Sums of squares, sums of cubes, and modern number theory

Kimball Martin*

Original version: October 14, 2015
Minor revisions: March 12, 2019

Abstract

These are notes which grew out of a talk for general math graduate students with the aim of starting from the questions "Which numbers are sums of two squares?" and "Which numbers are sums of two cubes?" and going on a tour of many central topics in modern number theory. In the notes, I discuss composition laws, class groups, L-functions, modular forms, and elliptic curves, ending with the Birch and Swinnerton-Dyer conjecture. The goal is not to explain any topic too deeply, but to provide some context for how these seemingly disparate topics piece together to (attempt to) satisfy the burning questions of classical number theory.

Contents

*Department of Mathematics, University of Oklahoma, Norman, OK 73019

Introduction

Number theory is about solving diophantine equations, usually in $\mathbb{Z}$ or in $\mathbb{Q}$. These are equations of the form

$$P(x_1, \ldots, x_k) = n, \tag{1}$$

where P is a polynomial with integer coefficients, and $n \in \mathbb{Z}$. Given such an equation, there are a couple of basic questions we can ask:

Question 1. *Does* (1) *have a solution? (over* $\mathbb{Z}$ *or over* $\mathbb{Q}$*)*

Question 2. *How many solutions does* (1) *have? (over* $\mathbb{Z}$ *or over* $\mathbb{Q}$*)*

Note the second question is a refinement of the first—the first question is just asking whether the answer to the second question is nonzero or not, and occasionally it turns out that the easiest way of answering the first question is by trying to answer the second.

We will focus on several explicit examples, such as $P(x, y) = x^2 + y^2$, $P(x_1, \ldots, x_k) = x_1^2 + \cdots + x_k^2$ and $P(x, y) = x^3 + y^3$. The first two, each term having degree 2, are called quadratic forms, and the latter example is called a cubic form. For these polynomials, the first question just reads: what numbers are sums of two squares? what numbers are sums of r squares? and what numbers are sums of two cubes?

Besides their aesthetic appeal, exploring these questions will lead us on a safari adventure where, if you keep your eyes open, you'll get a glimpse of many fascinating ideas in mathematics. During this tour, you can spy things like class groups, L-functions, modular forms and elliptic curves in their natural habitat, and gain an appreciation for how they interact and coexist within the world of number theory.

As mentioned in the abstract, these notes are based on a talk aimed at first and second year math grad students at the University of Oklahoma in October 2015. Of course, there are many more details here than what I could fit into a one-hour talk.[1] (The talk covered sums of two squares, brief remarks on more general binary quadratic forms, and then focused on sums of two cubes.) I hope that these notes may be of interest both to students without prior exposure to number theory, as well as those currently learning number theory, as number theory is big world with many different roads leading into it. (Since this was based on a talk for grad students, I assume some familiarity with abstract algebra, though a large part of the story can be understood without this.) Even many students who have taken a few number theory courses may not know every topic or connection I will mention here, so it may at least prove useful as a cartographical assistant. (And the details about sums of two cubes may not be known to many number theorists, including me.)

Feedback is welcome, as I hope to update and expand these notes someday, possibly including brief introductions to topics such as higher composition laws (à la Bhargava) and Siegel modular forms.

[1] I did not expect the number of pages of these notes to reach the number of lines on a cubic surface, but it is not an unpleasing coincidence.

1 Binary quadratic forms

My primary reference in preparing this section was my Number Theory II notes [Marb], which lists additional references.

1.1 Sums of two squares

Let's start with the basic question: which numbers are sums of two squares, i.e., what integers n satisfy

$$x^2 + y^2 = n, \qquad x, y \in \mathbb{Z}. \tag{2}$$

It is clear that we need $n \geq 0$. We do not require both x, y to be nonzero, so any square is a sum of two squares by taking y, say, to be 0. For $n \leq 10$, we see that 0, 1, 2, 4, 5, 8, 9, 10 are all sums of two squares, but 3, 6, 7 are not.

The first key to solving this problem comes from Brahmagupta's composition law (7th c.):

$$(x^2 + my^2)(z^2 + mw^2) = (xz + myw)^2 + m(xw - yz)^2. \tag{3}$$

We only need the $m = 1$ case (which goes back to Diophantus, 4th c.) for $x^2 + y^2$, but it may be interesting to see one gets a similar law for the forms $x^2 + my^2$. This law for $m = 1$ says that if two numbers are sums of two squares, then their product is also. In other words, we can *compose* solutions to $x^2 + y^2 = n_1$ and $x^2 + y^2 = n_2$ to get a solution to $x^2 + y^2 = n_1 n_2$.

Using this, we can reduce the problem to solving (2) when $n = p$ is prime. To explain the reduction in the most basic setting, let's just consider (2) when $n = pq$ is a product of two primes. First, if p and q are both sums of two squares then (3) tells us so is pq. Then we need to understand what happens when either one or both p and q are not sums of two squares. One case is obvious: for $p = q$, we get $pq = p^2$ is a sum of two squares whether p and q are sums of squares or not.

Here is a nice framework to understand this reduction. Having a solution to (2) means we have a factorization,

$$(x + iy)(x - iy) = n, \qquad x, y \in \mathbb{Z},$$

of n into 2 "smaller" numbers in $\mathbb{Z}[i]$. One way to measure the size of a number in $\mathbb{Z}[i]$ comes from undoing our factorization: the *norm* of $\alpha = x + iy \in \mathbb{Z}[i]$ (or more generally, in $\alpha \in \mathbb{Q}(i)$) is

$$N(\alpha) = \alpha\bar{\alpha} = (x + iy)(x - iy) = x^2 + y^2.$$

(Here $\alpha \mapsto \bar{\alpha}$ denotes Galois conjugation in $\mathbb{Q}(i)/\mathbb{Q}$, which happens to coincide with complex conjugation in this case, but it will be different for real quadratic extensions like $\mathbb{Q}(\sqrt{2})/\mathbb{Q}$.) Hence $N(\alpha)$ is the square of the length of the vector from 0 to α in the complex plane. Note we can rephrase (2) having a solution as saying n is a norm from $\mathbb{Z}[i]$. It is easy to check that the norm is multiplicative, which gives us Brahmagupta's composition law (3) when $m = 1$.

Now suppose $n = pq$ where p is not a sum of two squares and q is another prime. Then we have a factorization $n = (x + iy)(x - iy)$ with $x, y \neq 0$. This give us two factorizations of n:

$$n = (x + iy)(x - iy) = pq.$$

One can define a notion of primes in $\mathbb{Z}[i]$, and the important feature for us is that $\mathbb{Z}[i]$ *also has unique factorization into primes.* Since p is not a sum of two squares, it doesn't factor into "smaller" numbers in $\mathbb{Z}[i]$, i.e., p remains prime in $\mathbb{Z}[i]$. However p cannot divide $x \pm iy$ (i.e., $\frac{x \pm iy}{p} \notin \mathbb{Z}[i]$), so the prime factorization of $(x+iy)(x-iy)$ is "not compatible" with the factorization pq, contradicting the uniqueness of prime factorization. That is, pq is not a sum of two squares. More generally, this kind of argument shows that if p is not a sum of two squares, then $n = p^e m$ is not a sum of two squares if e is odd and $\gcd(p, m) = 1$.

Thus solving the two squares problem for $n = p$ will yield the answer for general $n \in \mathbb{N}$, and here is the answer.

Theorem 1.1 (Fermat (1640))**.** *A prime p is a sum of two squares if and only if $p \equiv 1 \bmod 4$ or $p = 2$. More generally, a positive integer n is a sum of two squares if and only if any prime factor of n which is $3 \bmod 4$ occurs to an even power in the prime factorization of n.*

One way to go about proving this is to think about which primes $p \in \mathbb{N}$ factor in $\mathbb{Z}[i]$, or equivalently, whether p is a norm from $\mathbb{Z}[i]$. Some work is involved, and we won't do it here. However, I want to point out that this gives us an example of an important principle, which we'll come back to in our next example. Namely, Fermat's theorem says that $x^2+y^2 = p$ has a solution if and only if there are no local obstructions, i.e., if and only if $x^2+y^2 \equiv p \bmod m$ has a solution for all m. In fact, there's only one important m to check: $m = |\Delta| = 4$, where $\Delta = -4$ is the discriminant (see the next section).

That is, consider $x^2 + y^2 \equiv p \bmod 4$. The only possibilities for $x^2, y^2 \bmod 4$ are 0 and 1, hence we always have $x^2 + y^2 \equiv 0, 1, 2 \bmod 4$. A prime p is never 0 mod 4, and only 2 mod 4 for $p = 2$. We call this kind of priniciple, where "local" (mod m) solvability is equivalent to "global" (over $\mathbb{Z}$ or $\mathbb{Q}$) solvability a *local-global principle.*[2]

Regarding our second question, up to interchanging x and y and multiplying x and y by ± 1, one can show that there is only one way to write $p = x^2 + y^2$ for $p \equiv 1 \bmod 4$.

1.2 Positive definite forms

Having treated x^2+y^2, we can ask about similar quadratic forms in x and y such as x^2+2y^2 or $x^2 - xy + y^2$. Given such a form Q, it is not true that a composition law like the one in (3), nor is it true that a local-global principle holds (i.e., one can not determine numbers, or even just primes, represented by general forms just by taking congruences). Nevertheless, Gauss in his youth discovered a miraculous composition law on collections of such forms. In modern terminology—he defined a group law on appropriate collections of quadratic forms. Moreover, one has a local-global principle for these groups of forms.

A *binary quadratic form* (over $\mathbb{Z}$) is a polynomial of the form $Q(x, y) = ax^2 + bxy + cy^2$, where $a, b, c \in \mathbb{Z}$ and not all of a, b, c are 0. An important invariant is the *discriminant* is $\Delta = b^2 - 4ac$.

Let $\Delta < 0$ and $\mathcal{F}(\Delta)$ be the set of *positive definite* binary quadratic forms of discriminant Δ. Positive definite means forms which only take on positive (or zero) values, so we exclude

[2]Note this local-global principle does not apply to the general equation $x^2 + y^2 \equiv n \bmod 4$. E.g., take $n = 3 \cdot 7$. Since 3 and 7 are both 3 mod 4, they are not sums of two squares, so neither is their product. However $3 \cdot 7 \equiv 1 \bmod 4$, which is a sum of two squares mod 4.

negative definite forms like $-x^2 - y^2$ (which only take on zero or negative values). The condition $\Delta < 0$ implies that our form is either positive or negative definite, and not something which takes on both positive and negative values like xy or $x^2 - y^2$, which are called *indefinite* forms.[3] (The theory of negative definite forms will follow from the that of positive definite forms.) Given one such form $Q(x, y)$ and $\tau = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ (the group of integral 2×2 matrices with determinant 1), we can define

$$Q^\tau(x, y) = Q(x', y'), \quad \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

In other words, Q^τ is just obtained from Q by an invertible linear change of variables. It will also be a positive definite binary quadratic form of discriminant Δ, and an integer n will be represented by Q^τ (i.e., $Q^\tau(x, y) = n$ has a solution) if and only if it is represented by Q.

