RATIONAL AND INTEGRAL POINTS ON CURVES

Andrew Granville

Abstract. These are notes from a course given in Autumn of 2011 at the Université de Montréal. I have written things fairly completely up to about chapter 15. There is a lot more to be added on elliptic curves, and then on higher degree curves. Also stuff on integral points via Diophantine approximation. Please email me with any corrections or remarks (andrew@dms.umontreal.ca).

Table of Contents

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Preface

It is natural for a mathematician to ask for the solutions to a given equation. To understand the solutions in complex numbers may require ideas from analysis; to understand the solutions in integers requires ideas from number theory. The main difference is that, in the complex numbers, one can solve equations by using limits. For example to find a complex number x for which $x^2 = 2$ we need to find increasingly good approximations to a solution, in order to prove that $\sqrt{2}$, the limit of the approximations, exists. However to prove that there is no rational number p/q such that $(p/q)^2 = 2$ takes rather different techniques, and these are what we focus on in these notes.

Finding integral or rational solutions to a polynomial equation (or finite set of polynomial equations) is known as *Diophantine arithmetic* in honour of Diophantus,[1] a Greek mathematician who lived in Alexandria in the third century A.D.[2] His thirteen volume *Arithmetica* dealt with solutions to equations in integers and rationals,[3] which are now known as *Diophantine equations*.

Typically the subject of Diophantine equations is motivated by the quest to prove that there are no integer solutions to certain Diophantine equations. For example, Catalan's 1844 conjecture that the only integer solution to

$$x^p - y^q = 1$$

with $x, y \geq 1$ and $p, q \geq 1$ is $3^2 - 2^3 = 1$; which was proved by Preda Mihăilescu in 2002. Also Fermat's assertion from around 1637, known as "Fermat's Last Theorem",[4] that there are no integers $x, y, z \geq 1$ and $n \geq 3$ for which

$$x^p + y^p = z^p,$$

which was finally proved in 1995 by Sir Andrew Wiles.

Our goals will not be so elevated. We will mostly focus on the beautiful theory for trying to find rational solutions to Diophantine equations in just two variables, of degrees one, two and three. Degree one equations are more-or-less completely understood, but there are fundamental questions for degrees two and three that remain open, and indeed understanding degree three equations is perhaps the most researched problem in modern number theory.

We shall assume that the reader is familiar with a first course in number theory. When we require more advanced concepts we shall explain them.

---

[1] Bachet's 1621 Latin translation of Diophantus's *Arithmetica* helped spark the mathematical Renaissance.

[2] Many of the most famous ancient "Greek mathematicians", such as Diophantus, Eratosthenes, Euclid, Hero, Pappus, and Ptolemy, actually worked in Alexandria, Egypt.

[3] Though only parts of six of the volumes have survived.

[4] Indeed he wrote "it is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain" in his copy of *Arithmetica*. A copy of *Arithmetica* with Fermat's notes was published after his death, by his son in 1670.

1. Linear equations

1.1. Linear equations in two unknowns. The simplest Diophantine problem one encounters is to ask whether, given non-zero integers a, b and integer c, there are solutions in integers m, n to

$$am + bn = c \ . \tag{1.1}$$

One encounters this problem in a first course in number theory, and proves that there are such solutions if and only if $\gcd(a, b)$ divides c. We shall investigate this question again now, with a slightly different tilt to our discussion.

The first step in any Diophantine problem is to note that we may divide the coefficients by their greatest common divisor, and not change the solutions to the equation; in this case the coefficients are a, b, c and so, after dividing through by their gcd, we may assume that $\gcd(a, b, c) = 1$. In this case, our goal is to prove that (1.1) has integer solutions m, n if and only if $\gcd(a, b) = 1$:

Suppose that there exists integers m, n for which $am + bn = c$. Now $\gcd(a, b)$ divides both a and b, and so divides any linear combination of them, in particular $am + bn = c$. But then $\gcd(a, b)$ divides $\gcd(\gcd(a, b), c) = \gcd(a, b, c) = 1$.

Now suppose that $\gcd(a, b) = 1$. Since $\gcd(a, b) = 1$, therefore there exists an inverse to $b \pmod a$; that is an integer v for which $bv \equiv 1 \pmod a$ and hence there exists an integer u for which $au + bv = 1$. Then we have a solution to (1.1) with $m = cu, \ n = cv$

This proof establishes that there are solutions to the Diophantine equation $am+bn = 1$ when $\gcd(a, b) = 1$ without necessarily finding any solutions. Of course we can explicitly determine a solution using the Euclidean algorithm (which we discuss below).

Once one has found one solution to this Diophantine equation, then naturally one wishes to find all solutions. This is not difficult by "comparing solutions". To do this we start with a base solution (r, s) and write every other solution as $(r + x, s + y)$, so that

$$ax + by = (a(r + x) + b(s + y)) - (ar + bs) = c - c = 0.$$

This is another Diophantine problem so dividing through a and b by $\gcd(a, b)$ we obtain $Ax = -By$ where $A = a/g, \ B = b/g$ with $g := \gcd(a, b)$. Now B divides Ax and $(B, A) = 1$ so that B divides x. Writing $x = tB$ for some integer t, we have $y = -tA$ and hence all solutions to $am + bn = c$ are given by

$$m = r + t\ \frac{b}{(a, b)}, \quad n = s - t\ \frac{a}{(a, b)} \quad \text{where } t \text{ is an integer.}$$

There is another way to write the result for solvability of (1.1):

The Local-Global Principal for Linear Equations. *Let a, b, c be given integers, with $a, b \neq 0$ and $\gcd(a, b, c) = 1$. There are solutions in integers m, n to $am+bn = c$ if and only if for all prime powers p^e there exist residue classes $u, v \pmod{p^e}$ such that $au + bv \equiv c \pmod{p^e}$ and $\gcd(u, v, p) = 1$.*

This is called a *Local-Global Principal* because the solubility of the Diophantine equation over the integers, an infinite (global) ring, is equivalent to the solubility of the Diophantine equation over the integers mod p^e, each a finite (local) ring.

Proof. Suppose that $am + bn = c$, so that $am + bn \equiv c \pmod{p^e}$ for all prime powers p^e. Either this solution satisfies $(m, n, p) = 1$, or we can take $a(m-b)+b(n+a) \equiv c \pmod{p^e}$ in which case $(m - b, n + a, p) = 1$, else p divides $(a, b, c) = 1$ which is impossible.

On the other hand suppose that for each prime power p^e dividing b we have a solution u_p, v_p to $au_p + bv_p \equiv c \pmod{p^e}$. By the Chinese Remainder Theorem there exist residue classes $u, v \pmod{b}$ such that $u \equiv u_p,\ v \equiv v_p \pmod{p^e}$ for each $p^e \| b$, and so $au + bv \equiv c \pmod{b}$. Now given any integer $m \equiv u \pmod{b}$ we note that $am \equiv au \equiv c \pmod{b}$ and so there exists an integer n for which $am + bn = c$.

At first sight a big drawback of this re-formulation is that we are going to determine solubility over the integers, an infinite number of possibilities, by studying an infinite number of finite questions (i.e. modular arithmetic). However the proof makes it clear that we actually only need to study a finite number of cases, those prime powers dividing b. Hence we might have simply stated that there are solutions in integers m, n to $am+bn = c$ if and only if there exist residue classes $u, v \pmod{b}$ such that $au + bv \equiv c \pmod{b}$ with $\gcd(u, v, b) = 1$.

We will take a similar approach to quadratic equations.

1.2. The Euclidean Algorithm. To determine the greatest common divisor of two integers $a \geq b \geq 1$, we write $a = qb + r$ where q and r are integers with $0 \leq r \leq q - 1$.

`Exercise` 1.2.1. Prove that $\gcd(a, b) = \gcd(b, r)$.

Now since $r < b \leq a$, our new pair of integers (b, r) is smaller than the original pair (a, b) and so we can iterate this observation, until the smaller of the two numbers divides the larger. In that case we know that $\gcd(b, a) = b$. In this way we can determine the gcd of any two given positive integers. For example:

$$\begin{aligned}
85 &= 1 \cdot 48 + 37, \text{ so that } \gcd(85, 48) = \gcd(48, 37);\\
48 &= 1 \cdot 37 + 11, \text{ so that } \gcd(48, 37) = \gcd(37, 11);\\
37 &= 3 \cdot 11 + 4, \quad \text{so that } \gcd(37, 11) = \gcd(11, 4);\\
11 &= 2 \cdot 4 \ \ + 3, \quad \text{so that } \gcd(11, 4) \ \ = \gcd(4, 3);\\
4 &= 1 \cdot 3 \ \ + 1, \quad \text{so that } \gcd(4, 3) \quad = \gcd(3, 1) = 1.
\end{aligned}$$

We now wish to show how to find integers m, n such that $am+bn = \gcd(a, b)$, by iterating backwards in the above algorithm. We start with the observation that if $a = bq$ then $0 \cdot a + 1 \cdot b = b = \gcd(a, b)$. We also note that if $a = qb + r$ and $bu - rv = 1$ for some integers u and v. Then

$$\gcd(a, b) = \gcd(b, r) = bu - rv = bu - (a - qb)v = b(u + qv) - av,$$

the desired linear combination of a and b. For the above example we start with $1 = 0 \cdot 3 + 1 \cdot 1$

and together

$$\begin{aligned}
\text{with } \ 4 = 1\cdot 3 \ \ +1 \ &\implies 1 = 0\cdot 3 + 1\cdot(4 - 1\cdot 3) = 1\cdot 4 - 1\cdot 3;\\
\text{with } 11 = 2\cdot 4 \ \ +3 \ &\implies 1 = 1\cdot 4 - 1\cdot(11 - 2\cdot 4) = 3\cdot 4 - 1\cdot 11;\\
\text{with } 37 = 3\cdot 11 + 4 \ &\implies 1 = 3\cdot(37 - 3\cdot 11) - 1\cdot 11 = 3\cdot 37 - 10\cdot 11;\\
\text{with } 48 = 1\cdot 37 + 11 \ &\implies 1 = 3\cdot 37 - 10\cdot(48 - 1\cdot 37) = 13\cdot 37 - 10\cdot 48;\\
\text{with } 85 = 1\cdot 48 + 37 \ &\implies 1 = 13\cdot(85 - 1\cdot 48) - 10\cdot 48 = 13\cdot 85 - 23\cdot 48.
\end{aligned}$$

One useful consequence of the Euclidean algorithm is an understanding of the set of linear combinations of two integers

$$I(a,b) := \{am + bn : \ m, n \in \mathbb{Z}\}.$$

This is called the *ideal generated by a and b over $\mathbb{Z}$*. We observe that if $a = bq + r$ then $I(a,b) = I(b,r)$: for then $am + bn = (bq + r)m + bn = b(n + qm) + rm$ so that $I(a,b) \subset I(b,r)$, and $bu + rv = bu + (a - bq)v = av + b(u - qv)$ so that $I(b,r) \subset I(a,b)$. Following through the steps of the Euclidean algorithm we then have that

$$I(a,b) = I(g), \quad \text{where } \ g := \gcd(a,b).$$

One can view the step as showing that there is a 1-to-1 correspondence between the integer solutions to $am + bn = c$ and the integer solutions to $bu + rv = c$, via the invertible linear transformation

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} m \\ n \end{pmatrix} \iff \begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}.$$

The notion of ideal generalizes to $I(a_1, \dots a_k) := \{a_1m_1 + \dots + a_km_k : \ m_1, \dots, m_k \in \mathbb{Z}\}$ for any finite set of generators $a_1, \dots, a_k$, and this equals $I(g)$ where $g = \gcd(a_1, \dots a_k)$; this implies there exist integers $m_1, \dots, m_k$ such that $a_1m_1 + \dots + a_km_k = g$.

There are compact (and useful) ways to represent the Euclidean algorithm. We begin again with integers $a > b > 1$ but now for which $(a,b) = 1$. Again if $a = bq + r$ with $b > r \geq 1$ then

$$\frac{a}{b} = \ q + \frac{r}{b} = \ q + \frac{1}{\frac{b}{r}}\ .$$

We can repeat this to replace b/r by such a fraction, and iterate, until our fraction is an integer. We therefore obtain the *continued fraction*

$$\frac{a}{b} = \ a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \dots + \frac{1}{a_k}}}}\ ,$$

often written $[a_0,\ a_1,\ a_2,\ \dots,\ a_k]$ for convenience, where the a_i are the quotients from the various divisions. In our example,

$$\frac{85}{48} = 1 + \frac{37}{48},$$

and then

$$\frac{48}{37}= 1+ \frac{11}{37}, \text{ so that } \frac{85}{48}= 1+ \frac{1}{\frac{48}{37}}= 1+ \frac{1}{1+ \frac{11}{37}} .$$

Continuing like this:

$$\frac{85}{48}= 1+ \frac{1}{1+ \frac{11}{37}}= 1+ \frac{1}{1+ \frac{1}{3+\frac{4}{11}}}= 1+ \frac{1}{1+ \frac{1}{3+\frac{1}{2+\frac{3}{4}}}}= 1+ \frac{1}{1+ \frac{1}{3+\frac{1}{2+\frac{1}{1+\frac{1}{3}}}}} .$$

Hence $\frac{85}{48}= [1, 1, 3, 2, 1, 3]$, the quotients in the sequence of divisions above.

`Exercise 1.2.3.` Show that if $a_k > 1$ then $[a_0,\ a_1, \ldots,\ a_k] = [a_0,\ a_1, \ldots,\ a_k - 1, 1]$. Prove that the set of positive rational numbers are in 1-to-1 correspondence with the finite length continued fractions that do not end in 1.

