.

[Hum75] James Humphreys. *Linear algebraic groups*, volume 21 of *Graduate Texts in Mathematics*. Springer-Verlag, 1975.

[Hum78] James Humphreys. *Introduction to Lie algebras and representation theory*, volume 9 of *Graduate Texts in Mathematics*. Springer-Verlag, 1978. Second printing, revised.

[KS99] Nicholas Katz and Peter Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45. American Mathematical Society, Providence, RI, 1999.

[KW09a] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. I. *Invent. Math.*, 178(3):485–504, 2009.

[KW09b] Chandrashekhar Khare and Jean-Pierre Wintenberger. Serre's modularity conjecture. II. *Invent. Math.*, 178(3):505–586, 2009.

[Lan80] Robert Langlands. *Base change for* GL(2), volume 96 of *Annals of Math. Studies*. Princeton Univ. Press, 1980.

[Mal02] Gunter Malle. On the distribution of Galois groups. *J. Number Theory*, 92(2):315–329, 2002.

[Mal04] Gunter Malle. On the distribution of Galois groups. II. *Experiment. Math.*, 13(2):129–135, 2004.

[MFK94] D. Mumford, J. Fogarty, and F. Kirwan. *Geometric invariant theory*, volume 34 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (2)*. Springer-Verlag, third edition, 1994.

[MP12] William McCallum and Bjorn Poonen. The method of Chabauty and Coleman. In *Explicit methods in number theory*, volume 36 of *Panor. Synthèses*, pages 99–117. Soc. Math. France, 2012.

[Nak89] Jin Nakagawa. Binary forms and orders of algebraic number fields. *Invent. Math.*, 97:219–235, 1989.

[Ner87] Yu Neretin. An estimate for the number of parameters defining an n-dimensional algebra. *Izv. Akad. Nauk SSSR Ser. Mat.*, 51(2):306–318, 447, 1987.

[O'N02] Catherine O'Neil. The period-index obstruction for elliptic curves. *J. Number Theory*, 95(2):329–339, 2002.

[Poo08] Bjorn Poonen. The moduli space of commutative algebras of finite rank. *J. Eur. Math. Soc.*, 10:817–836, 2008.

[Poo13] Bjorn Poonen. Average rank of elliptic curves [after Manjul Bhargava and Arul Shankar]. *Astérisque*, (352):187–204, 2013. Sémenaire Bourbaki Vol. 2011/2012. Exposés 1043–1058.

[PR12] Bjorn Poonen and Eric Rains. Random maximal isotropic subspaces and Selmer groups. *J. Amer. Math. Soc.*, 25(1):245–269, 2012.

[PS13] Bjorn Poonen and Michael Stoll. Most odd degree hyperelliptic curves have only one rational point. 2013. To appear in *Annals of Math*, `arXiv:1302.0061`.

[Ser79] Jean-Pierre Serre. *Local Fields*, volume 67 of *Graduate Texts in Mathematics*. Springer-Verlag, 1979. Translated from the French by Marvin Greenberg.

[Shi72] Takuro Shintani. On Dirichlet series whose coefficients are class numbers of integral binary cubic forms. *J. Math. Soc. Japan*, 24:132–188, 1972.

[Shi75] Takuro Shintani. On zeta-functions associated with the vector space of quadratic forms. *J. Fac. Sci. Univ. Tokyo Sect. I A Math.*, 22:25–65, 1975.

[Sim01] Denis Simon. The index of nonmonic polynomials. *Indag. Math. (N.S.)*, 12(4):505–517, 2001.

[Sim03] Denis Simon. La classe invariante d'une forme binaire. *C. R. Math. Acad. Sci. Paris*, 336(1):7–10, 2003.

[Sim05] Denis Simon. Solving quadratic equations using reduced unimodular quadratic forms. *Math. Comp.*, 74(251):1531–1543 (electronic), 2005.

[SK77] M. Sato and T. Kimura. A classification of irreducible prehomogeneous vector spaces and their relative invariants. *Nagoya Math. J.*, 65:1–155, 1977.

[Ski14] Christopher Skinner. A converse to a theorem of Gross, Zagier, and Kolyvagin. 2014. Preprint, `arXiv:1405.7294`.

[Spr09] T. A. Springer. *Linear algebraic groups*. Birkhäuser, Boston, MA, second edition, 2009.

[SS74] Mikio Sato and Takuro Shintani. On zeta functions associated with prehomogeneous vector spaces. *Ann. of Math. (2)*, 100:131–170, 1974.

[SU14] Christopher Skinner and Eric Urban. The Iwasawa main conjectures for GK_2. *Invent. Math.*, 195(1):1–277, 2014.

[Sup56] D. Supruneko. On maximal commutative subalgebras of the full linear algebra. *Uspehi Mat. Nauk (N.S.)*, 11:181–184, 1956.

[Tho13] Jack Thorne. Vinberg's representations and arithmetic invariant theory. 2013. Preprint, http://www.math.harvard.edu/~thorne/canonical.pdf.

[Tun81] Jerrold Tunnell. Artin's conjecture for representations of octahedral type. 5(2):173–175, 1981.

[uh] user76758 (http://mathoverflow.net/users/43107/user76758). Etale cohomology and restricted direct product. MathOverflow. http://mathoverflow.net/q/161783.

[Woo09] Melanie Matchett Wood. *Moduli spaces for rings and ideals*. ProQuest LCC, Ann Arbor, MI, 2009. Thesis (Ph.D.) – Princeton University.

[Woo11a] Melanie Matchett Wood. Gauss composition over an arbitrary base. *Adv. Math.*, 226(2):1756–1771, 2011.

[Woo11b] Melanie Matchett Wood. Parameterizing quartic algebras over an arbitrary base. *Algebra Number Theory*, 5(8):1069–1094, 2011.

[Woo11c] Melanie Matchett Wood. Rings and ideals parameterized by binary n-ic forms. *J. Lond. Math. Soc. (2)*, 83(1):208–231, 2011.

[Woo12a] Melanie Matchett Wood. The distribution of the number of points on trigonal curves over $\mathbf{F}_q$. *Int. Math. Res. Not. IMRN*, (23):5444–5456, 2012.

[Woo12b] Melanie Matchett Wood. Quartic rings associated to binary quartic forms. *Int. Math. Res. Not. IMRN*, (6):1300–1320, 2012.

[Woo14] Melanie Matchett Wood. Parameterization of ideal classes in rings associated to binary forms. *J. Reine Angew. Math.*, 689:169–199, 2014.

[WY92] David Write and Akihiko Yukie. *Prehomogeneous vector spaces and field extensions*, volume 110. 1992.

[You06] Matthew Young. Low-lying zeros of families of elliptic curves. *J. Amer. Math. Soc.*, 19(1):205–250, 2006.

[Yuk92] Akihiko Yukie. On the Shintani zeta function for the space of binary quadratic forms. *Math. Ann.*, 292(2):355–374, 1992.