Thus the action of $\mathrm{SL}_2(\mathbb{Z})$ defines an equivalence relation on $\mathcal{F}(\Delta)$, called *proper equivalence*. Define the set $\mathrm{Cl}(\Delta)$ of *form classes* to be the set of proper equivalence classes of $\mathcal{F}(\Delta)$. Any $Q \in \mathcal{F}(\Delta)$ is properly equivalent to exactly one *reduced* form $ax^2 + bxy + cy^2$, i.e., a form with $|b| \leq a \leq c$. It is easy to see the number of reduced forms of a fixed discriminant must be finite, whence $\mathrm{Cl}(\Delta)$ is finite.

Theorem 1.2 (Gauss (1798?))**.** $\mathrm{Cl}(\Delta)$ *is a finite abelian group, called the form class group of discriminant* Δ*.*

The order $h(\Delta)$ of $\mathrm{Cl}(\Delta)$ is called the *class number*.

Here is a classical way to define Gauss composition. Suppose $Q_1(x, y) = a_1x^2 + b_1xy + c_1y^2$ and $Q_2(x, y) = a_2x^2 + b_2xy + c_2y^2$ have discriminant Δ and satisfy $b_1 = b_2 = b$ for some b, $a_1|c_2$ and $a_2|c_1$. (The notation $a|c$ means a divides c.) Then $\frac{c_2}{a_1} = \frac{c_1}{a_2} = c$ for some c, and we define the composition $Q_1Q_2 = Q_3$ to be the form $Q_3(x, y) = a_1a_2x^2 + bxy + cy^2$. Then the identity

$$(a_1x_1^2 + bx_1y_1 + c_1y_1^2)(a_1x_2^2 + bx_2y_2 + c_2y_2^2) = a_1a_2x^2 + bxy + cy^2,$$

where $x = x_1x_2 - cy_1y_2$ and $y = a_1x_1y_2 + a_2y_1x_2 + by_1y_2$ tells us this is a composition law in Brahmagupta's sense, i.e., if Q_1 represents n_1 and Q_2 represents n_2, then one can compose the solutions to get that Q_3 represents n_1n_2. Then one can show that given any proper equivalence classes in $\mathrm{Cl}(\Delta)$, one can choose Q_1 and Q_2 as above, and the composition respects proper equivalence classes.

A more insightful way of understanding Gauss composition is in terms of ideal classes. Suppose Δ is the discriminant of the quadratic field $K = \mathbb{Q}(\sqrt{\Delta})$, then one can define a correspondence between $\mathcal{F}(\Delta)$ and ideals of the ring of integers $\mathcal{O}_K$ of K which induces an isomorphism of $\mathrm{Cl}(\Delta)$ with the ideal class group $\mathrm{Cl}(\mathcal{O}_K)$.[4] I will describe this correspondence in our example below.

[3] The most famous indefinite binary quadratic form is $x^2 - dy^2$, $d > 0$, which is appears in *Pell's equation* $x^2 - dy^2 = \pm 1$. Solutions to Pell's equation are interesting because they provide good rational approximations to $\sqrt{d}$ for large y, since $d = \frac{x^2}{y^2} \mp \frac{1}{y^2}$.

[4] This also works if Δ is not a fundamental discriminant, i.e., not the discriminant of some $\mathcal{O}_K$, by replacing the ring of integers in $K = Q(\sqrt{\Delta})$ with a quadratic order $\mathcal{O}_\Delta$ of discriminant Δ, which will be a subring of $\mathcal{O}_K$.

Example 1.3. *When $\Delta = -4$, we can compute there is only reduced form: $x^2 + y^2$, so $h(-4) = 1$ and $\mathrm{Cl}(-4)$ just consists of the (class of) $x^2 + y^2$.*

Example 1.4. *When $\Delta = -8$, there is only one reduced form $x^2 + 2y^2$.*

Example 1.5. *When $\Delta = -3$, again there is only one reduced form $x^2 - xy + y^2$.*

Example 1.6. *When $\Delta = -20$, there are two reduced forms, $Q_1(x, y) = x^2 + 5y^2$ and $Q_2(x, y) = 2x^2 + 2xy + 3y^2$, and the composition is such that $Q_2^2 = Q_1$. (Note, we know $Q_1^2 = Q_1$ by (3).)*

Example 1.7. *When $\Delta = -23$, we get $h(-23) = 3$ and the reduced forms are $Q_1(x, y) = x^2 + xy + 6y^2$, $Q_2(x, y) = 2x^2 - xy + 3y^2$, and $Q_3(x, y) = 2x^2 + xy + 3y^2$. So $\mathrm{Cl}(-23)$ is a cyclic group of order 3, generated by Q_2 or Q_3 and Q_1 is the identity.*

For the rest of this section, let me discuss the case of $\Delta = -23$, as I think it is more interesting than the $\Delta = -20$ example (or the class number 1 examples), though the general ideas apply to will arbitrary negative discriminants Δ. So take $\Delta = -23$ and Q_1, Q_2, Q_3 as in Example 1.7.

Here the associated imaginary quadratic field is $K = \mathbb{Q}(\sqrt{\Delta}) = \mathbb{Q}(\sqrt{-23})$. The right analogue of the integers $\mathbb{Z}$ is the *ring of integers* $\mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{-23}}{2}]$. We have a norm $N : K \to \mathbb{Q}$ given by

$$N(a + b\sqrt{-23}) = a^2 + 23b^2, \qquad a, b \in \mathbb{Q}$$

It's easy to check that for $\alpha \in \mathcal{O}_K$, $N(\alpha) \in \mathbb{Z}$. It turns out $\mathcal{O}_K$ is the largest ring in K for which this is true, and is one reason why $\mathcal{O}_K$ is nicer to work with than the subring $\mathbb{Z}[\sqrt{-23}]$.[5] We say two nonzero ideals $\mathcal{I}, \mathcal{J} \subset \mathcal{O}_K$ are *equivalent*, and write $\mathcal{I} \sim \mathcal{J}$, if $a\mathcal{I} = b\mathcal{J}$ for some $a, b \in \mathcal{O}_K - \{0\}$. Note $\mathcal{I} \sim \mathcal{O}_K$ if and only if $\mathcal{I} = a\mathcal{O}_K$ for some nonzero $a \in \mathcal{O}_K$, i.e., if and only if $\mathcal{I}$ is principal. The product $\mathcal{I}\mathcal{J}$ is the ideal generated by elements of the form ab, where $a \in \mathcal{I}$, $b \in \mathcal{J}$. This product defines a commutative group law on $\mathrm{Cl}(\mathcal{O}_K)$, the set of nonzero ideals of $\mathcal{O}_K$ modulo equivalence (i.e., modulo principal ideals). We call $\mathrm{Cl}(\mathcal{O}_K)$ the *(ideal) class group of K*. Write $[\mathcal{I}]$ for the equivalence class of $\mathcal{I}$. Then $[\mathcal{O}_K]$ is the group identity and $[\mathcal{I}]^{-1}$ means the class of some $\mathcal{J}$ such that $\mathcal{I}\mathcal{J}$ is principal.

An important theorem in algebraic number theory is that $\mathcal{O}_K$ has unique factorization of ideals into prime ideals. If all ideals are principal, then this means one has unique factorization of numbers in $\mathcal{O}_K$ into irreducible (prime) elements. In fact, $\mathcal{O}_K$ has unique factorization (of numbers) if and only if the *class number* $h_K := |\mathrm{Cl}(\mathcal{O}_K)| = 1$.

For our specific $K = \mathbb{Q}(\sqrt{-23})$, one can check that there are 3 ideal classes, represented by $\mathcal{I}_1 = \mathcal{O}_K$, $\mathcal{I}_2 = (2, \frac{1+\sqrt{-23}}{2})$ and $\mathcal{I}_3 = (2, \frac{1-\sqrt{-23}}{2})$. The group law is such that this is the group of order 3 with $\mathcal{I}_1$ being the identity. Given some ideal $\mathcal{I} = \mathbb{Z}\alpha \oplus \mathbb{Z}\beta$, we can associate the quadratic form $Q_{\mathcal{I}}(x, y) = N(\alpha x - \beta y)/N(\mathcal{I})$. (Really, $Q_{\mathcal{I}}$ depends on the basis $\{\alpha, \beta\}$ of $\mathcal{I}$, but the proper equivalence class of $Q_{\mathcal{I}}$ does not.) I won't define the norm of an ideal, but just tell you $N(\mathcal{I}_1) = 1$ and $N(\mathcal{I}_2) = N(\mathcal{I}_3) = 2$. Using this we compute

[5] Th ring $\mathbb{Z}[\sqrt{-23}]$ is the order of discriminant $-23 \cdot 4$, and would be the right thing to work with for forms like $x^2 + 23y^2$. Incidentally, $h(-92) = 6$.

that $Q_{\mathcal{I}_i} = Q_i$ for $i = 1, 2, 3$, for appropriate choice of bases. E.g., for $\mathcal{I}_1$, take $\alpha = 1$ and $\beta = -\frac{1+\sqrt{-23}}{2}$. Then

$$Q_{\mathcal{I}_1} = N(x + \frac{1+\sqrt{-23}}{2}y) = x^2 + xy + 6y^2 = Q_1.$$

This correspondence defines an isomorphism $\mathrm{Cl}(\mathcal{O}_K) \simeq \mathrm{Cl}(-\Delta)$ (or we can take this as the definition of composition of quadratic forms). The correspondence the other way is easier to describe: we can map quadratic forms to ideals by

$$Q : ax^2 + bxy + cy^2 \quad \mapsto \quad \mathcal{I}_Q = (a, \frac{b - \sqrt{\Delta}}{2}).$$

At the level of equivalence classes, this correspondence is the inverse to the map $\mathcal{I} \mapsto Q_{\mathcal{I}}$ above.

Our main motivating question is: when does some Q_i represent a prime p? Here the local-global principle that we used when $\Delta = -4$ need to be modified. We can't determine whether a single Q_i represent p by knowing if it does mod $|\Delta|$, but we can determine whether at least one of Q_1, Q_2, Q_3 represent p by considerations mod $|\Delta|$. In fact, one can prove a precise formula!

Let $r_{Q_i}(n)$ denote the number of representations of n by Q_i, i.e., the number of solutions in $\mathbb{Z} \times \mathbb{Z}$ to $Q_i(x, y) = n$. For simplicity, I will just state the following result in the case $n = p$ odd.

Theorem 1.8 (Dirichlet's mass formula)**.** *For $\Delta = -23$ and p odd, we have*

$$r_\Delta(p) := r_{Q_1}(p) + r_{Q_2}(p) + r_{Q_3}(p) = 2\left(1 + \left(\frac{\Delta}{p}\right)\right).$$

Here $\left(\frac{a}{p}\right)$ is the *Legendre symbol*, which is 0 if p divides a; otherwise it is ± 1 according to whether a is a square mod p or not. To compute this explicitly, we use *quadratic reciprocity* (also proved by Gauss, and widely considered the crown jewel of elementary number theory), which says $\left(\frac{p}{q}\right) = (-1)^{(p-1)(q-1)/4}\left(\frac{q}{p}\right)$ for p, q odd primes. In this case, it gives

$$\left(\frac{\Delta}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{23}{p}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{-1}{p}\right)\left(\frac{p}{23}\right) = \left(\frac{p}{23}\right).$$

(The last equality comes from the first supplementary law of quadratic reciprocity, which says $\left(\frac{-1}{p}\right) = 1$ if and only if $p \equiv 1 \bmod 4$.) Thus an odd prime p is represented by some form in $\mathrm{Cl}(\Delta)$ if and only if p is a square mod 23. Explicitly, the squares mod 23 are $\{0, 1, 2, 3, 4, 6, 8, 9, 12, 13, 16, 18\}$. See Table 1 for values of $r_{Q_i}(p)$ for $p < 150$ a square mod 23.