The rationals that correspond to the first few numbers of the continued fraction, are $[1] = 1$, $[1, 1] = 2$, $[1, 1, 3] =\frac{7}{4}$, $[1, 1, 3, 2] =\frac{16}{9}$, $[1, 1, 3, 2, 1] =\frac{23}{13}$, which yield the approximations $1,\ 2,\ 1.75,\ 1.777\ldots,\ 1.7692\ldots$ to $85/48 = 1.770833\ldots$. We call these the *convergents* $p_j/q_j,\ j \geq 1$ for a continued fraction, defined by $\frac{p_j}{q_j}= [a_0,\ a_1,\ a_2,\ \ldots,\ a_j]$, so that $a/b = p_k/q_k$.

We can also write the transformation from the pair b, r to $a = qb + r, b$ in the matrix form

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix}.$$

Now, if we follow the Euclidean algorithm when $(a, b) = 1$, we have $q = a_0$, and then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b \\ r \end{pmatrix} = \ldots = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

`Exercise 1.2.4.` Prove by induction that

$$\begin{pmatrix} p_j & p_{j-1} \\ q_j & q_{j-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_j & 1 \\ 1 & 0 \end{pmatrix} .$$

Deduce that $a = p_k$ and $b = q_k$.

Taking determinants we have $p_j q_{j-1} - p_{j-1} q_j = (-1)^{j+1}$; and in particular

$$au + bv = 1, \quad \text{where } \ u = (-1)^{k-1} q_{k-1} \text{ and } v = (-1)^k p_{k-1},$$

which is a convenient reworking of the Euclidean algorithm. Bachet introduced this method to Renaissance mathematicians in the second edition of his brilliantly named book *Pleasant and delectable problems which are made from numbers* (1624). Such methods had been known to 8th century Indian scholars, probably to Archimedes, and possibly to the Babylonians.

1.3. Several simultaneous linear equations in several unknowns. We can represent such a problem as being given by an m-by-n matrix A, and an m-by-1 vector b, with integer entries, and we wish to find an n-by-1 vector x, with integer entries, for which

$$Ax = b.$$

If we proceed by Gaussian elimination on the augmented matrix $(A|b)$ then we end up with a matrix with all rational coefficients, which represents a set of linear equations that have exactly the same solutions as the original equations. We multiply each row through a rational number to ensure that the coefficients are all integers, and have no common divisor > 1. Re-labeling the variables if necessary this means that we have independent variables $x_1, \dots x_r$, and dependent variables $x_{r+1}, \dots, x_n$, such that there exist integers m_i, b_i, and $c_{i,j}$ for $j = 1, \dots, r$, for which our linear equations can be written as

$$\sum_{j=1}^{r} c_{i,j}x_j + m_i x_{r+i} = b_i$$

for $i = 1, 2, \dots, n - r := n'$. Hence we wish to find integer solutions $x_1, \dots x_r$ to

$$\sum_{j=1}^{r} c_{i,j}x_j \equiv b_i \pmod{m_i}, \quad \text{for } i = 1, 2, \dots, n'; \tag{1.2}$$

and then each x_{r+i} is determined (to be an integer). So solving several linear equations in several unknowns in integers is equivalent to solving several simultaneous congruences in several unknowns. We also note that if, for each prime p, we find solutions modulo the highest power of p dividing each m_i, then we can find solutions modulo the m_i by the Chinese Remainder Theorem; so we may restrict our attention to where the m_i are prime powers.

We now wish to solve (1.2) where each $m_i = p^{e_i}$ for given prime p, and $e_1 \geq \dots \geq e_{n'}$, and we may assume that $\gcd(p, b_i, c_{i,j}, 1 \leq j \leq r) = 1$. We may assume that $p \not| c_{i,j}$ for some $j, 1 \leq j \leq r$, for each i, else $p \not| b_i$ and there are no solutions. Re-numbering the x_j if necessary we may assume that $p \not| c_{1,1}$ and so, multiplying the first equation by the inverse of $c_{1,1} \pmod{p^{e_1}}$, we have evaluated $x_1 \pmod{p^{e_1}}$ in terms of the other x_j. This allows us to do Gaussian elimination on this variable and we iterate this process to completely solve the system of equations.

`Exercise 1.3.1.` Find all $x, y, z \in \mathbb{Z}$ for which $\begin{pmatrix} 3 & -5 & 1 & -1 & 2 \\ 4 & 6 & 4 & 4 & 8 \\ 5 & 1 & -3 & -5 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 6 \\ -20 \end{pmatrix}$.
(*Need an example which works through all the ideas in the above proof.*)

We have provided a suitable algorithm to find all solutions using Gaussian elimination. We ended up finding that there is a solution to $Ax = b$ if and only if there is a solution to a suitable set of linear congruences. In fact we now use this to prove:

The Local-Global Principal for Simultaneous Linear Equations. *We are given an m-by-n matrix A, and an m-by-1 vector b, with integer entries. There exists a vector x with integer entries, such that $Ax = b$ if and only if, for all prime powers $q = p^e$, there exist a vector x_q with integer entries, such that $Ax_q \equiv b \pmod q$.*

Proof. If $Ax = b$ then $Ax \equiv b \pmod q$ for all integers q. In the other direction, let us suppose that we have a solution to $Ax_m \equiv b \pmod m$, where m is chosen so that when we follow the Gaussian elimination algorithm above, this yields a solution to (1.2). By the argument there this then gives a solution to $Ax = b$.

1.4. Rational solutions and lattices. In section 1.1 we saw that the integer solutions to $ax + by = c$ where $(a, b) = 1$ are given by $(x, y) = (r, s) + t(b, -a),\ t \in \mathbb{Z}$ where (r, s) is an "initial solution". This can be viewed as the set of points where the line $ax + by = c$ meets the integer lattice $\mathbb{Z}^2 = \{(x, y) :\ x, y \in \mathbb{Z}\}$, and hence forms a (1-dimensional) sub-lattice.

How about rational solutions r, s to $ar + bs = c$ where a, b, c are non-zero and have no common divisor. If we re-write $r = x/z,\ s = y/z$ where z is the least common denominator of r and s, then we are asking for the integer solutions x, y, z to $ax + by = cz$ with $(x, y, z) = 1$ and $z > 0$. If we drop these last two conditions, for convenience, then we wish to find all integers x, y, z for which $ax + by = cz$.

In the language of linear algebra, we want all $(x, y, z) \in \mathbb{Z}^3$ in the null space of $(a, b, -c)$. The null space has dimension two and is obviously generated by $(0, c, b)$ and $(c, 0, a)$ but it is not obvious which linear combinations of these only have integer entries. It is easiest to use what we have done before: We know that $g := \gcd(a, b)|ax + by = cz$, and so $\gcd(a, b)|z$ as $\gcd(a, b, c) = 1$. Now any such z can be written as $z = gm$ for some integer m. If we again let $ar + bs = g$ then for a given $z = gm$ we have the solution $a(crm) + b(csm) = c(gm)$. Hence we have the vectors $m(cr, cs, g)$ in the null space. Now, if we subtract this from any solution x, y, z then $a(x - crm) + b(y - csm) = c(z - gm) = 0$. This is again easily parametrized as $n(-b/(a, b), a/(a, b), 0)$, and so our solutions are given by the 2-dimensional lattice

$$m(cr, cs, g) + n(-b/g, a/g, 0) \in \mathbb{Z}^3.$$

To recover the conditions $(x, y, z) = 1$ and $z > 0$, we ensure that $m > 0$ (as $g > 0$ by definition) and then divide through by the gcd of the entries of the vector.

If we rewrite $a = gA,\ b = gB$, so that $(A, B) = 1$ then the above becomes $m(cr, cs, g)+ n(-B, A, 0)$ where $rA + sB = 1$.

In the language of congruences we want to find all solutions x, y to $ax+by \equiv 0 \pmod c$ and then we can take $z = (ax + by)/c$ for any such solution. Now the above implies that the set of solutions is $(x, y) = mc(r, s) + n(-B, A)$. We now look at this from a different perspective:

`Exercise 1.4.1.` Prove directly that if $ax + by \equiv 0 \pmod c$ then there exists an integer N such that $x \equiv -BN \pmod c$ and $y \equiv AN \pmod c$. This can be re-written in the form $(x, y) = N(-B, A) + c(u, v)$ for arbitrary integers u, v. Establish that all of these solutions are given by the formula in the last paragraph.

What proportion of the lattice points of $\mathbb{Z}^2$ belong to our lattice $\langle c(r,s),(-B,A)\rangle$ of solutions to $ax+by\equiv 0 \pmod c$? This is evidently the reciprocal of the area in a "cell" of the lattice. Now a cell is formed by 0, the two vectors $c(r,s),(-B,A)$, and their sum, and its area is given by the determinant $\begin{vmatrix} cr & cs \\ -B & A \end{vmatrix} = c(Ar+Bs) = c$. That is $1/c$ of the pairs x,y satisfy $ax+by\equiv 0 \pmod c$, which is what one might guess by supposing that $ax+by$ is distributed like a random number mod c.

In this discussion we "homogenized" the linear equation $ax+by=c$ to $ax+by=cz$ to most easily deal with the rationals solutions. When we do this there are several inconvenient conditions on x,y,z – it is easiest to work with all rational solutions to $ax+by=cz$, knowing that in each "equivalence class" of ratios $x:y:z$ there are (at most) two with integer co-ordinates and no common divisor.

`Exercise 1.4.2.` Suppose that the non-zero solutions $x:y:z$ and $x':y':z'$ are equivalent if and only if there exists a rational number q such that $x'=qx,\ y'=qy,\ z'=qz$. Prove that there are at most two representatives of each equivalence class of rational solutions with x,y,z integers with $\gcd(x,y,z)=1$.

However we do have some extraneous solutions, those with $z=0$, since these do not correspond to a rational solution to the original equation $ax+by=c$, rather to "solutions" where x and y are both ∞, which we have long been taught not to do. However, rather than think of these solutions as extraneous we will say that they are *points at infinity* on the curve $ax+by=c$. The great advantage of this is that when we later manipulate more complicated equations, we will not have to keep track of special cases as we change the form of the equation.

The Frobenius postage stamp problem: If we only have postage stamps worth a cents and b cents where $(a,b)=1$, what amounts can we make up? That is, what is the set

$$\mathcal{P}(a,b) := \{am+bn:\ m,n\in\mathbb{Z},\ m,n\geq 0\}\ \ ?$$

(Note that in $\mathcal{P}(a,b)$ we only allow non-negative coefficients for a and b in our linear combinations, whereas in $I(a,b)$ there is no such restriction.) Suppose that r is an integer with $0\leq r\leq b-1$. If $N=am+bn\in\mathcal{P}(a,b)$ with $N\equiv ar \pmod{ab}$ then $am\equiv N\equiv ar \pmod n$ so that $m\equiv r \pmod b$ and hence $m=r+bk$ for some integer $k\geq 0$. Therefore $N=am+bn=ar+b(n+ak)$, and so the elements of $\mathcal{P}(a,b)$ in the arithmetic progression $ar \pmod b$ are all those elements of the arithmetic progression that are $\geq ar$. Hence $a(b-1)-b=ab-a-b$ is the largest integer that is not in $\mathcal{P}(a,b)$.

`Exercise 1.4.3.` Show that if $1\leq M,N\leq ab$ with $(M,ab)=1$ and $M+N=ab$ then exactly one of M and N is in $\mathcal{P}(a,b)$. (Hint: Given a representation of M, find one of N.)

Determining, in general, the largest integer that does not belong $\mathcal{P}(a,b,c)$, is an open problem.

2. Some Useful mathematics

2.1. Rational approximations to real numbers. We are interested in how close the integer multiples of a given real number α can get to an integer; that is, are there integers m, n such that $n\alpha - m$ is small? It is obvious that if $\alpha = p/q$ is rational then $n\alpha = m$ whenever $n = kq$ for some integer k, so that $m = kp$. How about irrational α?

Dirichlet's Theorem. *Suppose that α is a given real number. For every integer $N \geq 1$ there exists a positive integer $n \leq N$ such that*

$$|n\alpha - m| < \frac{1}{N},$$

for some integer m.

Proof. Let $\{t\}$ denote the fractional part of t, that is $\{t\} = t - [t]$, where $[t]$ is the largest integer $\leq t$. The numbers $0, \{\alpha\},\ \{2\alpha\}, \dots,\ \{N\alpha\}$ are all in $[0, 1)$. If we order them as $0 = \{i_0\alpha\} \leq \{i_1\alpha\} \leq \dots \leq \{i_N\alpha\} < 1$, then there exists j such that $\{i_{j+1}\alpha\} - \{i_j\alpha\} < 1/N$, else

$$1 > \{i_N\alpha\} - \{i_0\alpha\} = \sum_{j=0}^{N-1} \{i_{j+1}\alpha\} - \{i_j\alpha\} \geq N \cdot \frac{1}{N} = 1,$$

a contradiction. Now $0 < (i_{j+1} - i_j)\alpha - ([i_{j+1}\alpha] - [i_j\alpha]) = \{i_{j+1}\alpha\} - \{i_j\alpha\} < 1/N$ and so the result follows with $n = |i_{j+1} - i_j|$, and $m = \pm|[i_{j+1}\alpha] - [i_j\alpha]|$.

Corollary 2.1.1. *If α is a real irrational number then there are infinitely many pairs m, n of coprime positive integers for which*

$$\left|\alpha - \frac{m}{n}\right| < \frac{1}{n^2}\ .$$

Proof. If there are only finitely many solutions then there is one with $|n\alpha - m|$ minimal. Select integer $N > 1/|n\alpha - m|$ and now select a/b as in Dirichlet's Theorem. Then $b \leq N$ and $|b\alpha - a| < \frac{1}{N} < |n\alpha - m|$, so that

$$\left|\alpha - \frac{a}{b}\right| < \frac{1}{bN} \leq \frac{1}{b^2}\ .$$

For irrational α one might ask how the numbers $\{\alpha\},\ \{2\alpha\}, \dots,\ \{N\alpha\}$ are distributed in $[0, 1)$ as $N \to \infty$, for α irrational. There is a beautiful theory that shows that these values are dense and eventually equidistributed. This ideas tie in with the geometry of the torus and exponential sum theory, which go in a different direction from these notes.

We know that $\sqrt{d}$ is irrational if d is a non-square integer that is not the square of an integer. We can also show that there exist irrational numbers simply by how well they can be approximated by rationals:

Proposition 2.1.2. *Suppose that α is a given real number. If for every integer $q \geq 1$ there exist integers m, n such that*

$$0 < |n\alpha - m| < \frac{1}{q},$$

then α is irrational.

Proof. If α is rational then $\alpha = p/q$ for some coprime integers p, q with $q \geq 1$. For any integers m, n we then have $n\alpha - m = (np - mq)/q$. Now, the value of $np - mq$ is an integer $\equiv np \pmod{q}$. Hence $|np - mq| = 0$ or is an integer ≥ 1, and therefore $|n\alpha - m| = 0$ or is $\geq 1/q$.

There are several other methods to prove that numbers are irrational, but more challenging is to prove that a number is *transcendental*; that is, that it is not the root of a polynomial with integer coefficients (such a root is called *an algebraic number*).

Liouville's Theorem. *Suppose that α is the root of an irreducible polynomial $f(x) \in \mathbb{Z}[x]$ of degree ≥ 2. There exists a constant $c_\alpha > 0$ such that for any rational p/q with $(p, q) = 1$ and $q \geq 1$ we have*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c_\alpha}{q^d}.$$

Proof. Since $I := [\alpha - 1, \alpha + 1]$ is a closed interval, there exists a bound $B \geq 1$ such that $|f'(t)| \leq B$ for all $t \in I$. Let $c_\alpha = 1/B$. If $p/q \notin I$ then $|\alpha - p/q| \geq 1 \geq c_\alpha \geq c_\alpha/q^d$ as desired. Henceforth we may assume that $p/q \in I$.

If $f(x) = \sum_{i=0}^{d} f_i x^i$ then $q^d f(p/q) = \sum_{i=0}^{d} f_i p^i q^{d-i} \in \mathbb{Z}$. Now $f(p/q) \neq 0$ since f is irreducible of degree ≥ 2 and so $|q^d f(p/q)| \geq 1$.

The mean value theorem tells us that there exists t lying between α and p/q, and hence in I, such that

$$f'(t) = \frac{f(\alpha) - f(p/q)}{\alpha - p/q}.$$

Therefore

$$\left|\alpha - \frac{p}{q}\right| = \frac{|q^d f(p/q)|}{q^d |f'(t)|} \geq \frac{1}{Bq^d} = \frac{c_\alpha}{q^d}.$$

One usually first proves that there exist transcendental numbers by simply showing that the set of real numbers is uncountable, and the set of algebraic numbers is countable, so that the vast majority of real numbers are transcendental. However it is unsatisfying that this method yields that most real numbers are transcendental, without actually constructing any! As a consequence of Liouville's Theorem it is not difficult to construct transcendental numbers, for example

$$\alpha = \frac{1}{10} + \frac{1}{10^{2!}} + \frac{1}{10^{3!}} + \dots$$

since if p/q with $q = 10^{(n-1)!}$ is the sum of the first $n - 1$ terms then $0 < \alpha - p/q < 2/q^n$, and α cannot be an algebraic number by Liouville's Theorem.

Liouville's Theorem has been improved to its, more-or-less, final form:

Roth's Theorem. (1955) *Suppose that α is a real algebraic number. For any fixed $\epsilon > 0$ there exists a constant $c_{\alpha,\epsilon} > 0$ such that for any rational p/q with $(p,q) = 1$ and $q \geq 1$ we have*

$$\left|\alpha - \frac{p}{q}\right| \geq \frac{c_{\alpha,\epsilon}}{q^{2+\epsilon}} .$$

Evidently this cannot be improved much since, by Corollary 2.1.1, we know that if α is real, irrational then there are infinitely many p, q with $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$. In Corollary 2.3.2 we will show that all p/q for which $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{2q^2}$ can be easily identified from the continued fraction of α. Moreover we will see that if α is a quadratic, real irrational then there exists a constant $c_\alpha > 0$ such that $\left|\alpha - \frac{p}{q}\right| \geq \frac{c_\alpha}{q^2}$ for all p/q. The most amusing example is where $\alpha = \frac{1+\sqrt{5}}{2}$ and the best approximations are given by F_{n+1}/F_n where F_n is the nth Fibonacci numbers. One can show (in exercise 2.3.8) that

$$\left|\frac{1+\sqrt{5}}{2} - \frac{F_{n+1}}{F_n} + \frac{(-1)^n}{\sqrt{5}F_n^2}\right| \leq \frac{1}{2F_n^4} .$$

2.2. Irrationality of $\sqrt{d}$. We shall prove that there are real irrational numbers, for example $\sqrt{2}$

Proposition 2.2.1. *$\sqrt{2}$ is irrational.*

This is a consequence of

Proposition 2.2.2. *If d is an integer for which $\sqrt{d}$ is rational, then $\sqrt{d}$ is an integer. Therefore if integer d is not the square of an integer than $\sqrt{d}$ is irrational.*

Proof of Proposition 2.2.2 by integer factorization. We may write $\sqrt{d} = a/b$ where a and b are coprime positive integers, and $a^2 = db^2$. Write $a = \prod_p p^{a_p}$, $b = \prod_p p^{b_p}$, $d = \prod_p p^{d_p}$ where each $a_p, b_p, d_p \geq 0$, so that $2a_p = 2b_p + d_p$ for each prime p, as $a^2 = db^2$. Therefore if $b_p > 0$ or $d_p > 0$ then $a_p = b_p + d_p/2 > 0$, and so $b_p = 0$ as $(a,b) = 1$; but then $d_p = 2a_p$. Therefore $b = 1$ and $d = a^2$.

Proof of Proposition 2.2.2 by polynomial factorization. Now $\sqrt{d}$ is a root of the polynomial $x^2 - d$. If it equals a/b with $(a,b) = 1$ then $bx - a$ is a factor of $x^2 - d$. We have seen that $x^2 - d$ must factor in $\mathbb{Z}[x]$ if it factors in $\mathbb{Q}[x]$ but then we must have that $x^2 - d = (ux - v)(bx - a)$ for some integers a, b, u, v. Therefore $bu = 1$ and so $b = -1$ or 1, and therefore d is the square of a or $-a$.

Proof of Proposition 2.2.2 by approximation. We will see later that for any positive integer d that is not a square there are infinitely many solutions in positive integers m, n to Pell's equation $m^2 - dn^2 = 4$. Note that $m^2 = dn^2 + 4 > dn^2$ and so $m > \sqrt{d}n$, and therefore $m + \sqrt{d}n > 2\sqrt{d}n$. But then $4 = m^2 - dn^2 = (m - \sqrt{d}n)(m + \sqrt{d}n) > 2\sqrt{d}n(m - \sqrt{d}n)$. Dividing through by $2\sqrt{d}n$ we find that

$$0 < |\sqrt{d}n - m| < \frac{2}{\sqrt{d}n} .$$

Since the values of n get arbitrarily large this proves that $\sqrt{d}$ cannot be rational by Proposition 2.1.2.

Proof of Proposition 2.2.1 by 2-divisibility. : Let us recall that if $\sqrt{2}$ is rational then we can write it as a/b so that $a^2 = 2b^2$. Let us suppose that (b, a) give the smallest non-zero solutions to $y^2 = 2x^2$ in non-zero integers. Now 2 divides $2b^2 = a^2$ so that $2|a$. Writing $a = 2A$, thus $b^2 = 2A^2$, and so $2|b$. Writing $b = 2B$ we obtain a solution $A^2 = 2B^2$ where A and B are half the size of a and b, contradicting minimality.

`Exercise 2.2.1.` Now prove Proposition 2.2.2 analogously. (Hint: Start by writing $d = Dm^2$ where D is squarefree. Then try the same argument using each prime dividing D.)

`Exercise 2.2.2.` Show that there are no non-zero integer solutions to $x^3 + 3y^3 + 9z^3 = 0$.

Proof of Proposition 2.2.1 by a different algebraic descent. As $a^2 = 2b^2$, so $a > b > a/2$. Now

$$(2b - a)^2 = a^2 - 4ab + 2b^2 + 2b^2 = a^2 - 4ab + 2b^2 + a^2 = 2(a - b)^2.$$

However $0 < 2b - a < a$ contradicting the minimality of a.

Proof of Proposition 2.2.1 by a geometric version of the last argument. [5] Again we may assume that $\sqrt{2} = a/b$ with a and b positive integers, where a is minimal. Hence $a^2 = 2b^2$ which gives rise to the smallest right-angle, isosceles triangle, OPQ with integer side lengths $\overline{OP} = \overline{OQ} = b$, $\overline{PQ} = a$ and angles $P\hat{O}Q = 90^o$, $P\hat{Q}O = Q\hat{P}O = 45^o$. Now mark a point R which is b units along PQ from Q and then drop a perpendicular to meet OP at the point S. Now $R\hat{P}S = Q\hat{P}O = 45^o$, and so $R\hat{S}P = 180^o - 90^o - 45^o = 45^o$ by considering the angles in the triangle RSP, and therefore this is a smaller isosceles, right-angled triangle. This implies that $\overline{RS} = \overline{PR} = a - b$. Now two sides and an angle are the same in OQS and RQS so these triangles are congruent; in particular $\overline{OS} = \overline{SR} = a - b$ and therefore $\overline{PS} = \overline{OP} - \overline{OS} = b - (a - b) = 2b - a$. Hence RSP is a smaller isosceles, right-angled triangle than OPQ with integer side lengths, giving a contradiction.

Proof of Proposition 2.2.2 by an analogous algebraic descent. Suppose that a is the smallest integer for which $\sqrt{d} = a/b$ with a and b positive integers. Let r be the smallest integer $\geq db/a$, so that $\frac{db}{a} + 1 > r \geq \frac{db}{a}$, and therefore $a > ra - db \geq 0$. Then

$$\begin{aligned}(ra - db)^2 &= da^2 - 2rdab + d^2b^2 + (r^2 - d)a^2 \\ &= da^2 - 2rdab + d^2b^2 + (r^2 - d)db^2 = d(rb - a)^2\end{aligned}$$

However $0 \leq ra - db < a$ contradicting the minimality of a, unless $ra - db = 0$. In this case $r^2 = d \cdot db^2/a^2 = d$.

2.3. Continued fractions for real numbers. One can define the continued fraction for any real number $\alpha = \alpha_0$: Let $a_0 := [\alpha_0]$. If $\alpha_0 - a_0 = 0$, that is α_0 is an integer we stop; otherwise we repeat the process with $\alpha_1 = 1/(\alpha_0 - a_0)$. This yields a unique continued

[5] Needs diagram

fraction for each real number α. In fact $\alpha_j - a_j = \alpha_j - [\alpha_j] = \{\alpha_j\} \in [0,1)$, so that each $\alpha_{j+1} \geq 1$ for all $j \geq 0$. Hence a_j is a positive integer for each $j \geq 1$.

`Exercise 2.3.1.` Prove that if α has a finite length continued fraction then the last term is an integer ≥ 2.

To determine the value of $[a_0,\ a_1,\ a_2,\ \ldots,]$, where the integer $a_0 \geq 0$ and each other integer $a_i \geq 1$, we define the *convergents* $p_n/q_n := [a_0,\ a_1,\ \ldots,\ a_n]$ for each $n \geq 0$ as above by

$$\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}.$$

Note that $p_n = a_n p_{n-1} + p_{n-2}$ and $q_n = a_n q_{n-1} + q_{n-2}$ for all $n \geq 2$, so that the sequences $p_1, p_2, \ldots$ and $q_1, q_2, \ldots$ are increasing. Taking determinants yields that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n+1}}{q_{n-1}q_n} \tag{2.3.1}$$

for each $n \geq 1$.

`Exercise 2.3.2.` Deduce that

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \ldots < \frac{p_{2j}}{q_{2j}} < \ldots\ldots < \frac{p_{2j+1}}{q_{2j+1}} < \ldots < \frac{p_3}{q_3} < \frac{p_1}{q_1};$$

and that p_n/q_n tends to a limit as $n \to \infty$.

We have proved that if n is finite then the value given by the continued fraction is indeed α, but this is not so obvious if n is infinite (i.e. α is irrational). We now prove this, and as a consequence one can deduce that the positive real numbers are in 1-to-1 correspondence with the continued fractions.

`Exercise 2.3.4.` Show that if a, b, A, B, u, v are positive reals then $\frac{au+Av}{bu+Bv}$ lies between $\frac{a}{b}$ and $\frac{A}{B}$.

Now $\alpha = [a_0,\ a_1,\ a_2,\ \ldots,\ a_n,\ \alpha_{n+1}]$, so that $\alpha = R/S$ where

$$\begin{pmatrix} R & p_n \\ S & q_n \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} \begin{pmatrix} \alpha_{n+1} & 1 \\ 1 & 0 \end{pmatrix},$$

and hence

$$\alpha = \frac{R}{S} = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} \tag{2.3.2}$$

lies between $\frac{p_{n-1}}{q_{n-1}}$ and $\frac{p_n}{q_n}$ for each $n \geq 1$, by the previous exercise.

`Exercise 2.3.5.` Deduce that $\alpha = \lim_{n\to\infty} p_n/q_n$.

`Exercise 2.3.6.` Also deduce that $\left|\alpha - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}}$ for all $n \geq 0$.