There are a couple of things we can observe from Table 1. If p is a square mod 23, either Q_1 represents p or *both* Q_2 and Q_3 do, but not all of Q_1, Q_2 and Q_3 do. One can prove this using the above theorem (since $r_\Delta(p) \leq 4$) together with a linear transformation (in $\mathrm{GL}_2(\mathbb{Z})$ but not $\mathrm{SL}_2(\mathbb{Z})$) relating Q_2 and Q_3. This is not true for general n—already $n = 4$ is represented by each Q_i. However, the residue class of p mod 23 does not determine whether Q_1 or Q_2 and Q_3 represent p: $59 \equiv 13 \bmod 23$ and Q_1 represents 59, whereas Q_2 and

Table 1: Number of representations of small n by reduced forms of discriminant $\Delta = -23$

p	p mod 23	$r_{Q_1}(p)$	$r_{Q_2}(p)$	$r_{Q_3}(p)$	$r_\Delta(p)$
2	2	0	2	2	4
3	3	0	2	2	4
13	13	0	2	2	4
23	0	2	0	0	2
29	6	0	2	2	4
31	8	0	2	2	4
41	18	0	2	2	4
47	1	0	2	2	4
59	13	4	0	0	4
71	2	0	2	2	4
73	4	0	2	2	4
101	9	4	0	0	4
127	12	0	2	2	4
131	16	0	2	2	4
139	1	0	2	2	4

Q_3 represent 13. So the local-global principle cannot be applied at the level of individual forms—each Q_i represents 13 in $\mathbb{Z}/23\mathbb{Z}$ (in fact, the forms are equivalent in $\mathbb{Z}/23\mathbb{Z}$), but not every prime $p \equiv 13 \bmod 23$ in $\mathbb{Z}$.

We remark that in general the local-global principle will apply at the level of an individual form if $\mathrm{Cl}(\Delta)$ has only one element, i.e., if the class number $h(\Delta) = 1$. So if Δ is the discriminant of some ring of integers (or order) $\mathcal{O}_K$, then $\mathcal{O}_K$ having unique factorization (like $\Delta = -4$ and $\mathcal{O}_K = \mathbb{Z}[i]$ in our sum of two squares example) implies there is only one proper equivalence class of forms $Q \in \mathrm{Cl}(\Delta)$ and the local-global principle will apply to the individual form Q.

In fact, one can refine this so that the local-global principle applies to certain nice subsets of $\mathrm{Cl}(\Delta)$. Each form class group can be partitioned into *genera* (plural of *genus*), and the local-global principle applies for each genus. Two forms being in the same genus means they are equivalent mod m for each m, and it suffices to check this for $m = |\Delta|$. In the case of $\Delta = -23$, there is only one genus, so we cannot separate Q_1, Q_2 and Q_3, but in the case of $\Delta = -20$, there are two classes of forms which are in separate genera, so one can apply the local-global principle for each form.[6] (This is why I think $\Delta = -23$ is more interesting.) In general, we can determine the primes represented by a given form of discriminant Δ just by looking at congruences mod $|\Delta|$ if each genus in $\mathrm{Cl}(\Delta)$ has size one. The phrase we usually use for this is "one class per genus." There is one class per genus if and only if $\mathrm{Cl}(\Delta)$ has no elements of order > 2, i.e., $\mathrm{Cl}(\Delta) \simeq (\mathbb{Z}/2\mathbb{Z})^m$ for some m.

[6] To go back to our earlier examples: for $\Delta = -8$, $p = x^2 + 2y^2$ if and only if $p = 2$ or $p \equiv 1, 3, \bmod 8$; for $\Delta = -3$, we have $p = x^2 - xy + y^2$ if and only if $p = 3$ or $p \equiv 1 \bmod 3$; for $\Delta = -20$, we have $p = x^2 + 5y^2$ if and only if $p = 5$ or $p \equiv 1, 9 \bmod 20$ and $p = 2x^2 + 2xy + 3y^2$ if and only if $p = 2$ or $p \equiv 3, 7 \bmod 20$.

Table 2: Class numbers for negative discriminants

Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$
-3	1	-52	2	-103	5	-152	6	-203	4	-252	10
-4	1	-55	4	-104	6	-155	4	-204	8	-255	12
-7	1	-56	4	-107	3	-156	8	-207	9	-256	8
-8	1	-59	3	-108	6	-159	10	-208	6	-259	4
-11	1	-60	4	-111	8	-160	6	-211	3	-260	8
-12	2	-63	5	-112	4	-163	1	-212	6	-263	13
-15	2	-64	4	-115	2	-164	8	-215	14	-264	8
-16	2	-67	1	-116	6	-167	11	-216	8	-267	2
-19	1	-68	4	-119	10	-168	4	-219	4	-268	4
-20	2	-71	7	-120	4	-171	5	-220	8	-271	11
-23	3	-72	3	-123	2	-172	4	-223	7	-272	12
-24	2	-75	3	-124	6	-175	7	-224	12	-275	5
-27	2	-76	4	-127	5	-176	10	-227	5	-276	8
-28	2	-79	5	-128	7	-179	5	-228	4	-279	15
-31	3	-80	6	-131	5	-180	6	-231	12	-280	4
-32	3	-83	3	-132	4	-183	8	-232	2	-283	3
-35	2	-84	4	-135	8	-184	4	-235	2	-284	14
-36	3	-87	6	-136	4	-187	2	-236	12	-287	14
-39	4	-88	2	-139	3	-188	10	-239	15	-288	9
-40	2	-91	2	-140	8	-191	13	-240	8	-291	4
-43	1	-92	6	-143	10	-192	8	-243	5	-292	4
-44	4	-95	8	-144	8	-195	4	-244	6	-295	8
-47	5	-96	6	-147	3	-196	5	-247	6	-296	10
-48	4	-99	3	-148	2	-199	9	-248	8	-299	8
-51	2	-100	3	-151	7	-200	7	-251	7	-300	10
-52	2	-103	5	-152	6	-203	4	-252	10	-303	10

1.3 Class numbers

A natural question to ask is: what is the behaviour of the class numbers $h(\Delta)$, where $\Delta < 0$ is a discriminant (meaning the discriminant of some binary quadratic form)? For instance, when is $h(\Delta) = 1$? In this case, there is only one positive definite form Q (up to proper equivalence) with discriminant Δ, so one has a composition law $Q^2 = Q$, and one can determine all n represented by Q by reducing to the $n = p$ case and using the local-global principle or Dirichlet's mass formula. Here one can get a completely elementary answer just as in the case of sums of two squares (see Footnote 6 for a couple of examples).

However, class numbers behave mysteriously, almost randomly, like prime numbers. Gauss, being a master of calculation, computed a large amount of class numbers, and based on this conjectured that $h(\Delta) \to \infty$ as $\Delta \to -\infty$. See Table 2 for some calculations. In particular, there are only a finite number of Δ with given class number h. This was proved by Heilbronn in 1934. However, the proof is not effective, meaning one cannot actually determine all Δ with a given class number.

Gauss further conjectured that there are exactly 13 negative discriminants of class number 1 (9 of which are *fundamental*, meaning the discriminant of some imaginary quadratic number field), the largest one (in absolute value) being -163.[7] (See Table 2 to determine Gauss' list.) This was essentially proved by Heegner (an "amateur" mathematician) in 1952 using modular forms, but his work wasn't understood or accepted until 1967 when Stark understood it and corrected a minor gap. In the meantime, it was settled (with an accepted proof) by Baker in 1966 using completely different methods.

The general problem of determining all $\Delta < 0$ with a given class number h was observed to be related to L-functions of elliptic curves by Goldfeld, and is currently solved for (at least) $h < 100$ by Watkins. We'll discuss elliptic curves and their L-functions in another context in Section 3.2.

Despite the class numbers behaving in essentially a random way, Dirichlet discovered a formula for the class numbers. One can reduce to the case of fundamental discriminant $\Delta < 0$, so we will assume that now. One defines what is now called a *Dirichlet character*

$$\chi_\Delta : \mathbb{N} \to \{1, 0, -1\} \qquad \chi_\Delta(n) = \left(\frac{\Delta}{n}\right),$$

where $\left(\frac{a}{b}\right)$ is a suitable, multiplicative extension of the Legendre symbol to arbitrary $b \in \mathbb{N}$. Then χ_Δ is multiplicative. For a character $\chi : \mathbb{N} \to \{1, 0, -1\}$, one defines the L-series

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

These kinds of series, i.e. those of the form $\sum \frac{a_n}{n^s}$ for some sequence (a_n), are now called *Dirichlet series.* The above series converges if $\mathrm{Re}(s) > 1$.

Furthermore, χ being multiplicative means there is an *Euler product*

$$L(s, \chi) = \prod_p (1 - \chi(p)p^{-s})^{-1},$$

which again is valid for $\mathrm{Re}(s) > 1$.

Let me sketch how one gets the Euler product. Consider the first two terms ($p = 2, 3$) of the Euler product above

$$\begin{aligned}\left(\frac{1}{1 - \frac{\chi(2)}{2^s}}\right)\left(\frac{1}{1 - \frac{\chi(3)}{3^s}}\right) &= \left(1 + \frac{\chi(2)}{2^s} + \frac{\chi(2)^2}{2^{2s}} + \cdots\right)\left(1 + \frac{\chi(3)}{3^s} + \frac{\chi(3)^2}{3^{2s}} + \cdots\right) \\ &= 1 + \frac{\chi(2)}{2^s} + \frac{\chi(3)}{3^s} + \frac{\chi(4)}{4^s} + \frac{\chi(6)}{6^s} + \frac{\chi(8)}{8^s} + \frac{\chi(9)}{9^s} + \cdots\end{aligned}$$

Here the first equality comes from the geometric series expansion, and the second comes from multiplying out terms and using the fact that $\chi(mn) = \chi(m)\chi(n)$. The expansion can be justified for $\mathrm{Re}(s)$ large by showing both expressions converge. Note that the right

[7] In terms of quadratic fields, class number one means the ring of integers has unique factorization: e.g., $h(-4) = 1$ means the ring of integers $\mathbb{Z}[i]$ in $\mathbb{Q}(\sqrt{-4}) = \mathbb{Q}(i)$ has unique factorization, where as $h(-20) = 2$ means the ring of integers $\mathbb{Z}[\sqrt{-5}]$ in $\mathbb{Q}(\sqrt{-20}) = \mathbb{Q}(\sqrt{-5})$ does not.

hand side will be the sum of the terms $\frac{\chi(n)}{n^s}$ over precisely the n of the form $2^e 3^f$ for some e, f. Inductively adding more primes in this product (and checking convergence) gives the equality

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

When $\chi \equiv 1$ is the trivial character, this is precisely the Riemann zeta function $\zeta(s) = L(s,1)$. For the zeta function, we get a pole at $s = 1$. This is seen from the series expansion: $\zeta(1)$ is the harmonic series. Note that any factor in the Euler product is a finite number when $s = 1$, so $\zeta(1) = \infty$ implies there must be infinitely many primes, and this is Euler's proof of the infinitude of primes. (This may seem like overkill, but besides being very cool, the idea is important: one can use the Riemann zeta function to estimate the number of primes $< x$, which is far from obvious using only elementary methods.)