Now $\pi = [3, 7, 15, 1, 292, 1, \ldots]$ which leads to the convergents

$$3 < \frac{333}{106} < \ldots < \pi < \ldots < \frac{355}{113} < \frac{22}{7}\ .$$

Archimedes knew that $|\pi - \frac{355}{113}| < 3 \cdot 10^{-7}$.[6] The continued fraction for e displays an interesting pattern: $e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \dots]$. One can generalize the notion of continued fractions to obtain

$$\frac{\pi}{4} = \frac{1}{1+\frac{1^2}{2+\frac{3^2}{2+\frac{5^2}{2+\dots}}}} \qquad \text{or} \qquad \pi = \frac{4}{1+\frac{1^2}{3+\frac{2^2}{5+\frac{3^2}{7+\dots}}}}.$$

By (2.3.2) and then (2.3.1) we have

$$\alpha - \frac{p_n}{q_n} = \frac{\alpha_{n+1}p_n + p_{n-1}}{\alpha_{n+1}q_n + q_{n-1}} - \frac{p_n}{q_n} = \frac{(-1)^n}{q_n(\alpha_{n+1}q_n + q_{n-1})}.$$

Now $a_{n+1} \le \alpha_{n+1} < a_{n+1} + 1$ and so $q_{n+1} \le \alpha_{n+1}q_n + q_{n-1} < q_{n+1} + q_n < 2q_{n+1}$, yielding

$$\frac{1}{2q_nq_{n+1}} < \left|\alpha - \frac{p_n}{q_n}\right| \le \frac{1}{q_nq_{n+1}}. \tag{2.3.4}$$

This is a good approximation, but are there better? Lagrange showed that there are not:

Theorem 2.3.1. *If $1 \le q < q_{n+1}$ then $|q_n\alpha - p_n| \le |q\alpha - p|$.*

Proof. Let $x = (-1)^n(pq_{n+1} - qp_{n+1})$ and $y = (-1)^n(pq_n - qp_n)$, so that $p_nx - p_{n+1}y = p$ and $q_nx - q_{n+1}y = q$ as $p_nq_{n+1} - q_np_{n+1} = (-1)^n$. We observe that $x \ne 0$ else q_{n+1} divides $q_{n+1}y = -q$ so that $q_{n+1} \le q$ contradicting the hypothesis.

Now $q_nx = q_{n+1}y + q$ where $q < q_{n+1} \le |q_{n+1}y|$ if $y \ne 0$, and so q_nx and $q_{n+1}y$ have the same sign, and therefore x and y have the same sign. We saw earlier that $q_n\alpha - p_n$ and $q_{n+1}\alpha - p_{n+1}$ have opposite signs, and so $x(q_n\alpha - p_n)$ and $y(q_{n+1}\alpha - p_{n+1})$ have opposite signs. Now $q\alpha - p = x(q_n\alpha - p_n) - y(q_{n+1}\alpha - p_{n+1})$ and so

$$|q\alpha - a| = |x(q_n\alpha - p_n)| + |y(q_{n+1}\alpha - p_{n+1})| \ge |q_n\alpha - p_n|,$$

with equality implying that $|x| = 1$ and $y = 0$, so that $\{p, q\} = \{p_n, q_n\}$.

`Exercise 2.3.7.` Deduce that if $1 \le q < q_n$ then $\left|\alpha - \frac{p_n}{q_n}\right| < \left|\alpha - \frac{p}{q}\right|$.

Corollary 2.3.2. *If $\left|\alpha - \frac{p}{q}\right| < \frac{1}{2q^2}$ then $\frac{p}{q}$ is a convergent for α.*

Proof. If $q_n \le q < q_{n+1}$ then $|q_n\alpha - p_n| < 1/2q$ by Theorem 2.3.1. Hence $p/q = p_n/q_n$ else

$$\frac{1}{qq_n} \le \left|\frac{p}{q} - \frac{p_n}{q_n}\right| \le \left|\alpha - \frac{p}{q}\right| + \left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{2q^2} + \frac{1}{2qq_n},$$

which is impossible.

---

[6] Around 1650 B.C., ancient Egyptians approximated π by regular octagons obtaining 256/81, a method developed further by Archimedes in the third century B.C. and Liu Hui in China in the third century A.D. In 1168 B.C. the Talmudic scholar Maimonides asserted that π can only be known approximately, perhaps a claim that it is irrational. In the ninth century B.C. the Indian astronomer Yajnavalkya arguably gave the approximation 333/106 in *Shatapatha Brahmana*; in the 14th century A.D., Madhava of the Kerala school in India indicated how to get arbitrarily good approximations to π.

Lemma 2.3.4. *The inequality* $\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{2q^2}$ *is satisfied for at least one of* $p/q = p_n/q_n$ *or* p_{n+1}/q_{n+1} *for each* $n \geq 0$.

Proof. If not then, since $\alpha - \frac{p_n}{q_n}$ and $\alpha - \frac{p_{n+1}}{q_{n+1}}$ have opposite signs, hence

$$\frac{1}{q_n q_{n+1}} = \left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| = \left|\alpha - \frac{p_n}{q_n}\right| + \left|\alpha - \frac{p_{n+1}}{q_{n+1}}\right| > \frac{1}{2q_n^2} + \frac{1}{2q_{n+1}^2},$$

which is false for any positive reals q_n, q_{n+1}.

Hurwitz showed that for at least one of every three convergents one can improve this to $\leq 1/(\sqrt{5}q^2)$ and that this is best possible, since this is the constant for the golden ratio $\frac{1+\sqrt{5}}{2}$:

`Exercise 2.3.8.` Show that $\frac{1+\sqrt{5}}{2} = [1, 1, 1, 1, \dots]$ and so the convergents are F_{n+1}/F_n where F_n is the nth Fibonacci numbers. By using the general formula for Fibonacci numbers, determine how good these approximations are; i.e. prove a strong version of the formula at the end of chapter 11:

$$\left|\frac{1+\sqrt{5}}{2} - \frac{F_{n+1}}{F_n} + \frac{(-1)^n}{\sqrt{5}F_n^2}\right| \leq \frac{1}{2F_n^4}\,.$$

2.4. Solving the cubic. Pre-cursors to Galois theory. The roots of a quadratic polynomial $ax^2 + bx + c = 0$ are

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\,.$$

The easy way to prove this is to put the equation into a form that is easy to solve. One begins by dividing through by a, to get $x^2 + (b/a)x + c/a = 0$, so that the leading coefficient is 1. Next we make a change variable, letting $y = x + b/2a$ to obtain

$$y^2 - (b^2 - 4ac)/4a^2 = 0.$$

Having removed the y^1 term, we can simply take square-roots to obtain the possibilities for y, and hence the possible values for x.

We call $\Delta := b^2 - 4ac$ the *discriminant* of our polynomial. Note that if $f(x) = ax^2 + bx + c$ then $f'(x) = 2ax + b$. We apply the Euclidean algorithm on these two polynomials: $2f(x) - xf'(x) = bx + 2c$ and so $2a(bx + 2c) - b(2ax + b) = -\Delta$, which yields

$$\Delta = -4a(ax^2 + bx + c) + (2ax + b)^2.$$

Thus Δ is the smallest positive integer in the ideal generated by f and f' over $\mathbb{Z}[a, b, x]$.

Can one similarly find the roots of a cubic? We can certainly begin the same way.

`Exercise 2.4.1.` Show that we can easily deduce the roots of any given cubic polynomial, from the roots of some cubic polynomial of the form $x^3 + ax + b$.

We wish to find the roots of $x^3 + ax + b = 0$. This does not look so easy since we cannot simply take cube roots unless $a = 0$. Cardano's trick (1545) is a little surprising: Write $x = u + v$ so that

$$x^3 + ax + b = (u + v)^3 + a(u + v) + b = (u^3 + v^3 + b) + (u + v)(3uv + a).$$

This equals 0 when $u^3 + v^3 = -b$ and $3uv = -a$; in other words

$$u^3 + v^3 = -b \text{ and } u^3v^3 = -a^3/27. \tag{2.4.1}$$

Hence $(X - u^3)(X - v^3) = X^2 + bX - a^3/27$ and so, using the formula for the roots of a quadratic polynomial, yields

$$u^3, v^3 = \frac{-b \pm \sqrt{b^2 + 4a^3/27}}{2}$$

Taking cube roots yield values for u and v for which (2.4.1) holds but it is not clear that $3uv = -a$. Indeed what we do have is that if $\alpha = -3uv/a$ then $\alpha^3 = -27u^3v^3/a^3 = 1$ by (2.4.1), and so α is one of the three cube roots of unity, and not necessarily 1. To rectify this we replace v by $\alpha^2 v$. Hence the roots of $x^3 + ax + b$ are given by

$$u + v,\ \omega u + \omega^2 v,\ \omega^2 u + \omega v,$$

where ω is a primitive cube root of unity. Now the discriminant is

$$\Delta := 4a^3 + 27b^2 = (6ax^2 - 9bx + 4a^2)(3x^2 + a) - 9(2ax - 3b)f(x),$$

where $f(x) = x^3 + ax + b$, the smallest positive integer in the ideal generated by f and f' over $\mathbb{Z}[a, b, x]$, and so $u^3, v^3 = \frac{-b \pm \sqrt{\Delta/27}}{2}$.

The important thing to notice here is that the solution to a cubic is given in terms of both cube roots and square roots, not just cube roots.

How about the roots of a quartic polynomial? Can these be found in terms of fourth roots, cube roots and square roots? And similarly roots of quintics and higher degree polynomials?

The general quartic: We begin, as above, by rewriting the equation in the form $x^4 + ax^2 + bx + c = 0$. Following Ferrari (1550s) we add an extra variable y to obtain the equation

$$(x^2 + a + y)^2 = (a + 2y)x^2 - bx + ((a + y)^2 - c),$$

and then select y to make the right side the square of a linear polynomial in x (and so we would have $(x^2 + a + y)^2 = (rx + s)^2$ and hence x can be deduced as a root of one of the quadratic polynomials $(x^2 + a + y) \pm (rx + s)$). The right side is a square of a linear polynomial if and only if its discriminant is 0, that is $b^2 - 4(a + 2y)((a + y)^2 - c) = \Delta = 0$. But this is a cubic in y, and we have just seen how to find the roots of a cubic polynomial.

Example: We want the roots of $X^4 + 4X^3 - 37X^2 - 100X + 300$. Letting $x = X + 1$ yields $x^4 - 43x^2 - 18x + 360$. Proceeding as above leads to the cubic equation $2y^3 - 215y^2 + 6676y - 64108 = 0$. Dividing through by 2 and then changing variable $y = t + 215/6$ gives the cubic $t^3 - (6169/12)t - (482147/108) = 0$. This has discriminant $-4(6169/12)^3 + 27(482147/108)^2 = -(2310)^2$. Hence $u^3, v^3 = (482147 \pm 27720\sqrt{-3})/216$. Unusually this has an exact cube root in terms of $\sqrt{-3}$; that is $u, v = \omega^*(-37 \pm 40\sqrt{-3})/6$. Now $-3(-37 + 40\sqrt{-3})/6 \cdot (-37 - 40\sqrt{-3})/6 = -6169/12 = a$. Therefore we can take $u, v = (-37 \pm 40\sqrt{-3})/6$, and the roots of our cubic are $t = u + v = -37/3$, $\omega u + \omega^2 v = 157/6$, $\omega^2 u + \omega v = -83/6$ so that $y = 47/2, 62, 22$. From these Ferrari's equation becomes $(x^2 - 39/2) = \pm(2x + 9/2)$ for $y = 47/2$ and so the possible roots $-5, 3; -4, 6$; or $(x^2 + 19) = \pm(9x + 1)$ for $y = 62$ and so the possible roots $-5, -4; 3, 6$; or $(x^2 - 21) = \pm(x + 9)$ for $y = 22$ and so the possible roots $-5, 6; 3, -4$. For each such y we get the same roots $x = 3, -4, -5, 6$, yielding the roots $X = 2, -5, -6, 5$ of the original quartic.

Example: Another fun example is to find the fifth roots of unity. That is those x satisfying $\phi_5(x) = \frac{x^5-1}{x-1} = x^4 + x^3 + x^2 + x + 1 = 0$. Proceeding as above we find the roots

$$\frac{\sqrt{5} - 1 \pm \sqrt{-2\sqrt{5} - 10}}{4}, \quad \frac{-\sqrt{5} - 1 \pm \sqrt{2\sqrt{5} - 10}}{4} \quad .$$

Example: Gauss's favourite example was the expression in surds was for $\cos \frac{2\pi}{2^k}$, which we will denote by $c(k)$. A double angle formula states that $\cos 2\theta = 2\cos^2 \theta - 1$, and so taking $\theta = 2\pi/2^k$ we have $c(k-1) = 2c(k)^2 - 1$, or $c(k) = \frac{1}{2}\sqrt{2 + 2c(k-1)}$. Note that $c(k) \geq 0$ for $k \geq 2$ and $c(2) = 0$. Hence $c(3) = \frac{1}{2}\sqrt{2}$, $c(4) = \frac{1}{2}\sqrt{2 + \sqrt{2}}$, $c(5) = \frac{1}{2}\sqrt{2 + \sqrt{2 + \sqrt{2}}}$ and, in general,

$$\cos\left(\frac{2\pi}{2^k}\right) = \frac{1}{2}\underbrace{\sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \ldots\sqrt{2}}}}}}_{k-2 \text{ times}} \qquad \text{for each } \; k \geq 3.$$

Why surds? (A *surd* is a number of the form $n^{1/k}$ where n and k are positive integers.) Why do we wish to express roots of all polynomials in terms of square roots, cube roots, etc.? That is, surds. After all, is a solution in terms of $\sqrt{7}$ any more enlightening then in terms of the second-largest root of $x^5 - 3x^2 + 2x - 11$? By this I mean we have an expression that gives each of these numbers "exactly", though that expression is not something that is a solid integer, just something that one can approximate speedily and accurately. But there are methods to quickly approximate the roots of any given polynomial to any desired level of accuracy, so why the obsession with surds? The answer is more aesthetic than anything else – we have a comfort level with surds that we do not have with more complicated expressions. One can rephrase the question: *Can we describe the roots of any given polynomial, $x^d + a_1 x^{d-1} + \ldots + a_d$ as a polynomial, with rational coefficients, in roots of polynomials of the form $x^k - n$?* One can see that this is probably wishful thinking since we wish to express a root that is given in terms of d coefficients, in terms of something

much simpler, and it is perhaps miraculous that we have succeeded with all polynomials where $d \leq 4$. After this discussion it may not come as such a surprise that there are degree five polynomials whose roots cannot be expressed as a rational expression in surds. This was understood by Gauss in 1804, but waited for a magnificent proof by Galois in 1829 at the age of 18. More on that in a moment.