On the other hand, for a fundamental discriminant Δ, $L(s,\chi_\Delta)$ can be extended to an entire function on $\mathbb{C}$—i.e., the Dirichlet L-functions $L(s,\chi_\Delta)$ have no poles. In particular, the series expansion converges (conditionally) at $s = 1$, and Dirichlet showed the value at $s = 1$ is related to the class number. We only state Dirichlet's class number formula for $\Delta < -4$ for simplicity.

Theorem 1.9 (Dirichlet)**.** *Suppose $\Delta < -4$. Then*

$$h(\Delta) = \frac{\sqrt{|\Delta|}}{\pi} L(1,\chi_\Delta).$$

Furthermore, one can prove a formula for $L(1,\chi_\Delta)$ in terms of the values of χ_Δ. This gives a more computationally explicit form of Dirichlet's class number formula (again just stated for $\Delta < -4$):

$$h(\Delta) = \frac{1}{2 - \chi_\Delta(2)} \Big| \sum_{1 \le k < \frac{|\Delta|}{2}} \chi_\Delta(k) \Big|.$$

While it is nice to have this explicit formula, this formula was not directly useful in answering Gauss' class number questions, as χ_Δ will oscillate between positive and negative, so it is hard to determine the size of the right hand side (and as you see from class number tables, it fluctuates quite a bit).

We remark that one can do something similar for $\Delta > 0$, where $\mathrm{Cl}(\Delta)$ will be a class group of indefinite binary quadratic forms which correspond to ideal class groups for real quadratic fields (at least for the fundamental $\Delta > 0$). In contrast to the negative discriminant case, much less is known about the positive discriminant case. A table of class numbers for positive fundamental discriminants is given in Table 3.

As you can see, class numbers tend to be a lot smaller for positive discriminants. Gauss conjectured that, in contrast to the negative discriminant case, class number 1 occurs for infinitely many positive discriminants. Again, there is a Dirichlet class number formula in this case, but it is complicated by occcurences of log and sin. However, knowing some things about elliptic curve L-functions could tell us that we get class number 1 infinitely often for positive discriminants.

Table 3: Class numbers for positive *fundamental* discriminants

Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$	Δ	$h(\Delta)$
5	1	57	1	105	2	161	1	213	1	268	1
8	1	60	2	109	1	165	2	217	1	269	1
12	1	61	1	113	1	168	2	220	2	273	2
13	1	65	2	120	2	172	1	221	2	277	1
17	1	69	1	124	1	173	1	229	3	280	2
21	1	73	1	129	1	177	1	232	2	281	1
24	1	76	1	133	1	181	1	233	1	284	1
28	1	77	1	136	2	184	1	236	1	285	2
29	1	85	2	137	1	185	2	237	1	293	1
33	1	88	1	140	2	188	1	241	1	296	2
37	1	89	1	141	1	193	1	248	1	301	1
40	2	92	1	145	4	197	1	249	1	305	2
41	1	93	1	149	1	201	1	253	1	309	1
44	1	97	1	152	1	204	2	257	3	312	2
53	1	101	1	156	2	205	2	264	2	313	1
56	1	104	2	157	1	209	1	265	2	316	3

2 Higher dimensional quadratic forms

Now we move on to the problem: what numbers are sums of k squares, i.e., when does

$$Q_k(x_1, \dots, x_k) = x_1^2 + \cdots + x_k^2 = n \qquad x_1, \dots, x_k \in \mathbb{Z}$$

have a solution. This form is called a k-ary quadratic form (i.e, there are k variables and each monomial has degree 2), and again one can consider arbitrary k-ary forms, but we will just focus on the forms Q_k above.

When $k = 2$, we get binary forms. Simlarly, we call the forms for $k = 3$ and $k = 4$ ternary and quaternary quadratic forms.

We remark that a k-ary quadratic form Q can be viewed a function $Q : \mathbb{R}^n \to \mathbb{R}$ (or from $\mathbb{Q}^n \to \mathbb{Q}$, or $\mathbb{C}^n \to \mathbb{C}$). Then $B(u, v) = (Q(u+v) - Q(u) - Q(v))/2$ defines a symmetric bilinear form on $\mathbb{R}^n$ such that $B(v, v) = Q(v)$ and makes $\mathbb{R}^n$ into what is called a quadratic space. In linear algebra, one often looks at the orthogonal group $O(Q)$ of this space, i.e., $O(Q)$ is the group of invertible linear operators on $\mathbb{R}^n$ which preserve Q. For $Q = Q_k$, one gets the usual orthogonal group $O(k)$, which is the isometry group of the k-dimensional sphere $\{v \in \mathbb{R}^n : Q_k(v) = 1\}$. Much of the algebraic theory of quadratic forms is devoted to a study of these quadratic spaces and their orthogonal groups. While this theory is important also in the arithmetic of quadratic forms, we will not focus on it here.

Primary references for this section are [Gro85] and my modular forms notes [Mar], which almost list additional references. (In fact, Section 2.2 is largely taken from my introduction in [Mar].)

2.1 Ternary and quaternary quadratic forms

Now you might think the next easiest case after sums of two squares is sums of three squares, but actually it's sums of four squares.

Theorem 2.1 (Lagrange (1770))**.** *Every integer is the sum of four squares.*

There are many ways to prove this. One approach is to use quaternions. The connection with quaternions hints at a higher-dimensional generalization of the connection between binary quadratic forms and ideals in quadratic fields.

Let

$$\mathbb{H} = \mathbb{R}[i, j, k] = \{a + bi + cj + dk | a, b, c, d \in \mathbb{R}\},$$

where

$$i^2 = j^2 = k^2 = ijk = -1.$$

This is a *non-commutative* division algebra,[8] known as Hamilton's quaternions. It is some sort of generalization of the complex numbers. Just as $\mathbb{C}$ is 2-dimensional over $\mathbb{R}$, this is 2-dimensional over $\mathbb{C}$, or 4-dimensional over $\mathbb{R}$ (as a vector space). $\mathbb{H}$ has many applications, and is closely related to the algebra of 2×2 real matrices. Note that going from $\mathbb{R}$ to $\mathbb{C}$, one loses the natural well-ordering one had on $\mathbb{R}$, and going from $\mathbb{C}$ to $\mathbb{H}$ one loses commutativity. Incidentally, one can extend the quaternions to the octonions $\mathbb{O}$, but then one loses associativity. The octonions also have numerous applications, and can be used to study sums of 8 squares.

The norm on $\mathbb{H}$ is given by

$$N(a + bi + cj + dk) = a^2 + b^2 + c^2 + d^2.$$

The integers of (the rational subfield of) $\mathbb{H}$, called the *Hurwitz integers* are

$$\mathcal{O}_{\mathbb{H}} = \mathbb{Z}\big[i, j, k, \frac{1 + i + j + k}{2}\big] = \left\{ \frac{a + bi + cj + dk}{2} \middle| a, b, c, d \in \mathbb{Z}, a \equiv b \equiv c \equiv d \bmod 2 \right\}.$$

Then

$$N : \mathcal{O}_{\mathbb{H}} \to \mathbb{Z}.$$

One can check that n is a sum of four squares if and only if $n = N(\alpha)$ for some $\alpha \in \mathcal{O}_{\mathbb{H}}$, which means that n factors in $\mathcal{O}_{\mathbb{H}}$. The fact that N is multiplicative means we have a composition law for Q_4, i.e., the product of sums of four squares is again a sum of four squares. Thus we can reduce Lagrange's four square theorem to showing every prime is a sum of four squares.

Suppose p is not a sum of four squares, so it does not factor in $\mathcal{O}_{\mathbb{H}}$, i.e., it is irreducible in $\mathcal{O}_{\mathbb{H}}$. A lemma of Lagrange says that p divides $1 + a^2 + b^2 = (1 + ai + bj)(1 - ai - bj)$ for some $a, b \in \mathbb{Z}$. It is a fact that $\mathcal{O}_{\mathbb{H}}$ possesses unique factorization (suitably defined), which is equivalent to saying every ideal in $\mathcal{O}_{\mathbb{H}}$ is principal, i.e., it has class number 1. Hence p being irreducible in $\mathcal{O}_{\mathbb{H}}$ means p must divide $1 + ai + bj$ or $1 - ai - bj$ in $\mathcal{O}_{\mathbb{H}}$, but $\frac{1 \pm ai \pm bj}{p} \notin \mathcal{O}_{\mathbb{H}}$, a contradiction. This gives Lagrange's theorem.

[8] An *algebra* is a vector space (over some field) that is simultaneously a ring. E.g., any field is an algebra. So are matrix algebras, such as $\mathrm{Mat}_{n \times n}(\mathbb{R})$.

I'll give you a formula for the number of representations of n as a sum of four squares in the next section.

In general, you might wonder about composition laws for general quaternary quadratic forms. For binary forms, the correspondence was with quadratic number fields K. An analogous correspondence for $k = 4$ should associate certain 4-dimensional algebras like $\mathbb{H}$ to quaternary forms.. The right algebras to look at are those known as quaternion algebras $B/\mathbb{Q}$, and there are infinitely many. The analogue of the ring of integers (or more generally, an order) in K is the notion of an order $\mathcal{O}$ in B, such as $\mathcal{O}_{\mathbb{H}}$. However B being noncommutative means that one cannot always multiply ideals in $\mathcal{O}$ (the ideal classes still make sense, and they are finite in number, but they do not form a group). Nevertheless, for suitably compatible ideals $\mathcal{I}$ and $\mathcal{J}$, the product $\mathcal{I}\mathcal{J}$ does make sense. By associating quaternary quadratic forms to these ideals, one gets a partial composition law on certain quaternary quadratic forms. Note that one does not get all quaternary quadratic forms this way, only those of "quaternionic type." See [KOK+86] for more on these partial composition laws.

For three squares, we have the following theorem.

Theorem 2.2 (Legendre (1798?))**.** *A natural number n is the sum of three squares if and only if $n \neq 4^j(8k+7)$ for all j, k. In particular, a prime p is the sum of three squares if and only $p \not\equiv 7 \bmod 8$.*

The prime case of Legendre's three square theorem can be easily reduced to the problem for binary quadratic forms. One can check that 7 is not a sum of three squares mod 8, so one direction is easy, and $p = 2$ is simple. If $p \equiv 1, 5 \bmod 8$, i.e., $p \equiv 1 \bmod 4$, then $p = x^2 + y^2$ by Fermat's two square theorem, and thus p is also a sum of three squares (one square being 0). An analogue of Fermat's two square theorem is that $p = x^2 + 2y^2$ if and only if $p \equiv 1, 2, 3 \bmod 8$. So if $p \equiv 3 \bmod 8$, then $p = x^2 + y^2 + z^2$ with $y = z$. This yields the prime case of Legendre's theorem.