The theory of symmetric polynomials. It is difficult to work with algebraic numbers since one cannot necessarily evaluate them precisely. However for many of the reasons we use them we do not need to actually work with complex numbers, but rather we work with the set of roots of a polynomial. It was Sir Isaac Newton who recognized the following result. We say that $P(x_1, x_2, \dots, x_n)$ is a *symmetric polynomial* if $P(x_k, x_2, \dots, x_{k-1}, x_1, x_{k+1}, \dots, x_n) = P(x_1, x_2, \dots, x_n)$ for each k.

`Exercise 2.4.2.` Show that for any permutation σ if $1, 2, \dots, n$ and any symmetric polynomial P we have $P(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)}) = P(x_1, x_2, \dots, x_n)$.

The fundamental theorem of symmetric polynomials. *Let $f(x) = \prod_{i=1}^{d}(x - \alpha_i) = \sum_{i=0}^{d} a_i x^i$ be a monic polynomial with integer coefficients. Then any symmetric polynomial in the roots of f can be expressed as a polynomial in the a_i.*

Let us see the idea. First, we know by multiplying out $\prod_{i=1}^{d}(x - \alpha_i)$ that

$$\sum_i \alpha_i = -a_1, \ \sum_{i<j} \alpha_i\alpha_j = a_2, \ \sum_{i<j<k} \alpha_i\alpha_j\alpha_k = -a_3, \ \dots, \ \alpha_1\alpha_2\dots\alpha_n = \pm a_n.$$

Now define $s_k := \sum_{i=1}^{d} \alpha_i^k$. Since $\frac{f'(x)}{f(x)} = \sum_{i=1}^{d} \frac{1}{x-\alpha_i}$ we have

$$\frac{\sum_{j=0}^{d} j a_j x^{d-j}}{\sum_{i=0}^{d} a_i x^{d-i}} = \frac{x^{d-1}}{x^{d-1}} \cdot \frac{f'(1/x)}{x f(1/x)} = \sum_{i=1}^{d} \frac{1}{1-\alpha_i x} = \sum_{i=1}^{d} \sum_{k\geq 0} (\alpha_i x)^k = \sum_{k\geq 0} s_k x^k.$$

This implies that $\sum_{j=0}^{d}(d-j)a_{d-j}x^j = \sum_{i=0}^{d} a_{d-i}x^i \cdot \sum_{k\geq 0} s_k x^k$, so that, comparing the coefficients of x^k, we obtain (as $a_d = 1$)

$$s_k = -\sum_{i=1}^{\min\{d,k\}} a_{d-i}s_{k-i} + \begin{cases} (d-k)a_{d-k} & \text{if } k < d; \\ 0 & \text{if } k \geq d. \end{cases}$$

Hence, by induction on k, we see that the s_k are polynomials in the a_j.

`Exercise 2.4.3.` If f is not monic, develop analogous results by working with $g(x)$ defined by $g(a_d x) = a_d^{d-1} f(x)$.

Now that we have obtained the s_k, we can prove Newton's result for arbitrary symmetric polynomials, by showing that every symmetric polynomial is a polynomial in the s_k, which implies the theorem. We proceed by induction on the number of variables in the monomials of the symmetric polynomial. The result for the s_k is precisely the case where each monomial has one variable. Now, for the proof by induction, suppose that

the symmetric polynomial under question has monomial $\alpha_{i_1}^{k_1}\alpha_{i_2}^{k_2}\dots\alpha_{i_r}^{k_r}$ with each $k_i \geq 1$ and summed over all possibilities of $i_1, i_2, \dots, i_r$ being distinct elements of $1, 2, \dots, n$. We subtract $s_{k_1}s_{k_2}\dots s_{k_r}$,[7] and we are left with various cross terms, in which two or more of the variables α_j are equal. Hence in the remaining expression each monomial contains fewer variables and the result follows by induction.

Example: Look at $\sum_{i,j,k}\alpha_i\alpha_j^2\alpha_k^3$. Subtract $s_1s_2s_3$ and we have to account the cases where $i=j$ or $i=k$ or $j=k$. Hence what remains is $-\sum_{i,k}\alpha_i^3\alpha_k^3 - \sum_{i,j}\alpha_i^4\alpha_j^2 - \sum_{i,j}\alpha_i\alpha_j^5 + 2s_6$ where in the first sum we have $i=j$, in the second $i=k$, then $j=k$ and $i=j=k$. Proceeding the same way again we have $\sum_{i,j}\alpha_i^4\alpha_j^2 = s_4s_2 - s_6$, $\sum_{i,j}\alpha_i\alpha_j^5 = s_1s_5 - s_6$ and $\sum_{i,k}\alpha_i^3\alpha_k^3 = (s_3^2 - s_6)/2$, the last since in s_3^2 the cross term $\alpha_i^3\alpha_k^3$ appears also as $\alpha_k^3\alpha_i^3$. Collecting this all together yields $\sum_{i,j,k}\alpha_i\alpha_j^2\alpha_k^3 = s_1s_2s_3 - s_1s_5 - s_2s_4 - s_3^2/2 + 9s_6/2$. Notice that in each term here the sum of the indices is 6, the degree of the original polynomial.

Some special cases: If α is a root of an irreducible polynomial $f(x) = a\prod_{i=1}^{d}(x-\alpha_i)$ then there are two particular symmetric polynomials of the roots of special interest:

The *trace* of α is $\alpha_1 + \alpha_2 + \dots + \alpha_d$, the sum of the roots of f.

The *norm* of α is $\alpha_1\alpha_2\dots\alpha_d$, the product of the roots of f.

2.5. Constructibility and Galois theory. The ancient Greeks were interested in what could be constructed using only an unmarked ruler (i.e. a straight edge) and a compass. Three questions stumped them:

(1) *Quadrature of the circle*: Draw a square that has area equal to that of a given circle.
(2) *Duplication of the cube*: Construct a cube that has twice the volume of a given cube.
(3) *Trisection of the angle*: Construct an angle which is one third the size of a given angle.

Let us formulate these problems algebraically.

(1) The area of the square is π, so we need to be able to find a root of $x^2 - \pi$.

(2) Starting with a cube of side length 1, the new cube would have side length $2^{1/3}$; in other words we need to find the real root of $x^3 - 2$.

(3) Constructing an angle θ as is difficult as constructing a right angled triangle containing that angle, and so the triangle with side lengths $\sin\theta$, $\cos\theta$, 1. Hence if we start with angle 3θ and wish to determine θ, we will need to be able to determine $\cos\theta$ from $\cos 3\theta$ and $\sin 3\theta$. But these are linked by the formula $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$, that is we need to find the root $x = 2\cos\theta$ of $x^3 - 3x - A$ where $A = 2\cos 3\theta$. So if $\theta = \pi/9$ this yields the equation $x^3 - 3x - 1$.

The first is impossible because π is transcendental (something that requires proof, and is by no means easy). The next two, whether we can construct the roots of the polynomials $x^3 - 2$ and $x^3 - 3x - 1$, we shall discuss now:

---

[7] Actually this is only really correct if the k_j are distinct. To correct we divide through by $\prod_i m_i!$ where m_i is the number of k_j that equal i.

Our first goal is to understand the algebra of a new point constructed from given points and lengths.

Proposition 2.5.1. *Given a set of known points on known lines, and a set of lengths, any new points that can be constructed using ruler and compass will have coordinates that can be determined as roots of degree one or two polynomials whose coefficients are rational functions of the already known coordinates.*

Proof. Lines are defined by pairs of points: Given the points $A = (a_1, a_2)$ and $B = (b_1, b_2)$ the line between them is $(b_1 - a_1)(y - a_2) = (b_2 - a_2)(x - a_1)$.

`Exercise 2.5.1.` Show that the coefficients of the equation of this line can be determined by a degree one equation in already known coordinates.

`Exercise 2.5.2.` Prove that any two (non-parallel) lines intersect in a point that can be determined by a degree one equation in the coefficients of the equations of the lines.

Given a point $C = (c_1, c_2)$ and a radius r, we can draw a circle $(x-c_1)^2+(y-c_2)^2 = r^2$.

`Exercise 2.5.3.` Prove that the points of intersection of this circle with a given line can be given by a degree two equation in already known coordinates. (Hint: Substitute the value of y given by the line, into the equation of the circle.)

`Exercise 2.5.4.` Prove that the points of intersection of two circles can be given by a degree two equation in already known coordinates. (Hint: Subtract the equations for the two circles.)

So to show that one cannot duplicate the cube, or trisect an angle, we need to have a theory that shows that the roots of irreducible polynomials of degree three cannot be determined in terms of a (finite) succession of roots of linear or quadratic polynomials whose coefficients are already constructed. This is the beginning of Galois theory.

Fields. A field is a set of objects amongst which we can apply the usual operations of arithmetic (i.e. addition, subtraction, multiplication and division).[8] In number theory, the basic field is the set of rationals, $\mathbb{Q}$. We can adjoin an irrational to $\mathbb{Q}$, like $\sqrt{2}$, to obtain $\mathbb{Q}(\sqrt{2})$, the set of all arithmetic expressions in $\sqrt{2}$ with rational coefficients.

`Exercise 2.5.5.` Show that $\mathbb{Q}(\sqrt{2}) = \{r + s\sqrt{2} : \ r, s \in \mathbb{Q}\}$.

One can even adjoin several irrationals to $\mathbb{Q}$, for example to obtain $\mathbb{Q}(2^{1/2}, 3^{1/3}, 5^{1/5}, 7^{1/7})$. One might ask whether there exists α such that this field can be written as $\mathbb{Q}(\alpha)$ (so that every element of the field can be given as a polynomial in α with rational coefficients)? If so then the field extension is *simple*, but this is not always the case. In our cases of interest, like in the example of the fifth roots of unity, we start with $K = \mathbb{Q}(\sqrt{5})$, and then the fifth roots of unity all live in $L = K(\sqrt{-2\sqrt{5} - 10}, \sqrt{2\sqrt{5} - 10})$. In fact if we call these two elements α, β then $\alpha\beta = 4\sqrt{5} \in K$, and so $L = K(\alpha)$ so L/K is a simple extension.

`Exercise 2.5.6.` Verify that $\sqrt{5} = -\alpha^2/2 - 5$ and show how to find β as a function of α with rational coefficients. Deduce that $L/\mathbb{Q}$ is a simple extension.

[8] Technically, the objects are organized into both additive and multiplicative groups — see section 2.8 for more details.

The degree of the *field extension* $\mathbb{Q}(\alpha)$ of $\mathbb{Q}$ is the degree of the minimal polynomial for α over $\mathbb{Q}$. Since any field extension K of $\mathbb{Q}$ can be obtained as $K = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_n)$, let $K_0 = \mathbb{Q}$, $K_1 = \mathbb{Q}(\alpha_1)$, $K_2 = K_1(\alpha_2), \dots, K_n = K$, we obtain the degree of $K/\mathbb{Q}$ as the product of the degrees of K_{j+1}/K_j. If L is a subfield of K then we can find numbers $\beta_1, \dots, \beta_m$ such that $K = L(\beta_1, \dots, \beta_m)$.

With this we can give a more precise description of constructibility: A number γ is *constructible* (by ruler and compass) if there exists a field K such that $\gamma \in K$ and K can be written as $K = \mathbb{Q}(\alpha_1, \alpha_2, \dots, \alpha_n)$ where each K_{j+1}/K_j has degree 2. Hence the degree of K over $\mathbb{Q}$ is a power of 2. Let $L = \mathbb{Q}(\gamma)$. We see from the above that the degree of $L/\mathbb{Q}$, times the degree of K/L equals the degree of $K/\mathbb{Q}$, which is a power of 2. Hence the degree of $L/\mathbb{Q}$ must itself be a power of 2; that is the degree of the minimum polynomial of γ must be a power of 2. We deduce that if the minimum polynomial of γ has degree 3 then γ is not constructible.

`Exercise 2.5.7.` Deduce that one cannot duplicate the cube nor trisect the angle $\pi/3$. (You will need to show that the relevant cubic polynomials are irreducible over $\mathbb{Z}$.)

An *algebraic number* is a number $\alpha \in \mathbb{C}$ which satisfies a polynomial with integer coefficients. An *algebraic integer* is a number $\alpha \in \mathbb{C}$ which satisfies a *monic* polynomial with integer coefficients.

`Exercise 2.5.8.` Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$ of minimal degree for which $f(\alpha) = 0$, where the gcd of the coefficients of f is 1.

(1) Show that if $g(x) \in \mathbb{Z}[x]$ with $g(\alpha) = 0$ then $f(x)|g(x)$.
(2) Deduce that $f(x)$ is well-defined and unique, and so can be called the *minimum polynomial* of α.
(3) Show that if f has leading coefficient a then $a\alpha$ is an algebraic integer.
(4) Show that if $g(x) \in \mathbb{Z}[x]$ is monic, and $g(\alpha) = 0$, then α is an algebraic integer.

If α is an algebraic integer then so is $m\alpha + n$ for any integers m, n; for if $f(x)$ is the minimal polynomial of α and has degree d then $F(x) := m^d f(\frac{x-n}{m})$ is a monic polynomial in $\mathbb{Z}[x]$ with root $m\alpha + n$.

Suppose that α and β are algebraic integers with minimal polynomials f and g. Then

$$\prod_{\substack{u:\ f(u)=0 \\ v:\ g(v)=0}} (x - (u+v)) = \prod_{u:\ f(u)=0} g(x-u).$$

By the fundamental theorem of symmetric polynomials this has rational coefficients, and so $\alpha + \beta$ is an algebraic number.