The general case is more difficult essentially because there is no nice composition law for ternary quadratic forms. However, Gauss still used his theory of binary quadratic forms to prove a quantitative version of Legendre's theorem: the number of "primitive solutions" (meaning $\gcd(x, y, z) = 1$) to $x^2 + y^2 + z^2 = n$ is (for $n > 3$)

$$R_3(n) = \begin{cases} 12h(-4n) & n \equiv 1, 2 \bmod 4 \\ 24h(-n) & n \equiv 3 \bmod 8 \\ 0 & \text{else.} \end{cases}$$

In case you like the idea of using quaternions for four squares (I do), you can also use them to address 3 squares. Here the way to associate a ternary quadratic form to a quaternion algebra is to restrict it to the "pure quaternions," e.g., restricting the norm to

$$\mathbb{H}^0 = \{xi + yj + zk : x, y, z \in \mathbb{R}\} \subset \mathbb{H}$$

yields the quadratic form $x^2+y^2+z^2$. In some sense the reason that there is no composition law for $x^2 + y^2 + z^2$ is because the product of two pure quaternions is not in general a pure

quaternion.[9]

Another proof (for 3 or 4 squares) is to use the local-global principle, which applies to quadratic forms in dimensions higher than 2 as well. See, e.g., my Number Theory II notes [Marb] or Serre's classic [Ser73].

2.2 Quadratic forms in arbitrary dimension

Since any natural number is a sum of four squares, we know any natural number is a sum of k squares for $k \geq 4$, and thus Question 1 is solved for every Q_k. (See [Gro85] for a statement, without proof, in the case $k = 1$.) Now we want to say something about the Question 2. Denote by $r_k(n)$ the number of representations of n by Q_k.

Here is a general approach to determine $r_k(n)$. Jacobi considered the *theta function*

$$\vartheta(z) = \sum_{n=-\infty}^{\infty} q^{n^2}, \quad q = e^{2\pi i z}. \tag{4}$$

This function is well defined for $z \in \mathfrak{H} = \{x + iy : x, y \in \mathbb{R}, y > 0\}$. We call $\mathfrak{H}$ the *upper half plane*. Then

$$\vartheta^2(z) = \left(\sum_{\ell=-\infty}^{\infty} q^{\ell^2}\right)\left(\sum_{m=-\infty}^{\infty} q^{m^2}\right) = \sum_{\ell,m} q^{\ell^2+m^2} = \sum_{n\geq 0} r_2(n) q^n.$$

Similarly,

$$\vartheta^k(z) = \sum_{n\geq 0} r_k(n) q^n. \tag{5}$$

It is not too difficult to see that ϑ^k satisfies the identities

$$\vartheta^k(z+1) = \vartheta^k(z), \quad \vartheta^k\left(\frac{-1}{4z}\right) = \left(\frac{2z}{i}\right)^{\frac{k}{2}} \vartheta^k(z). \tag{6}$$

Indeed, the first identity is obvious because q is invariant under $z \mapsto z + 1$.

The space of (holomorphic) functions on $\mathfrak{H}$ satisfying the transformation properties is (6) is defined to be the space of *modular forms* $M_{k/2}(4)$ of *weight* $k/2$ and *level* 4. The theory of modular forms will tell us that $M_{k/2}(4)$ is a finite-dimensional vector space.

Let me outline the details in essentially the simplest case, $k = 4$ (Lagrange's case). Here $M_2(4)$ is a 2-dimensional vector space, and one can find a basis in terms of *Eisenstein series*. Specifically, consider the Eisenstein series

$$G(z) = -\frac{1}{24} + \sum_{n=1}^{\infty} \sigma(n) q^n, \tag{7}$$

where $\sigma(n)$ is the divisor function $\sigma(n) = \sum_{d|n} d$. Then a basis of $M_2(4)$ is

$$f(z) = G(z) - 2G(2z), \quad g(z) = G(2z) - 2G(4z).$$

[9] There is a composition law for $x_1^2 + \cdots + x_k^2$ if and only if $k = 1, 2, 4, 8$. This has to do with the existence of suitable algebras of real dimension k, which only occur for these values of k, namely $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and the octonions $\mathbb{O}$.

Hence $\vartheta^4(z)$ is a linear combination of $f(z)$ and $g(z)$. How do we determine what combination? Simply compare the first two coefficients of q^n in $af(z) + bg(z)$ with $\vartheta^4(z)$, and one sees that

$$\vartheta^4(z) = 8f(z) + 16g(z).$$

Expanding this out, one sees that

$$\vartheta^4(z) = \sum_{n\geq 0} r_4(n)q^n = 1 + 8\sum_{n\geq 1}\sigma(n)q^n - 32\sum_{n\geq 1}\sigma(n)q^{4n}. \tag{8}$$

Consequently

$$r_4(n) = \begin{cases} 8\sigma(n) & 4 \nmid n \\ 8\sigma(n) - 32\sigma(n/4) & 4|n. \end{cases}$$

If one wishes, one can write this as a single formula

$$r_4(n) = 8\left(2 + (-1)^n\right)\sum_{d|n, 2\nmid d} d.$$

In particular, it is obvious that $r_4(n) > 0$ for all n, in other words, we have Lagrange's theorem that every positive integer is a sum of four squares. Furthermore, we have a simple formula for the number of representations of n as a sum of four squares, in terms of the divisors of n.

This approach adapts to larger k as well (the k even case being easier), however $M_{k/2}(4)$ is no longer generated by linear combinations of the Eisenstein series $G(Nz)$ for various N, but rather generated (linearly) by Eisenstein series and *cusp forms*. The formula for $r_4(n)$ came out quite simple because the Fourier coefficients of the Eisenstein series (i.e., the coefficients of q^n in (7)) have a simple expression. On the other hand the Fourier coefficients of the cusp forms are more complicated and mysterious, but are asymptotically smaller than the Fourier coefficients of Eisenstein series. In general ϑ^k will be a linear combination of Eisenstein series and cusp forms (though for $k = 8$ one also only needs Eisenstein series),[10] so one can get a simple asymptotic formula for $r_k(n)$, but not a simple exact formula. Nevertheless, for specific values of k, one can work out more complicated exact formulas for $r_k(n)$ by, say, writing cusp forms as *polynomial* combinations of Eisenstein series. These expressions involve polynomial combinations of the higher power divisor functions $\sigma_m(n) = \sum_{d|n} d^m$.[11]

3 Binary cubic forms

Now that we have some ideas about representability questions for quadratic forms, let's explore a little in higher degree. A *binary cubic form* is a polynomial of the form

$$F(x, y) = ax^3 + bx^2y + cxy^2 + dy^3,$$

[10] One can also get an elementary formula for $r_8(n)$ using octonions. However, for general values of k there is no nice algebra to associate with the form Q_k.

[11] For special values of k, other formula may exist. For instance, for $k = 4m^2$ or $4m^2 + 4m$, Milne [Mil96] discovered interesting combinatorial expressions for $r_k(n)$ using connections with Lie algebras.

where we will take $a, b, c, d \in \mathbb{Z}$. Here I will just focus on the form $x^3 + y^3$, first over $\mathbb{Z}$, then over $\mathbb{Q}$, but I hope to add a section on composition laws for binary cubic forms later. (These composition laws are of a different nature than Gauss's.)

I drew the material here from a smattering of sources, such as the elementary article [Sil93], the research articles [RVZ95] and [DV09] and various (basic and advanced) facts about elliptic curves, though many of the results are also summarized in [Coh07, Section 6.4.6]. If I convince you to learn more about elliptic curves, there are many good sources. For instance, [ST15] is a particularly nice elementary introduction and [Sil09] is the gold standard for a more serious study.[12]

3.1 Sums of two integer cubes

A natural generalization of the sum of two squares question is: what numbers n are sums of two (or three, or four) cubes, i.e., when is $n = x^3 + y^3$ for $x, y \in \mathbb{Z}$?[13] Note this question makes sense for $n \in \mathbb{Z}$, but the answer for n is the same as the answer for $-n$, so we may assume $n > 0$. Recall that for the two squares problem, we used the factorization $x^2 + y^2 = (x + y)(x - y)$. In the case at hand, we have the factorization[14]

$$x^3 + y^3 = (x + y)(x^2 - xy + y^2) = n. \tag{9}$$

Thus, if this is solvable, we have a factorization of $n = rs$ with $r = x+y$ and $s = x^2-xy+y^2$. Writing $y = r - x$ and substituting yields the quadratic equation $s = 3x^2 - 3rx + r^2$, so there are at most two solutions for any (positive or negative) divisor s of n, and we can algorithmically check whether a given n is a sum of two cubes. (Note here x and y are allowed to be positive and negative integers, so one can't just check x, y up to $\sqrt[3]{n}$.)

While we don't have a composition law that will allow us to reduce the question for composite n to the question for n prime,[15] let's just focus on the case $n = p$ is prime for simplicity. Here there is esssentially one factorization of n giving the four possibilities $r = \pm 1$ and $r = \pm p$. If $r = \pm 1$ so $s = \pm p$, then the question is just what primes p satisfy $\pm p = 3x^2 \mp 3x + 1$. Since the polynomial on the right is is always positive, we can only have $s = +p$, and we get the primes of the form

$$p = 3x^2 - 3x + 1, \qquad x \in \mathbb{Z} \tag{10}$$

Note that $f(x) = 3x^2 - 3x + 1 = 3x(x - 1) + 1$ satisfies $f(-x) = f(x + 1)$, so the set of values of f at positive integers are the same as the set of values of f at negative integers,

[12] There don't seem to be too many introductory sources that treat sums of two cubes. Personally I think it is a great problem for motivating the study of elliptic curves, or at least to give as an application, but it seems to have gotten lost behind things like applications of elliptic curves to Fermat's last theorem, the congruent number problem (what numbers occur as areas of right triangles with rational length sides), and cryptography.

[13] If you're curious about an analogue of Lagrange's four squares theorem, for cubes it known that every integer is a sum of five integer cubes, and it is conjectured that four suffice (cf. [Coh07, Section 6.4.6]). If you want to restruct to positive numbers, it is known that every natural number is a sum at most 9 cubes of natural numbers. The analogous question for general k-th powers is called *Waring's problem*.

[14] One can further factor into linear factors over $\mathbb{Q}(\sqrt{-3})$ as $x^3 + y^3 = (x + y)(x - \frac{1+\sqrt{-3}}{2}y)(x - \frac{1-\sqrt{-3}}{2}y)$, however this is not necessary for our current discussion.