`Exercise 2.5.9.` Prove that $\alpha\beta$ is an algebraic number.

In the fundamental theorem of arithmetic we ignored negative integers. If we seek to generalize the fundamental theorem then we cannot do this. The right way to think about this is that every non-zero integer is of the form un where n is a positive integer and $u = -1$ or 1. These two values for u are the only integers that divide 1, and it is for this reason they are a bit exceptional. In general we define a *unit* to be an algebraic integer that divides 1, that is an algebraic integer u is a unit if and only if there exists an algebraic integer v such that $uv = 1$.

`Exercise 2.5.10.` Show that if $f(x)$, the minimum polynomial for u, has degree d, then $x^d f(1/x)$ is the minimum polynomial for $1/u$. Deduce that u is a unit if and only if $f(0)$ equals 1 or -1.

2.6. Resultants and Discriminants. Suppose that we have two polynomials $f(x) = f_0x^D + \dots$ and $g(x) = g_0x^d + \dots \in \mathbb{Z}[x]$ where $D \geq d$ and f_0, g_0 are non-zero. We can apply the Euclidean algorithm even in $\mathbb{Z}[x]$, subtracting an appropriate polynomial multiple of the polynomial of smaller degree, from a constant multiple of the polynomial of larger degree, to reduce the degree of the polynomial of larger degree; that is take $h(x) = g_0f(x) - f_0x^{D-d}g(x)$ to get a new polynomial in $\mathbb{Z}[x]$ of degree $< D$. Moreover if we define $\gcd(f(x), g(x))$ to be the polynomial in $\mathbb{Z}[x]$ of largest degree that divides both $f(x)$ and $g(x)$, then the same proof as in the integers yields that $\gcd(f(x), g(x)) = \gcd(g(x), h(x))$, so we can iterate our procedure until one of the two entries is 0. Evidently $\gcd(f(x), 0) = f(x)$. Hence this implies (as in the integers) that we have polynomials $a(x), b(x) \in \mathbb{Z}[x]$ such that

$$a(x)f(x) + b(x)g(x) = R \gcd_{\mathbb{Z}[x]}(f(x), g(x))$$

for some constant R. One can show that $\deg a < \deg g$ and $\deg b < \deg f$

The most interesting case for us is when $\gcd(f(x), g(x)) = 1$, that is when f and g have no common root, and we divide any common integer factors out from the three terms, to obtain

$$a(x)f(x) + b(x)g(x) = R,$$

where R is the *resultant* of a and b. Now, let us suppose that there exists an integer m such that $f(m) \equiv g(m) \equiv 0 \pmod p$. Substituting in $x = m$ we see that p divides R. This argument can be generalized, using some algebraic number theory, to show that if f and g have any common factor mod p (not just a linear polynomial) than p divides R.

Now suppose that prime p divides R so that $a(x)f(x) \equiv -b(x)g(x) \pmod p$. Hence $f(x)$ divides $b(x)g(x) \pmod p$, but f has higher degree than b and so it must have some factor in common with $g(x) \pmod p$. Thus we have an "if and only if" criterion:

Proposition 2.6.1. *Suppose that $f(x), g(x) \in \mathbb{Z}[x]$ have no common roots. Then prime p divides the resultant of f and g if and only if f and g have a common polynomial factor mod p.*

A particularly interesting special case of Proposition 2.6.1 is where we take $g(x) = f'(x)$. The resultant of f and f' is the *discriminant* of f. Let us check this: If $f(x) = ax^2 + bx + c$ then $f'(x) = 2ax + b$ and so

$$(2ax + b)(2ax + b) - 4a(ax^2 + bx + c) = b^2 - 4ac.$$

If $f(x) = x^3 + ax + b$ then $f'(x) = 3x^2 + a$ and so

$$9(3b - 2ax)(x^3 + ax + b) + (6ax^2 - 9bx + 4a^2)(3x^2 + a) = 4a^3 + 27b^2.$$

Corollary 2.6.2. *Suppose that $f(x) \in \mathbb{Z}[x]$ has no repeated roots. Then prime p divides Δ, the discriminant of f if and only if f has a repeated polynomial factor mod p.*

We should also note that the polynomial common factor of highest degree of f and f' can be obtained by using the Euclidean algorithm but can also be described as

$$\gcd_{\mathbb{Z}[t]}(f(t), f'(t)) = c' \prod_{i=1}^{k} (t-\alpha_i)^{e_i - 1}, \text{where } f(t) = c \prod_{i=1}^{k} (t-\alpha_i)^{e_i}$$

and c' divides c. In the case that $f(x)$ has no repeated roots, so that $\gcd_{\mathbb{Z}[t]}(f(t), f'(t)) \in \mathbb{Z}$, let us write

$$a(x)f(x) + b(x)f'(x) = \Delta$$

so that $f'(x) = c\sum_{j=1}^{d} \prod_{1 \le i \le d,\ i \ne j} (x - \alpha_i)$. Hence $f'(\alpha_j) = c\prod_{i:\ i \ne j} (\alpha_j - \alpha_i)$ and note that $\Delta = b(\alpha_j)f'(\alpha_j)$ so that $f'(\alpha_j)$ divides Δ. In fact one can determine the discriminant of f as

$$\pm c^{2d-2} \prod_{1 \le i < j \le d} (\alpha_i - \alpha_j)^2 \ = \ \pm c^{d-2} \prod_{j=1}^{d} f'(\alpha_j).$$

`Exercise 2.6.1.` By multiplying $f(x)$ through by a constant, establish that if such a formula for the discriminant holds then one must have an initial term of a^{2d-2}.

`Exercise 2.6.2.` Show that if $f(t) = \prod_{i=1}^{k} (t-\alpha_i)^{e_i}$ then $\prod_{j=1}^{d} f'(\alpha_j)$ is an integer, by using the theory of symmetric polynomials.

2.7. Möbius transformations: Lines and circles go to lines and circles. In both the Euclidean algorithm and in working with binary quadratic forms we have seen maps $(x, y) \to (\alpha x + \beta y, \gamma x + \delta y)$. These *linear transformations* have various nice properties, one of which is that a line is mapped to a line under such transformations.

A *Möbius transformation* acts on the complex plane (plus the "point" ∞). It is a map of the form

$$z \ \to \ \frac{\alpha z + \beta}{\gamma z + \delta} \quad \text{where} \quad \alpha\delta - \beta\gamma \ne 0.$$

Hence we see that $\infty \to \alpha/\gamma$ (and $\infty \to \infty$ if $\gamma = 0$), and $-\delta/\gamma \to \infty$.

`Exercise 2.7.1.` Show that if one composes two Möbius transformations one gets another one.

`Exercise 2.7.2.` Show that the Möbius transformations $z \to z + 1$, $z \to -1/z$ and $z \to \lambda z$ compose to give all Möbius transformations. (You might read section 4.3 before trying this.)

Let's study the two basic shapes, a line and a circle, and how they map under Möbius transformations. Certainly under translations $z \to z + k$, and dilations $z \to \lambda z$, it is clear geometrically that lines map to lines and circles map to circles.

We now focus on the map $z \to -1/z$. Notice that if we apply the map twice then we get back to the original point: A circle centered at the origin of radius r has equation $|z| = r$, and is mapped to the circle, $|z| = 1/r$, centered at the origin of radius $1/r$.

`Exercise 2.7.3.` Show that A line through the origin has equation $\overline{z} = \alpha z$ where $|\alpha| = 1$. Hence this gets mapped to the line $\overline{z} = (1/\alpha)z$.

`Exercise 2.7.4.` Show that any line in the complex plane that does not go through the origin can be viewed as the set of points equi-distant from 0 and some other point $\alpha \neq 0$.

This last exercise implies that any line that does not go through the origin may be written as $|z| = |\alpha - z|$ for some $\alpha \neq 0$. Under the map $z \to -1/z$ we get $|z - \beta| = |\beta|$ where $\beta = -1/\alpha$, the circle centered at $-1/\alpha$ that goes through the origin. Applying the map again we find that any circle that goes through the origin gets mapped back to a line that does not pass through the origin.

Finally we must deal with circles that do not pass through the origin nor have their centers at the origin; that is $|z - \alpha| = r$, where $|\alpha| \neq 0, r$. Under the map $z \to -1/z$ this goes to $|z - \beta| = t|z|$ where $\beta = -1/\alpha$ and $t = r/|\alpha| \neq 1$.

`Exercise 2.7.5.` Show that if $\beta = (1 - t^2)\gamma$ with $t \neq 1$ then $|z - \beta| = t|z|$ is the same as $|z - \gamma| = t|\gamma|$, and is therefore a circle.

`Exercise 2.7.6.` Prove that to determine a Möbius transformation one need only know the pre-images of $0, 1$ and ∞.

2.8. Groups. We discuss the abstract notion of a *group* because it is a structure that occurs often in number theory (and throughout mathematics). We can prove results for groups in general, and then these results apply for all examples of groups that arise (one can waste a lot of energy giving the same proof, with minor variations, in each case that a group arises). Many of the main theorems about groups were first proved in a number theory context and then found to apply elsewhere. The main examples of groups are additive groups such as the integers, the rationals, the complex numbers, the integers mod p, the polynomials of given degree, matrices of given size, etc, and multiplicative groups such as the rationals, the complex numbers, the integers mod p, invertible matrices of given dimensions, but not the integers or polynomials.

A *group* is defined to be a set of objects G, and an operation, call it $*$, such that:

(i) If $a, b \in G$, then $a * b \in G$. We say that G is *closed* under $*$.

(ii) If $a, b, c \in G$, then $(a * b) * c = a * (b * c)$; that is, when we multiply three elements of G together it does not matter which pair we multiply first. We say that G is *associative*.

(iii) There exists an element $0 \in G$ such that for every $a \in G$ we have $a * 0 = a$. We call 0 the *identity element* of G for $*$.

(iv) For every $a \in G$ there exists $b \in G$ such that $a * b = b * a = 0$. We say that b is the *inverse* of a. We sometimes write $-a$ or a^{-1}.

One can check that the examples of groups given above satisfy these criteria. We see that there are both finite and infinite groups. However, there is one property that one is used to with numbers and polynomials that is not used in the definition of the a group, and that is that $a * b = b * a$, that a and b *commute*. Although this often holds, there are some simple counterexamples, for instance 2-by-2 matrices:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -1 & 2 \end{pmatrix} \quad \text{whereas} \quad \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

We develop the full theory for 2-by-2 matrices at the end of this subsection. If all pairs of elements of a group commute then we call the group *commutative* or *abelian*.

A given group G can contain other, usually smaller, groups H, which are called *subgroups*. Every group G contains the subgroup given by the identity element, $\{0\}$, and also the subgroup G. It can also contain others. For example the additive group of integers mod 6 with elements $\{0,1,2,3,4,5\}$ contains the four subgroups $\{0\}$, $\{0,3\}$, $\{0,2,4\}$, $\{0,1,2,3,4,5\}$. Note that every group, and so subgroup, contains the identity element. Infinite groups can also contain subgroups, indeed

$$\mathbb{C} \supset \mathbb{R} \supset \mathbb{Q} \supset \mathbb{Z}.$$

If H is a subgroup of G then we define a *left coset* to be the set $a * H = \{a * h : h \in H\}$ for any $a \in G$. (Right cosets are analogously defined, and the two types are indistinguishable if G is a commutative group). One of Gauss's proofs of Fermat's Little Theorem is the prototype of the following result:

Proposition 2.8.1. *Let H be a subgroup of G. The left cosets of H in G are disjoint. Moreover if G is finite then they partition G, and hence the size of H, $|H|$, divides $|G|$.*

Proof. Suppose that $a*H$ and $b*H$ have a common element c. Then there exists $h_1, h_2 \in H$ such that $a*h_1 = c = b*h_2$. Therefore $b = a*h_1*(h_2)^{-1}$ so that $b \in a*H$ as $h_1*(h_2)^{-1} \in H$ since H is closed. Writing $b = a * k$, $k \in H$, suppose that $g \in b * H$ so that $g = b * h$ for some $h \in H$. Then $g = (a * k) * h = a * (k * h) \in a * H$ by associativity and closure of H. Hence $b * H \subset a * H$. By an analogous proof we have $a * H \subset b * H$, and hence $a * H = b * H$. Therefore any two left cosets of H in G are either disjoint or identical.

Suppose that G is finite, and let $a_1 * H, a_2 * H, \ldots, a_k * H$ be a maximal set of disjoint cosets of H inside G. If their union does not equal G then there exists $a \in G$ which is in none of these cosets. But then the coset $a * H$ is disjoint from these cosets (by the first part), and this contradicts maximality.

We have encountered the cosets of the subgroup $\mathbb{Z}$ of the additive group $\mathbb{R}$ in the proof of Dirichlet's Theorem. Since the cosets look like $a + \mathbb{Z}$, they are all represented by a number in $[0,1)$, that is by $\{a\}$, the fractional part of a. We denote this by $\mathbb{R}/\mathbb{Z}$, which is also an additive group. This can be represented by wrapping the real numbers around the unit circle; the line segment from 0 to 1 representing one complete revolution. Hence to find the coset representation of a given real number t we simply go round the circle this many times. We are familiar with this when working with the exponential function, since $e^{2i\pi t} = e^{2i\pi\{t\}}$ as $e^{2i\pi} = 1$. (For convenience we will often write $e(t)$ in place of $e^{2i\pi t}$.)

`Exercise 2.8.1.` Show that $a \equiv b \pmod{m}$ if and only if $\frac{a}{m} = \frac{b}{m}$ in $\mathbb{R}/\mathbb{Z}$; that is a/m and b/m belong to the same coset of $\mathbb{R}/\mathbb{Z}$.