[15] For instance, $1729 = 7 \cdot 13 \cdot 19$ is famously the sum of two cubes (see Footnote 18), but of these three prime factors, only 7 is itself a sum of two cubes, as we will see momentarily.

meaning we can just check this for $x \in \mathbb{N}$. In fact we can replace (10) with the condition that $p = f(x+1) = 3x^2 + 3x + 1$ for some $x \in \mathbb{N}$. (My tiny brain finds polynomials with positive coefficients easier to understand, so excuse this indulgence.) Clearly any number of the form $3x^2+3x+1$ must be 1 mod 3—one can further check it satisfies the finer condition $\equiv 1, 7 \bmod 9$.[16]

On the other hand, if $r = \pm p$, then we need $\pm 1 = 3x^2 \mp 3px + p^2$. Again, if we assume $p > 0$, then a solution with $r = -p$ is clearly impossible, so we would need $3x^2 - 3px + p^2 = 1$. The polynomial $f(x) = 3x^2 - 3px + p^2$ has minimum $f(\frac{p}{2}) = \frac{p^2}{4}$, so we only get the solution $x = 1$, $p = 2$. This shows

Proposition 3.1. *A prime p is a sum of two integer cubes if and only if $p = 2$ or $p = 3x^2 + 3x + 1$ (or equivalently $p = 3x^2 - 3x + 1$) for some $x \in \mathbb{Z}$ (or equivalently for some $x \in \mathbb{N}$). A necessary condition for $p > 2$ is $p \equiv 1, 7 \bmod 9$.*

Note the proof also shows that every (not necessarily prime) value of $3x^2 + 3x + 1$ is a sum of two integer cubes, though not that every sum of two cubes is represented by this polynomial. We also get from the factorization that *no prime $p > 2$ is a sum of two positive cubes*, because we need $y = 1 - x$.

Determining what primes are attained by quadratic polynomials at the integers is a hard problem in number theory. For instance, it is not even known if any quadratic polynomials take on prime values infinitely often, though Hardy and Littlewood have conjectures about how often you should get prime values (in particular, infinitely often if you avoid obvious counterexamples like $f(x) = x^2$ or $f(x) = 5x^2 + 10$). It turns out the first four values of $3x^2 + 3x + 1$ are prime: 7, 19, 37, 61, but the next, 91, is not. (This is another reason why I prefer $3x^2 + 3x + 1$ to $3x^2 - 3x + 1$.) Note 43 is the first example of a prime $\equiv 1, 7 \bmod 9$ which is not represented as a sum of two cubes. See Table 4 for which small primes are sums of two cubes.

From what we've done, we can conclude that most primes $p \equiv 1, 7 \bmod 9$ cannot be sums of two cubes. Namely, the Prime Number Theorem[17] says that number of primes less than x is asymptotic to $\frac{x}{\log x}$. Then Dirichlet's density theorem tells us that, asymtotically, the primes fall along the classes $1, 2, 4, 5, 7, 8 \bmod 9$ equally often. Hence, for instance, the number of primes $\equiv 1 \bmod 9$ less than x is asymptotic to $\frac{x}{6 \log x}$. On the other hand the number of integers represented by the quadratic polynomial $3t^2 + 3t + 1$ less than x is bounded by $\sqrt{\frac{x}{3}}$. Therefore the proportion of primes $\equiv 1 \bmod 9$ (or $\equiv 7 \bmod 9$) which are sums of two cubes must be asymptotically 0.

Having said all this, in some sense the above proposition is an analogue of our answer to the sum of two squares question (and easier to prove!). We can restate the two squares case as: p is a sum of two squares if and only if $p = 2$ or $p = 4x + 1$ for some x. This expresses what primes are represented by a binary quadratic form in terms of what primes are represented by a univariate linear polynomial. Proposition 3.1 relates what primes are

[16] Going back to the factorization of $x^3 + y^3$, we conclude any prime which is the sum of two cubes is represented by the positive definite binary quadratic form $x^2 - xy + y^2$ from Example 1.5. Since this discriminant $\Delta = -3$ has class number 1, we can apply the local-global principle to see that $p = x^2 - xy + y^2$ if and only if $p = 3$ or $p \equiv 1 \bmod 3$. The additional condition $x + y = 1$ precludes most of these p from being sums of two cubes, in particular $p = 3$ or any $p \equiv 4 \bmod 7$.

[17] This is one of the results we indicated earlier that can be proved using the Riemann zeta function.

Table 4: Small primes $\equiv 1, 7 \bmod 9$: bold entries are sums of 2 integer cubes

1 mod 9	7 mod 9	1 mod 9	7 mod 9	1 mod 9	7 mod 9
19	**7**	523	439	1063	1069
37	43	541	457	1117	1087
73	**61**	577	**547**	1153	1123
109	79	613	601	1171	1213
127	97	**631**	619	1279	1231
163	151	739	673	1297	1249
181	223	757	691	1423	1303
199	241	811	709	1459	1321
271	277	829	727	1531	1429
307	313	883	853	1549	1447
379	**331**	**919**	907	1567	1483
397	349	937	997	1621	1609
433	367	991	1033	**1657**	1627
487	421	1009	1051	1693	1663
523	439	1063	1069	1747	1699

represented by a binary cubic form in terms of what primes are represented by a univariate quadratic polynomial.

3.2 Sums of two rational cubes

Now we move on to a related question.[18]

Question 3. *What integers are sums of two rational cubes?*[19] *That is, for what n does*

$$x^3 + y^3 = n, \qquad x, y \in \mathbb{Q} \tag{11}$$

have a solution?

The answer may also give us more information about what numbers are sums of two integral cubes, as anything that's not a sum of two rational cubes can't be a sum of two integral cubes. Further, a rational solution $x = \frac{a}{b}$, $y = \frac{c}{d}$ to (11) means we have an integer solution $(x, y) = (a, c)$ to the curve $x^3 + y^3 = n(bd)^3$. Consequently, knowing everything about integer solutions to $x^3 + y^3 = n$ for all n is equivalent to knowing everything about rational solutions to the same equation for all n.

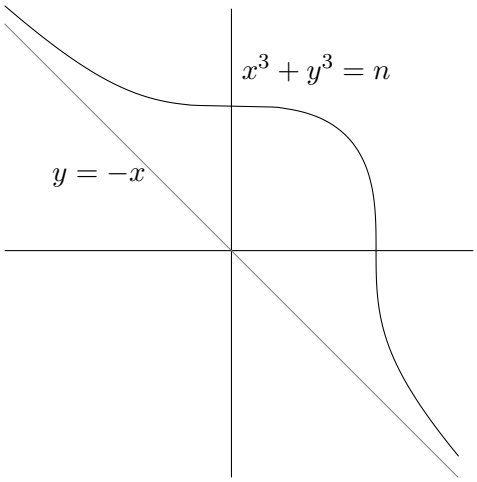
On the other hand, given some n, it is not as easy to algorithmically determine whether (11) has a rational solution. We of course still have the factorization $x^3 + y^3 = (x +$

[18] The nice, elementary article [Sil93] looks at a different related question—how many integer solutions can $n = x^3 + y^3$ have? (By earlier considerations, it is at most four times the number of divisors of n.) This is related to what are called taxicab numbers, in light of Ramanujan's remark to Hardy on riding on a taxi numbered 1729 that 1729 is very interesting because it is the smallest number which is the sum of two (integer) cubes in two different ways: $1729 = 1^3 + 12^3 = 9^3 + 10^3$. It turns out that given some r, there are infinitely many n which are expressible as the sum of two cubes in at least r different ways.

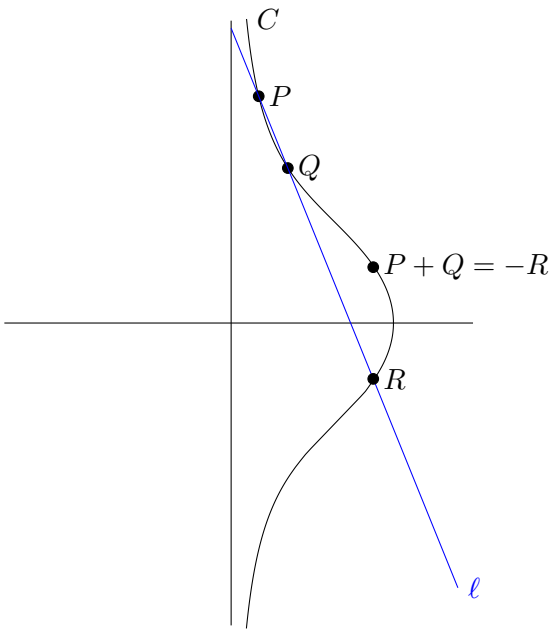
[19] All rationals are sums of three rational cubes. See, e.g., [HW08, Section 13.6] or [Coh07, Section 6.4.6].

$y)(x^2 - xy + y^2)$, however, if we are working with rational numbers, there are infinitely many possibilities for a factorization into two rationals. Hence our previous approach for determining what numbers are sums of two integer cubes does not give us a finite number of possibilities. So, computationally this is a harder problem.

The benefit, though, of asking the question about rational cubes is that this allows us to put some structure on the set of solutions to (11). Let me draw a picture of the curve defined by (11) over $\mathbb{R}$.



One can put a group structure on the set of points of this curve in $\mathbb{R}^2$. There are different ways to define it. I'll do it from an elementary geometric perspective (the usual first definition one sees, though not the most enlightening). To make it slightly easier to describe, let me rotate the graph so the asymptote is the y-axis rather than the line $y = -x$. Denote the rotated curve in $\mathbb{R}^2$ by C.



Now the (additive) group law on C is defined as follows. For any point P on C, we define $-P$ to be the reflection of P about x-axis. By the symmetry of C, we see that $-P$ also lies on C. Hence the map $P \mapsto -P$ is an involution of C (right now, just as a set—it's a bijective function which is its own inverse).

Given two points P and Q, draw the secant line ℓ through P and Q. Then, apart from exceptional cases, since this is a degree 3 curve, ℓ will intersect C in exactly one other point R. We define the sum $P + Q = -R$.

The notation $-R$ indicates that $-R$ should be the additive inverse of R. (I know we haven't defined the identity of the group yet, but it will come out of these considerations.) If we draw the secant line through R and $-R$, it won't intersect C in a third point—this is one of the exceptional cases. So we artificially add a point O to C, called the *point at infinity*, and we think of it as being infinitely far up, and also down, along the y axis. I.e, you can think of the coordinates of O as $(x, \pm\infty)$ for any x. Then reflection about the x-axis should preserve O, and we take $-O = O$. Then we see $-R + R = -O = O$. Hence O should be the additive identity. Indeed, the "secant line" ℓ through O and any point P on C is just the vertical line through P. By symmetry, the other point of intersection of ℓ and C must be $-P$, so $O + P = -(-P) = P$. We also define $O + O = O$.

The only remaining cases in which to define addition are when $P = Q$ or the line ℓ is tangent to P or Q. In the former case, we take ℓ to be the tangent line through P, and $-R$ will be the unique other point on C intersecting ℓ (this point will be O if P is the point of intersection of C and the x-axis). In the latter case, the line ℓ will also not intersect C in 3 distinct points. If, say, ℓ is tangent to P, then we take $-R$ to be P. These rules make C (together with O) an abelian group. We call this group an *elliptic curve*.

A less haphazard way to think about this is in terms of projective geometry. The point at infinity O wraps up the curve C into a topological loop (as well as any vertical line). We call the resulting curve the *projective curve* or *projectivization* E of C. Bezout's theorem from algebraic geometry[20] tells us that, over $\mathbb{C}$, E intersects any line in exactly 3 points, counting multiplicity (suitably defined—e.g., $y = x^n$ intersects the x-axis with multiplicity n, and if a line is tangent to a curve at a point, the multiplicity is at least 2). Using this, we can define a group law on E by specifying that if P, Q, R are three (counting multiplicity) collinear points on E, then $P + Q + R = O$, i.e., $P + Q = -R$. Over $\mathbb{R}$ or $\mathbb{Q}$, even though Bezout's theorem fails, this construction still works. Namely, there are many lines which intersect the curve E in at most one point, but it is still true that any line which intersects E in *at least* 2 points, must intersect E in exactly 3, counting multiplicity.

What's important for us is that if P and Q are *rational points*, i.e., have rational x- and y-coordinates (or are the point at infinity), then the line ℓ through P and Q has rational (or infinite) slope. Then $P + Q$ is also rational (or O), and therefore you get a group structure just considering the rational points. (This does not work for integral points—if P and Q have integer coordinates, then at most we know $P+Q$ has rational coordinates.) This group will be called an *elliptic curve over* $\mathbb{Q}$.

From now on, denote the elliptic curve over $\mathbb{Q}$ associated to the equation (11) by $E_n = E_n(\mathbb{Q})$.[21] That is $E_n(\mathbb{Q})$ is the group structure on the set of rational solutions (x, y) to (11)

[20] This theorem says that two *complex projective* curves of degree m and n interesect in exactly mn points, counting multiplicity.

[21] If you're familiar with elliptic curves and are used to Weierstrass form, the curve E_n can, via the change

together with the point at infinity O.[22] Thus we can rephrase Question 3 as: when does $E_n(\mathbb{Q})$ have a nontrivial (not O) rational point?

We have the following general theorem about possible group structures of elliptic curves.

Theorem 3.2 (Mordell (1922); Mazur (1977))**.** *Let E be a elliptic curve over $\mathbb{Q}$. Then $E(\mathbb{Q}) \simeq \mathbb{Z}^r \oplus E(\mathbb{Q})_{\text{tors}}$, where $E(\mathbb{Q})_{\text{tors}}$ is a finite abelian group of order at most 16.*

The number r is called the *(algebraic) rank* of the elliptic curve, and the subgroup $E(\mathbb{Q})_{\text{tors}}$ is called the *torsion subgroup* of $E(\mathbb{Q})$. Mordell proved $E(\mathbb{Q})$ is a finitely generated abelian group and Mazur classified the possible torsion subgroups ($\mathbb{Z}/m\mathbb{Z}$ for $m \leq 10$ or $m = 12$ and $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2m\mathbb{Z})$ for $m \leq 4$). The 15 possible torsion subgroups in Mazur's theorem all occur for some (in fact, infinitely many) elliptic curves over $\mathbb{Q}$, but not for our special elliptic curves E_n:

Proposition 3.3 (cf. [DV09])**.** *If n is not a cube or twice a cube, then $E_n(\mathbb{Q})_{\text{tors}} = \{O\}$. For such n, (11) has a rational solution if and only if the rank of $E_n(\mathbb{Q})$ is at least one, i.e., if and only if there are infinitely many rational solutions.*

Contrast this to the case of integral points, where there are at most finitely many integer solutions to $x^3 + y^3 = n$. From now on, we will assume that n is neither a cube nor twice a cube. (In these cases, n is trivially a sum of two (integer) cubes.)

We remark that if we have one rational solution to (11), i.e., a rational point P ($\neq O$) on $E_n(\mathbb{Q})$ we can explicitly construct infinitely many solutions by taking multiples mP of P, i.e., $2P = P + P$, $3P = P + P + P$, ... Here the multiples mP must all be distinct since the torsion subgroup is trivial. The benefit of this proposition is that it may hard to detect if you have only a finite number of solutions, but if you have infinitely many solutions, it's probably easier to see them. Thus we can translate Question 3 into a special case of the general question:

Question 4. *Given an elliptic curve E over $\mathbb{Q}$, how can you determine if $E(\mathbb{Q})$ is finite or infinite?*

If you believe in miracles, you might hope for a local-global principle. What would this mean? Well, we can reduce the equation for E_n mod p and count solutions. Of course there won't be infinitely many solutions mod p, but maybe if there are a lot of solutions for many p, then there should be infinitely many points over $\mathbb{Q}$.

Assume $p \nmid 6n$. Then taking E_n mod p gives an elliptic curve mod p, i.e., an elliptic curver over the finite field $\mathbb{F}_p$. (Part of the general definition of elliptic curves is that they should be nonsingular, so if $p|n$, then E_n mod p is given by the equation $x^3 + y^3 = 0$, which has a singularity at the origin.)

Theorem 3.4 (Hasse)**.** *For $p \nmid 6n$ and $E = E_n$,*

$$p + 1 - 2\sqrt{p} < \#E(\mathbb{F}_p) < p + 1 + 2\sqrt{p},$$

of variables $(x, y) \mapsto (\frac{16n}{x+y}, 36n\frac{x-y}{x+y})$, be put in the *Weierstrass form* $y^2 = x^3 - 432n^2$ (e.g., [HW08, Section 25.3]).

[22] Note rational points on E_n don't correspond to rational points on C—however, the points with coordinates in $\mathbb{Q}(\sqrt{2})$ do correspond.

i.e., the quantity $a_p = p + 1 - \#E(\mathbb{F}_p)$ *satisfies*

$$|a_p| < 2\sqrt{p}.$$

It turns out that quantities a_p determine much of the arithmetic of elliptic curves. One of the brilliant insights of analytic number theory (based off work of people like Euler and Dirichlet) is that given some interesting arithmetic sequence (a_n) associated to some object X, one can often read deep arithmetic information about X off of the analytic properties of the Dirichlet series $L(s) = \sum \frac{a_n}{n^s}$. Here we only have defined a_n for (most) n prime, but we can use the idea of an Euler product to define an nice Dirichlet series associated to E.

Somewhat more precisely, one defines the *L-function* (or *L-series*) associated to an elliptic curve E by

$$L(s, E) = \sum_n \frac{a_n}{n^s} = \prod_p{}' \frac{1}{1 - a_p p^{-s} + p^{1-2s}}, \qquad \mathrm{Re}(s) > \frac{3}{2}. \tag{12}$$

Here, for $E = E_n$ the product on the right is over the factors written for $p \nmid 6n$ and certain simpler factors for $p|n$. Then one can expand the product into a series of the form $\sum \frac{b_n}{n^s}$ where $b_p = a_p$ and one defines the remaining a_n's by this identity of the series and the infinite product. Hasse's bound on a_p implies these expressions converge for $\mathrm{Re}(s) > \frac{3}{2}$.

The factors $1 - a_p p^{-s} + p^{1-2s}$ appearing in the denominators on the right might appear unnatural on first sight, but they arise as traces of certain operators on elliptic curves. That this is the right choice for the factor also comes out of the analogy with modular forms, where the L-function of a modular form $f(z) = \sum a_n q^n$ is $L(s, f) = \sum \frac{a_n}{n^s}$ and has an Euler product of the same form as the one for elliptic curves.

In our case of $E = E_n$, the L-function $L(s, E_n)$ is essentially a product of two L-functions of Dirichlet characters, and from this one gets the following analytic properties.

Theorem 3.5. *The function* $L(s, E)$ *extends to an entire function on* $\mathbb{C}$ *and satisfies a functional equation*

$$L(2 - s, E) = \gamma(s, E)L(s, E), \tag{13}$$

where $\gamma(s, E)$ *is an entire function that is never zero.*

For $E = E_n$, this follows from relating it to the L-function of a Dirichlet character as in, say, [IR90, Chapter 18]. (For the connection with [IR90], one first needs to transform E_n into Weierstrass form as in Footnote 21.

For general elliptic curves E, the L-function cannot be expressed in terms of L-functions of Dirichlet characters. (The elliptic curves for which one can do this are the *CM elliptic curves*, or the elliptic curves with *complex multiplication*.) In this case the above analytic properties are still true, but now require the use of the deep Modularity Theorem (i.e., the Shimura–Taniyama–Weil conjecture) which was proved by Wiles (1995), Taylor–Wiles (1995) and Breuil–Conrad–Diamond–Taylor (2001).

In retrospect, this provides other evidence that these funny factors $1 - a_p p^{-s} + p^{1-2s}$ (polynomials in p^{-s}) must be natural things to take in the Euler product (12). First, if you have an infinite product of functions, you have almost no chance of getting something that is entire (first of all, continues, and then has no poles—even the Riemann zeta has

a pole). Second, L-functions match up with either L-functions of Dirichlet characters, or L-functions of modular forms, where the a_n's in the Dirichlet series are the obvious sequence of numbers to associate with your character or modular form.

Now the L-function is, in some sense, easy to understand on the right half plane $\mathrm{Re}(s) > \frac{3}{2}$ by its Dirichlet series or Euler product. Consequently, by the functional equation (13), it is also easy to understand on the left half plane $\mathrm{Re}(s) < \frac{1}{2}$. What happens inside the critical strip $\frac{1}{2} < \mathrm{Re}(s) < \frac{3}{2}$ is more mysterious (the values are just known to exist and vary analytically by analytic continuation). Of special interest, is what happens along the center of the critical strip, the *critical line* $\mathrm{Re}(s) = 1$, and in particular at the *central value* $s = 1$.

This is addressed by the following famous conjecture (one of the 7 Clay Millennium Problems), based off of numerical calculations:

Conjecture 3.6 (Birch–Swinnerton-Dyer (1965?))**.** *$\#E(\mathbb{Q}) = \infty$ if and only if $L(1, E) = 0$. More generally, the rank of $E(\mathbb{Q})$ equals the order of vanishing of $L(s, E)$ at $s = 1$.*

Consequently, for $p > 2$, the BSD conjecture says that p should be a sum of two cubes if and only if $L(1, E_p) = 0$.

This order of vanishing in the BSD conjecture is called the *analytic rank* of E, and so the conjecture can be reformulated as saying the algebraic rank equals the analytic rank. There is also a more precise version of the conjecture that describes what the actual value of the first nonvanishing derivative (so the central value $L(1, E)$ in the rank 0 case) in terms of fundamental arithmetic invariants of E. This more precise version can be viewed as an analogue of Dirichlet's class number formula (Theorem 1.9).