`Exercise 2.8.2.` Prove that if H is a subgroup of a finite abelian group G then the cosets $a*H$ themselves form a group. We call this the *quotient group* G/H. (We just encountered the example $\mathbb{R}/\mathbb{Z}$.) Show that every element G can be written in a unique way as $a * h$ where $h \in H$ and $a \in G/H$, which we write as $G \cong H \oplus G/H$ (we say that G is the *direct sum* of H and G/H). If G is finite show that $|G/H| = |G|/|H|$.

The most common type of group encountered in number theory is the additive group of integers mod m. One way to view this is as a map from the integers onto the residue

classes mod m, from integer a to its congruence class mod m. We write this $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$. Here $m\mathbb{Z}$ denotes "m times the integers", that is the integers divisible by m, each of which map to 0 (mod m). The Chinese Remainder Theorem states that there is a 1-to-1 correspondence between the residue classes $a \pmod m$, and the "vector" of residue classes $(a_1 \pmod{m_1},\ a_2 \pmod{m_2}, \dots, a_k \pmod{m_k})$, when the m_is are pairwise coprime and their product equals m. This is usually written

$$\begin{aligned} \mathbb{Z}/m\mathbb{Z} \quad &\cong \quad \mathbb{Z}/m_1\mathbb{Z} \quad \oplus \quad \mathbb{Z}/m_2\mathbb{Z} \quad \oplus \dots \oplus \quad \mathbb{Z}/m_k\mathbb{Z} \\ a \pmod m \leftrightarrow (a_1 \pmod{m_1},\ &a_2 \pmod{m_2}, \dots, a_k \pmod{m_k}). \end{aligned}$$

The beauty of this is that most arithmetic operations mod m can be "broken down" into the same arithmetic operations modulo each m_i performed componentwise. This is particularly useful when $m = \prod_p p^{e_p}$ and then the m_i are the individual p^{e_p}, since some arithmetic operations are much easier to do modulo prime powers than modulo composites. Besides addition the most important of these operations is multiplication. Thus the above correspondence gives a 1-to-1 correspondence between the reduced residue classes mod m, and the reduced residue classes mod the m_i; we write this as

$$(\mathbb{Z}/m\mathbb{Z})^* \quad \cong \quad (\mathbb{Z}/m_1\mathbb{Z})^* \quad \oplus \quad (\mathbb{Z}/m_2\mathbb{Z})^* \quad \oplus \quad \dots \quad \oplus \quad (\mathbb{Z}/m_k\mathbb{Z})^*,$$

considering these now as groups under multiplication; and again the operation (of multiplication) can be understood componentwise. Typically we write 0 for the identity of an additive group, and 1 for the identity of a multiplicative group.

We say that two groups G and H are *isomorphic*, and write $G \cong H$ if there is a 1-to-1 correspondence $\phi : G \to H$ such that $\phi(a *_G b) = \phi(a) *_H \phi(b)$ for every $a, b \in G$, where $*_G$ is the group operation in G, and $*_H$ is the group operation in H.

`Exercise 2.8.3.` Let H be a subgroup of $(\mathbb{Z}/m\mathbb{Z})^*$.

(1) Prove that if n is an integer coprime to m but which is not in a residue class of H, then n has a prime factor which is not in a residue class of H.
(2) Show that if integers $q = p_1 \cdots p_k$ and a are coprime to m then there are infinitely many integers $n \equiv a \pmod m$ such that $(n, q) = 1$.
(3) Prove that if H is not all of $(\mathbb{Z}/m\mathbb{Z})^*$ then there are infinitely many primes which do not belong to any of the residue classes of H. (Hint: Modify Euclid's proof of the infinitude of primes.)

Proposition 2.8.1 implies the following generalization of Fermat's Little Theorem:

Corollary 2.8.2. (Lagrange's Theorem) *For any element a of any finite multiplicative group G we have $a^{|G|} = 1$.*

Proof. Let m be the order of a in G; that is, the least positive integer for which $a^m = 1$.

`Exercise 2.8.4.` Prove that $H := \{1, a, a^2, \dots, a^{m-1}\}$ is a subgroup of G.

By Proposition 2.8.1 we know that $m = |H|$ divides $|G|$, and so

$$a^{|G|} = (a^m)^{|G|/m} = 1^{|G|/m} = 1.$$

To deduce Euler's Theorem let $G = (\mathbb{Z}/m\mathbb{Z})^*$ so that $|G| = \phi(m)$.

`Exercise 2.8.5.` Deduce that if $|G|$ is a prime then G is cyclic.

Wilson's Theorem for finite abelian groups. *The product of the elements of any given finite abelian group equals* 1 *unless the group contains exactly one element, ℓ, of order two, in which case the product equals ℓ.*

Proof. As in the proof of Wilson's Theorem we partition the elements, each element with its inverse, providing that they are distinct, since these multiply together to give 1, and hence the product of all of them gives 1. This leaves the product of the elements which are their own inverses; that is the roots of $x^2 = 1$ in the group. Now if $\ell \neq 1$ and $\ell^2 = 1$ then we partition these elements into pairs, $x, \ell x$. The product of each such pair equals ℓ, and therefore the product of all the $2N$ roots of $x^2 = 1$ equals ℓ^N. Now if N is even this equals 1, as $\ell^2 = 1$, and if N is odd then this equals ℓ. In this case the only roots of $x^2 = 1$ are 1 and ℓ, for if $m^2 = 1,\ m \neq 1, \ell$, then the product would also equal m and hence $m = \ell$, a contradiction.

The group $\mathbb{Z}/m\mathbb{Z}$, sometimes written C_m, is called the *cyclic* group of order m, which means that the elements of the group are precisely $\{0 \cdot a, 1 \cdot a, 2 \cdot a, \dots, (m-1) \cdot a\}$, the multiples of the *generator* a (in this case we can take $a = 1$). We now find the structure of all finite abelian groups:

Fundamental Theorem of Abelian Groups. *Any finite abelian group G may be written as*

$$\mathbb{Z}/m_1\mathbb{Z} \oplus \mathbb{Z}/m_2\mathbb{Z} \oplus \dots \mathbb{Z}/m_k\mathbb{Z}.$$

In other words every element of G may be written in the form $g_1^{e_1} g_2^{e_2} \dots g_k^{e_k}$ where g_j has order m_j. We write $G = \langle g_1, g_2, \dots, g_k \rangle$.

Proof. By induction on the size of G. Let a be the element of highest order in G, say of order m. By exercise 2.8.4 we know that $H := \{1, a, a^2, \dots, a^{m-1}\}$ is a subgroup of G. If $1 < m < |G|$ then, by induction, both H and G/H can be written as the direct sum of cyclic groups, and therefore G since $G \cong H \oplus G/H$ by exercise 2.8.2.

We saw above that we can write each of the $\mathbb{Z}/m_1\mathbb{Z}$ as a direct sum of cyclic groups of prime power order. Hence, by the Fundamental Theorem of Abelian Groups, we can write any finite abelian group as a direct sum of cyclic groups of prime power order. For each given prime we can put the powers of that prime in descending order, that is the p-part of the group is $\mathbb{Z}/p^{e_1}\mathbb{Z} \oplus \mathbb{Z}/p^{e_2}\mathbb{Z} \oplus \dots \mathbb{Z}/p^{e_\ell}\mathbb{Z}$ where $e_1 \geq e_2 \geq \dots$ Now if we take the components of largest prime power orders we can recombine these, and then those of second highest order, etc, that is we can write

$$\oplus_p\ \mathbb{Z}/p^{e_r}\mathbb{Z} \cong \mathbb{Z}/n_r\mathbb{Z}$$

for $r = 1, 2, 3, \dots$ so that

$$G \cong \mathbb{Z}/n_1\mathbb{Z} \oplus \mathbb{Z}/n_2\mathbb{Z} \oplus \dots \mathbb{Z}/n_\ell\mathbb{Z} \quad \text{where} \quad n_\ell | n_{\ell-1} | \dots | n_2 | n_1.$$

In this case we call ℓ the *rank* of G.

Explicit decomposition of $(\mathbb{Z}/m\mathbb{Z})^*$ as a direct sum of cyclic groups: We saw above that, via the Chinese Remainder Theorem, this is the direct sum of groups of the form $(\mathbb{Z}/p^r\mathbb{Z})^*$ for each prime power $p^r \| m$.

Now if p is an odd prime, then there is a primitive root mod p^r, that is a generator for the group of elements of $(\mathbb{Z}/p^r\mathbb{Z})^*$, and so

$$(\mathbb{Z}/p^r\mathbb{Z})^* \text{ as a multiplicative group } \cong \mathbb{Z}/\phi(p^r)\mathbb{Z} \text{ as an additive group.}$$

If $r \geq 3$ then the elements of $(\mathbb{Z}/2^r\mathbb{Z})^*$ can all be written in the form $\pm g^k \pmod{2^r}$ for some integer $k,\ 0 \leq k \leq 2^{r-2}-1$, for some integer $g \equiv \pm 3 \pmod 8$ which has order 2^{r-2} mod 2^r. This implies that

$$(\mathbb{Z}/2^r\mathbb{Z})^* \text{ as a multiplicative group } \cong \mathbb{Z}/2^{r-2}\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \text{ as an additive group.}$$

Let G be a finite abelian group, written additively. We wish to understand the structure of $2G$. Now if m is odd then every element of $\mathbb{Z}/m\mathbb{Z}$ is twice an element, since for any $a \pmod m$ we have $2 \cdot (\frac{m+1}{2})a \equiv a \pmod m$. On the other hand if $a \pmod{2^e}$ is twice another value then 2 divides a, which yields the residues $\{0, 2, 4, \dots, 2^e - 2\}$. Hence if $H = \mathbb{Z}/2^e\mathbb{Z}$ with $e \geq 1$ then $H/2H$ has two elements, with representatives $\{0, 1\}$, and so this quotient group is isomorphic to $\mathbb{Z}/2\mathbb{Z}$. By our structure theorem we may write any finite abelian group G as $\mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/2^{e_1}\mathbb{Z} \oplus \mathbb{Z}/2^{e_2}\mathbb{Z} \oplus \ldots \oplus \mathbb{Z}/2^{e_\ell}\mathbb{Z}$ for some integers $e_1, e_2 \dots, e_\ell \geq 1$, and so by the Chinese remainder theorem $G/2G \cong (\mathbb{Z}/2\mathbb{Z})^\ell$. We call ℓ the *2-rank of G*. Note that if m is odd then $\{a \pmod m :\ 2a \equiv 0 \pmod m\} = \{0\}$, and $\{a \pmod{2^e} :\ 2a \equiv 0 \pmod{2^e}\} = \{0, 2^{e-1}\}$. In both cases this has the same size as $H/2H$, and so by the Chinese Remainder Theorem we deduce that

$$\#\{g \in G :\ 2g = 0\} = |G/2G|.$$

We know many examples of infinite (additive) abelian groups like $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \dots$ We often are interested in the structure of *finitely generated abelian groups G*. Let T be the elements of G of finite order; so that T must be a finite abelian group and thus has the structure above (we call T *the torsion subgroup of G*). Then we can write

$$G \cong T \oplus \mathbb{Z}^r;$$

in other words there exist linearly independent $g_1, \dots, g_r$ elements of G of infinite order such that every element G may be written as

$$g = t + e_1g_1 + e_2g_2 + \ldots + e_rg_r \ \text{ with } \ t \in T, \text{ and each } e_i \in \mathbb{Z}.$$

We call r the *rank of the infinite part of G*. The rank of G is, of course, r plus the rank of T (that is, the minimal number of elements whose linear combinations generate G). It is another application of the Chinese Remainder Theorem to prove that

$$G/2G \cong T/2T \oplus (\mathbb{Z}/2\mathbb{Z})^r$$

Hence the 2-rank of G equals r plus the 2-rank of T, and above we saw that $T/2T$ equals the number of elements of T of order dividing 2. In the theory of elliptic curves we often

wish to determine the *rank of the infinite part of* some given group G. One method to do so is to find the 2-rank of G, and subtract the log, in base 2, of the number of elements of G of order dividing 2.

As promised we finish this section by determining:
What commutes with a given 2-by-2 matrix? We will now explore which 2-by-2 matrices commute with a given 2-by-2 matrix, M.

`Exercise 2.8.7.` Prove that if A and B commute with M then so does $rA + sB$ for any real numbers r and s.

It is evident that I and M commute with M, and hence any linear combinations of I and M. We will show that this is all, unless M is a multiple of the identity. Let $\mathcal{M}_2$ be the set of 2-by-2 matrices with entries in $\mathbb{C}$.

Proposition 2.8.3. *Given $M \in \mathcal{M}_2$, let $C(M) := \{A \in \mathcal{M}_2 : \ AM = MA\}$. If $M = aI$ for some constant a then $C(M) = \mathcal{M}_2$. Otherwise $C(M) = \{rI + sM : \ r, s \in \mathbb{C}\}$.*

Proof. Let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $A \in C(M)$ then so is $B = A - rI - sM$ for any r and $s \in \mathbb{C}$.

`Exercise 2.8.8.` Prove that if $a \neq d$ then we can select r and s so that the diagonal of B is all 0s.

If $a \neq d$ then write $B = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix}$, so that

$$\begin{pmatrix} cx & dx \\ ay & by \end{pmatrix} = \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = BM = MB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & x \\ y & 0 \end{pmatrix} = \begin{pmatrix} by & ax \\ dy & cx \end{pmatrix}.$$

The off diagonal terms yield that $x = y = 0$ and so $A = rI + sM$.

Now suppose that $a = d$ and $M \neq aI$. If $b \neq 0$ then we can select r and s so that the top row of B is all 0s; that is $B = \begin{pmatrix} 0 & 0 \\ x & y \end{pmatrix}$; then the top row of $MB = BM$ yields that $x = y = 0$ and so $A = rI + sM$. Otherwise $c \neq 0$ and an analogous argument works.