We can think of the BSD conjecture as a local-global principle in the following way. The L-series $L(s, E)$ is defined purely in terms of local data, namely the a_p's (and some finite amount of other data which is simple, but I will not describe). So the order of vanishing of $L(1, E)$ is determined just by the a_p's. Recall that for $p \nmid 6n$ and $E = E_n$ is, a_p is just $p + 1$ minus the number of ways to write n as a sum of two cubes mod p. The conjecture then says that whether E_n has (infinitely many) rational points is determined by the a_p's (and in principle all of the a_p's are determined by n and suitable finite collections). However, this is a much more subtle local-global principle than what we saw for binary quadratic forms. This is because the central point $s = 1$ is past the range of convergence for the Euler product/Dirichlet series, so the connection between the a_p's and $L(1, E)$ is rather indirect.[23]

While there have been a lot of exciting breakthroughs about elliptic curves and number theory recently, I would venture that we're still a ways off from solving BSD. Nevertheless, there is a spectacular partial result:

Theorem 3.7 (Gross–Zagier (1986), Kolyvagin (1988))**.** *Suppose that the analytic rank of E is at most 1. Then the algebraic rank equals the analytic rank, i.e., BSD holds.*

There has also been very recent work such as Bhargava–Skinner–Zhang (2015) proving statistical results like BSD holds for a large percentage of elliptic curves. Their approach is to show that a large percentage (at least 66%) of elliptic curves have analytic rank ≤ 1.

[23] This is entirely analogous to the situation of the Riemann hypothesis (another Clay Millennium problem), where the problem is to understand the zeroes of $\zeta(s)$ inside the critical strip, where ζ is just defined by meromorphic continuation and the usual sum and product formulas do not make sense.

By this theorem, if we can prove $L(1, E_n) \neq 0$ (analytic rank 0), we know n is not a sum of two rational cubes,[24] Moreover, if we can show $L(1, E_n) = 0$ but $L'(1, E_n) \neq 0$ (analytic rank 1), we get that n is a sum of two rational cubes (in infinitely many ways). While it is known that elliptic curves in general can have large rank, conjecturally almost all (statistically 100%) of elliptic curves should have rank 0 or rank 1, with half having rank 0 and half having rank 1.

There is a simple sufficient criterion for concluding that $L(1, E) = 0$ (in which case the analytic rank should be 1 almost all the time, but one doesn't know this in general, so we can only say that conjecturally $\#E(\mathbb{Q}) = \infty$). This comes from our functional equation (13). Plugging $s = 1$ in the functional equation gives

$$L(1, E) = \epsilon(E)L(1, E), \qquad \text{where } \epsilon(E) = \gamma(1, E).$$

We call $\epsilon(E)$ the *root number* of E and it can be shown that $\epsilon(E) = \pm 1$. Furthermore, the root number can be determined from local conditions. If $\epsilon(E) = +1$, the above equality is just $L(1, E) = L(1, E)$ and we get no information, but if $\epsilon(E) = -1$, then we $L(1, E) = -L(1, E)$ implies $L(1, E) = 0$.

For $E = E_n$ when $n = p$, the local root numbers are not hard to determine and we have the following (cf. [DV09]):

Proposition 3.8. *For $n = p > 3$, we have*

$$\epsilon(E_p) = \begin{cases} 1 & p \equiv 1, 2, 5 \bmod 9 \\ -1 & p \equiv 4, 7, 8 \bmod 9. \end{cases}$$

Consequently, when $p \equiv 4, 7, 8 \bmod 9$, we have $L(1, E_p) = 0$ and, if BSD is true, then p is a sum of two rational cubes.

The statement that any $p \equiv 4, 7, 8 \bmod 9$ is a sum of two rational cubes seems to have originally been conjectured by Sylvester around 1847. (Contrast this with the case of sums of two integral cubes, where no prime $\equiv 4, 8 \bmod 9$ appears and asymptotically 0% of primes $\equiv 7 \bmod 9$ are sums of two integral cubes.) Elkies (1994) announced a proof of this conjecture in the cases $p \equiv 4, 7 \bmod 9$, but this has not been published (cf. [DV09]).

For general elliptic curves, there are no simple criteria to conclude that $L(1, E) \neq 0$ or not when $\epsilon(E) = +1$, however things are easier to analyze in the case of CM elliptic curves such as E_n. In fact, the following result (excluding the statement about $L(1, E_p)$) is classical.

Theorem 3.9 (Pépin, Lucas, Sylvester (1879?))**.** *If $p = 3$ or $p \equiv 2, 5 \bmod 9$, then p is not the sum of two rational cubes (and thus $L(1, E_p) \neq 0$).*

Apart from the case $p \equiv 8 \bmod 9$, where BSD and Sylvester conjecture that p is a sum of two rational cubes, the only remaining p is the case $p \equiv 1 \bmod 9$. Here the root number $\epsilon(E_p) = +1$, so a priori $L(1, E_p)$ could be 0 or not, but this case turns out to be subtler than the $p \equiv 2, 5 \bmod 9$ cases. Indeed, when $p \equiv 1 \bmod 9$, $L(1, E_p)$ is sometimes 0

[24] For our particular elliptic curves E_n, because they have CM, this does not actually require the full result of Gross–Zagier and Kolyvagin, but follows from an earlier theorem of Coates–Wiles (1977).

($p = 19, 37, 127, 163, 271, \ldots$) and sometimes not ($p = 73, 109, 181, 199, 307, \ldots$). This case is studied in [RVZ95], which gives several equivalent criteria for $L(1, E_p)$ to be 0 or not for $p \equiv 1 \bmod 9$, but they are rather involved. Here I will just state one of the more elementary criteria:

Theorem 3.10 ([RVZ95])**.** *Define polynomials $f_k(t)$ by $f_0(t) = 1$, $f_1(t) = t^2$ and*

$$f_{k+1}(t) = (1 - t^3)f_k'(t) + (2k + 1)t^2 f_k(t) - k^2 t f_{k-1}(t), \quad k \geq 1.$$

Let $A_k = f_{3k}(0)$ and $p \equiv 1 \bmod 9$. Then $L(1, E_p) = 0$ if and only if $p | A_{2(p-1)/9}$. Hence if $p \nmid A_{2(p-1)/9}$ implies p is not a sum of two rational cubes; otherwise, if BSD is true, then p is a sum of two rational cubes.

The first several polynomials f_k are

$$\begin{aligned}
f_0(t) &= 1, \\
f_1(t) &= t^2, \\
f_2(t) &= t^4 + t, \\
f_3(t) &= t^6 + 4t^3 + 1, \\
f_4(t) &= t^8 + 13t^5 + 10t^2, \\
f_5(t) &= t^{10} + 44t^7 + 71t^4 + 4t, \\
f_6(t) &= t^{12} + 161t^9 + 480t^6 + 74t^3 + 4.
\end{aligned}$$

We remark that if $3 \nmid k$, then $f_k(0) = 0$. The first several A_k's are

$$\begin{aligned}
A_1 &= 1, \\
A_2 &= 4, \\
A_3 &= 64, \\
A_4 &= 23104, \\
A_5 &= 1537600, \\
A_6 &= 46895104, \\
A_7 &= 386187673600, \\
A_8 &= 65663406063616.
\end{aligned}$$

Then, for instance, we see that $p = 19 | A_4 = 2^6 \cdot 19^2$ and $p = 37 | A_8 = 2^{14} \cdot 29^2 \cdot 37^2 \cdot 59^2$, so these primes should be sums of two cubes by BSD. In fact, they are sums of two integral cubes $19 = 3^2 + (-2)^3$ and $37 = 4^3 + (-3)^3$. (You may remember we already saw these were sums of two integral cubes because they are the values of $3x^2 + 3x + 1$ for $x = 2, 3$.)

We remark there are some other results known when n is not prime. For instance, Satgé (1987) showed that if $p \equiv 2 \bmod 9$, then $2p$ is a sum of two rational cubes. Also, Elkies (1994, unpublished) and [RVZ95] can treat the case of p^2 in their works. See [DV09] for a list of some similar results.

In summary, we saw a convoluted (conjectural) local-global principle (BSD) for solving $x^3 + y^3 = p$ over $\mathbb{Q}$, which happens to translate into a clean local-global statement (look

mod 9) in most cases ($p \not\equiv 1 \bmod 9$), but is rather complicated in the remaining case, and partial results are known without relying on BSD. In contrast, sums of two *integral* are too rare to fill out any congruence class of primes. So while the question of which numbers are sums of two integral cubes seems to be a closer analogue of the question "which numbers are sums of two (integral) squares" than the question for rational cubes, the answer for rational cubes is closer to the answer for two integral squares (Theorem 1.1). On the other hand, it is conjectured that n is a sum of *three integral cubes* if and only if $n \not\equiv 4, 5 \bmod 9$ (it is known that $n \not\equiv 4, 5 \bmod 9$ is a sum of four integral cubes, and conjectured that every n is a sum of four integral cubes [Coh07, Section 6.4.6]).

Well, that's the end, at least for now. I hope you enjoyed your adventure of number theory for the day. Good bye, and good luck on your future adventures!

References

[Coh07] Henri Cohen, *Number theory. Vol. I. Tools and Diophantine equations*, Graduate Texts in Mathematics, vol. 239, Springer, New York, 2007. MR2312337 (2008e:11001)

[DV09] Samit Dasgupta and John Voight, *Heegner points and Sylvester's conjecture*, Arithmetic geometry, 2009, pp. 91–102. MR2498056 (2010j:11088)

[Gro85] Emil Grosswald, *Representations of integers as sums of squares*, Springer-Verlag, New York, 1985. MR803155 (87g:11002)

[HW08] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Sixth edition, Oxford University Press, Oxford, 2008. Revised by D. R. Heath-Brown and J. H. Silverman, With a foreword by Andrew Wiles. MR2445243 (2009i:11001)

[IR90] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, Second edition, Graduate Texts in Mathematics, vol. 84, Springer-Verlag, New York, 1990. MR1070716 (92e:11001)

[KOK+86] M. Kneser, M. Ojanguren, M.-A. Knus, R. Parimala, and R. Sridharan, *Composition of quaternary quadratic forms*, Compositio Math. **60** (1986), no. 2, 133–150. MR868134 (88a:11037)

[Mara] Kimball Martin, *Modular Forms course notes (Spring 2011)*. http://www.math.ou.edu/~kmartin/mfs/.

[Marb] ———, *Number Theory II course notes (Spring 2010)*. http://www.math.ou.edu/~kmartin/ntii/.

[Mil96] Stephen C. Milne, *New infinite families of exact sums of squares formulas, Jacobi elliptic functions, and Ramanujan's tau function*, Proc. Nat. Acad. Sci. U.S.A. **93** (1996), no. 26, 15004–15008 (electronic). MR1629061 (99k:11162)

[RVZ95] Fernando Rodríguez Villegas and Don Zagier, *Which primes are sums of two cubes?*, Number theory (Halifax, NS, 1994), 1995, pp. 295–306. MR1353940 (96g:11049)

[Ser73] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York-Heidelberg, 1973. Translated from the French, Graduate Texts in Mathematics, No. 7. MR0344216 (49 #8956)

[Sil09] Joseph H. Silverman, *The arithmetic of elliptic curves*, Second edition, Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009. MR2514094 (2010i:11005)

[Sil93] ———, *Taxicabs and sums of two cubes*, Amer. Math. Monthly **100** (1993), no. 4, 331–340. MR1209462 (93m:11025)

[ST15] Joseph H. Silverman and John T. Tate, *Rational points on elliptic curves*, Second edition, Undergraduate Texts in Mathematics, Springer, Cham, 2015. MR3363545