2.9. Unique Factorization and ideals. The proof of the Fundamental Theorem of Arithmetic appears to use very few ideas, and so one might expect that it generalizes into all sorts of other domains. For example, do polynomials factor in a unique way into irreducibles? Or numbers of the form $\{a + b\sqrt{d} : \ a, b \in \mathbb{Z}\}$? Or other simple arithmetic sets?

Before embarking on this we need to clarify how to think about factorizations like $3 = (-1)\cdot(-3)$ and $-3 = (-1)\cdot 3$, which make it appear that every integer can be factored into at least two others. The issue here is that one of the factors, -1, divides 1 in $\mathbb{Z}$, that is $(-1)\cdot(-1) = 1$, so division by such a number does not really reduce the size of numbers that we are working with. Elements of a ring that divide 1 are called *units* and will be excluded from the notion of factorization.

The rings $\{a + b\sqrt{d} : \ a, b \in \mathbb{Z}\}$ are closed under addition as well as multiplication. In $R := \{a + b\sqrt{-5} : \ a, b \in \mathbb{Z}\}$ we have the example

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

Now suppose that prime $p = (a+b\sqrt{-5})(c+d\sqrt{-5})$ for $p = 2$ or 3. Then $(a,b)(c,d)$ divides p, so at least one of these gcds equals 1, say $(a,b) = 1$. The coefficient of the imaginary part is $ad + bc = 0$, and so $a|bc$ and therefore $a|c$. Writing $c = ak$ we have $d = -bk$ and therefore $p = k(a+b\sqrt{-5})(a-b\sqrt{-5}) = k(a^2+5b^2)$. Therefore $a^2+5b^2 = 1, 2$ or 3, so that $b = 0,\ a = \pm 1$, yielding the uninteresting factorization $(\pm 1)(\pm p)$, and hence neither 2 nor 3 can be factored in R. Therefore 2 and 3 are irreducible in R. Also, by taking complex conjugates $1+\sqrt{-5} = (a+b\sqrt{-5})(c+d\sqrt{-5})$ if and only if $1-\sqrt{-5} = (a-b\sqrt{-5})(c-d\sqrt{-5})$, and hence $6 = (1+\sqrt{-5})(1-\sqrt{-5}) = (a+b\sqrt{-5})(c+d\sqrt{-5})(a-b\sqrt{-5})(c-d\sqrt{-5}) = (a^2+5b^2)(c^2+5d^2)$. Since a^2+5b^2 and c^2+5d^2 are both positive and $\neq 2$ or 3, with product 6, one must equal 1, the other 6, say $|a| = |b| = |c| = 1,\ d = 0$, which shows that $1+\sqrt{-5}$ and $1-\sqrt{-5}$ are both irreducible in R. So we *do not have* unique factorization of elements of R.

Let R be a set of numbers that is closed under addition and subtraction; for example $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ or $\mathbb{C}$, but not $\mathbb{N}$. We define the *ideal* generated by $a_1, \ldots, a_k$ over R, to be the set of linear combinations of $a_1, \ldots, a_k$ with coefficients in R; that is

$$I_R(a_1, \ldots, a_k) = \{r_1a_1 + r_2a_2 + \ldots + r_ka_k : \ r_1, \ldots, r_k \in R\}.$$

(Note $a_1, \ldots, a_k$ are not necessarily in R.) We have seen that any ideal over $\mathbb{Z}$ can be generated by just one element. The reason we take such interest in this definition is that this is not necessarily true when the a_i are taken from other domains. For example when $R = \mathbb{Z}[\sqrt{-5}]$, that is the numbers of the form $u + v\sqrt{-5}$ where u and v are integers, we have just seen that the ideal $I_R(2, 1+\sqrt{-5})$ cannot be generated by just one element.

A *principal ideal* is an ideal that can be generated by just one element.

`Exercise 2.9.1.` Show that the only units of R are 1 and -1. What if we allow $d = 1$?

We need arithmetic to work in $R = \mathbb{Z}[\sqrt{-d}]$ even though, as we have seen we cannot always factor uniquely into irreducibles. It turns out that the way to proceed is to replace all the numbers in the ring by the ideals that they generate. To do so we need to be able to multiply ideals, and it is easy to show from their definition that this works out by multiplying generators: For any $\alpha, \beta, \gamma, \delta \in R$ we have

$$I_R(\alpha, \beta) \cdot I_R(\gamma, \delta) = I_R(\alpha\gamma, \alpha\delta, \beta\gamma, \beta\delta).$$

Therefore if $n = ab$ in R then $I_R(n) = I_R(a)I_R(b)$. There are several desirable properties of ideals: All issues with units disappear for if I is an ideal and u a unit then $I = uI$. Ideals can be factored into prime ideals in a unique way; in all "number rings" R we get unique factorization. Note though that primes are no longer elements of the ring, or even necessarily principal ideals of the ring.

In our example $6 = 2 \cdot 3 = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$ above, all of $2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}$ are irreducibles of $\mathbb{Z}[\sqrt{-5}]$ but none generate prime ideals. In fact we can factor the ideals they generate into prime ideals as

$$\begin{aligned}
I_R(2) &= I_R(2, 1+\sqrt{-5}) \cdot I_R(2, 1-\sqrt{-5}) \\
I_R(3) &= I_R(3, 1+\sqrt{-5}) \cdot I_R(3, 1-\sqrt{-5}) \\
I_R(1+\sqrt{-5}) &= I_R(2, 1+\sqrt{-5}) \cdot I_R(3, 1+\sqrt{-5}) \\
I_R(1-\sqrt{-5}) &= I_R(2, 1-\sqrt{-5}) \cdot I_R(3, 1-\sqrt{-5}).
\end{aligned}$$

None of these prime ideals are principal, as we saw above. We do, though, call an element of R prime if it generates a prime ideal.

Here we see that the notion of "irreducible" and "prime" are not in general the same. In fact any prime of R is irreducible, but not vice-versa. One fun question of Davenport is to determine the most prime ideal factors an irreducible integer can have in R.

The *ring of integers* of a field are the algebraic integers in that field. For quadratic fields, these must satisfy an equation of the form $x^2 - bx + c = 0$ where $b, c \in \mathbb{Z}$, and thus can be written as $\frac{b+\sqrt{b^2-4c}}{2}$. If $b = 2B$ is even, then this becomes $B + \sqrt{B^2 - c}$; and any $B + C\sqrt{d}$ can written in this form. If b is odd, then $b^2 - 4c$ is also odd, and so the discriminant d must also be odd (note that $b^2 - 4c = C^2 d$ for some integer C). On the other hand if b and C are odd integers with $d \equiv 1 \pmod 4$ then $\frac{b+C\sqrt{d}}{2}$ is an algebraic integer, taking $c = (b^2 - C^2 d)/4$. Hence we have proved the ring of integers is

$$\begin{cases} \{B + C\sqrt{d} : \ B, C \in \mathbb{Z}\} = I_{\mathbb{Z}}(1, \sqrt{d}) & \text{if } d \text{ is even;} \\ \{\frac{B+C\sqrt{d}}{2}: \ B, C \in \mathbb{Z}, \ B \equiv C \pmod 2\} = I_{\mathbb{Z}}(1, \frac{1+\sqrt{d}}{2}) & \text{if } d \equiv 1 \pmod 4; \end{cases}$$

3. Quadratic equations

3.1. The Pythagorean equation. We wish to find all solutions in integers x, y, z to

$$x^2 + y^2 = z^2.$$

We may assume that x, y, z are all positive and so $z > x, y$. Given any solution we may divide through by any common factor of x, y and z to obtain a solution where $(x, y, z) = 1$.

`Exercise 3.1.1.` Prove that if $(x, y, z) = 1$ and $x^2 + y^2 = z^2$ then x, y and z are pairwise coprime.

Now x and y cannot both be odd else $z^2 = x^2 + y^2 \equiv 1 + 1 \equiv 2 \pmod 8$, which is impossible. Interchanging x and y if necessary we may assume that x is even and y and z are odd. Now

$$(z - y)(z + y) = z^2 - y^2 = x^2.$$

We prove that $(z - y, z + y) = 2$: Since y and z are both odd, we know that 2 divides $(z-y, z+y)$. Moreover $(z-y, z+y)$ divides $(z+y)-(z-y) = 2y$ and $(z+y)+(z-y) = 2z$, and $(2y, 2z) = 2(y, z) = 2$. Therefore there exist integers r, s such that

$$z - y = 2s^2 \text{ and } z + y = 2r^2,$$

so that

$$x = 2rs,\ y = r^2 - s^2,\ \text{and } z = r^2 + s^2.$$

To ensure these are pairwise coprime we need $(r, s) = 1$ and $r + s$ odd. If we now add back in any common factors we get the general solution

$$\boxed{x = 2grs,\ y = g(r^2 - s^2),\ \text{and } z = g(r^2 + s^2).} \tag{3.1}$$

There is also a nice geometric proof of this parametrization:

`Exercise 3.1.3.` Prove that the integer solutions to $x^2 + y^2 = z^2$ with $z \neq 0$ and $(x, y, z) = 1$ are in 1-to-1 correspondence with the rational solutions u, v to $u^2 + v^2 = 1$.

Where else does a line going though $(1, 0)$ intersect the circle $x^2 + y^2 = 1$? Unless the line is vertical it will hit the unit circle in exactly one other point, which we will denote (u, v). Note that $u < 1$. If the line has slope t then $t = v/(u - 1)$ is rational if u and v are. In the other direction, the line through $(1, 0)$ of slope t is $y = t(x - 1)$ which intersects $x^2 + y^2 = 1$ where $1 - x^2 = y^2 = t^2(x - 1)^2$ so that either $x = 1$ or $1 + x = t^2(1 - x)$. Hence

$$u = \frac{t^2 - 1}{t^2 + 1} \text{ and } v = \frac{-2t}{t^2 + 1}$$

are both rational if t is. We have therefore proved that $u, v \in \mathbb{Q}$ if and only if $t \in \mathbb{Q}$. Writing $t = -r/s$ where $(r, s) = 1$ we have $u = \frac{r^2 - s^2}{r^2 + s^2}$ and $v = \frac{2rs}{r^2 + s^2}$, the same parametrization to the Pythagorean equation as in (3.1) when we clear out denominators.

3.2. Fermat's "infinite descent".

In around 1637, Pierre de Fermat was studying the proof of (3.1) in his copy of Bachet's translation of Diophantus's *Arithmetica*. In the margin he wrote:

> "*I have discovered a truly marvelous proof that it is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second into two like powers. This margin is too narrow to contain it.*" — by P. DE FERMAT (1637), in his copy of *Arithmetica*.

In other words, Fermat claimed that for every integer $n \geq 3$ there do not exist positive integers x, y, z for which

$$x^n + y^n = z^n.$$

Fermat did not subsequently mention this problem or his truly marvelous proof elsewhere, and the proof has not, to date, been re-discovered, despite many efforts. However he did exhibit a proof for $n = 4$, based on (3.1), as we now show.

Theorem 3.1. *There are no solutions in non-zero integers x, y, z to*

$$x^4 + y^4 = z^2.$$

Proof. Let x, y, z give the solution in positive integers with z minimal.. We may assume that $\gcd(x, y) = 1$ else we can divide out the common factor. Here we have

$$(x^2)^2 + (y^2)^2 = z^2 \quad \text{with} \quad \gcd(x^2, y^2) = 1,$$

and so, by (3.1), there exist positive integers r, s with $(r, s) = 1$ and $r + s$ odd such that

$$x^2 = 2rs, \ y^2 = r^2 - s^2, \ \text{and} \ z = r^2 + s^2.$$

Now $s^2 + y^2 = r^2$ with y odd and $(r, s) = 1$ and so, by (3.1), there exist positive integers a, b with $(a, b) = 1$ and $a + b$ odd such that

$$s = 2ab, \ y = a^2 - b^2, \ \text{and} \ r = a^2 + b^2,$$

and so

$$x^2 = 2rs = 4ab(a^2 + b^2).$$

Now a, b and $a^2 + b^2$ are pairwise coprime, positive integers whose product is a square so they must each be squares, say $a = u^2, \ b = v^2$ and $a^2 + b^2 = w^2$ for some positive integers u, v, w. Therefore

$$u^4 + v^4 = a^2 + b^2 = w^2$$

yields another solution to the original equation with

$$w \leq w^2 = a^2 + b^2 = r < r^2 + s^2 = z,$$

contradicting the minimality of z.

Corollary 3.2. *There are no solutions in non-zero integers x, y, z to*

$$x^4 + y^4 = z^4.$$

`Exercise 3.2.` Prove this!

3.3. Squares mod m. We are interested in understanding the squares mod m; that is the residues $a \pmod m$ for which there exists $b \pmod m$ with $b^2 \equiv a \pmod m$. By the Chinese Remainder Theorem we know that a is a square mod m if and only if a is a square modulo every prime power factor of m, so it suffices to study only the case where m is a prime power. We show how to recognize squares modulo prime powers, in terms of the squares mod p:

Proposition 3.3. *Suppose that r is not divisible by prime p. If r is a square mod p^k then r is a square mod p^{k+1} whenever $k \geq 1$, except perhaps in the cases $p^k = 2$ or 4.*

Proof. Let x be an integer, coprime with p, such that $x^2 \equiv r \pmod{p^k}$, so that there exists an integer n for which $x^2 = r + np^k$. Therefore

$$(x - jp^k)^2 \;=\; x^2 - 2jxp^k + x^2p^{2k} \;\equiv r + (n - 2jx)p^k \pmod{p^{k+1}};$$

and this is $\equiv r \pmod{p^{k+1}}$ for $j \equiv n/2x \pmod p$ when p is odd. If $p = 2$ then

$$(x - n2^{k-1})^2 \;=\; x^2 - nx2^k + x^2 2^{2k-2} \;\equiv r \pmod{2^{k+1}},$$

provided $k \geq 3$.

`Exercise 3.3.1.` Deduce that integer r is a quadratic residue mod p^k if and only if r is a quadratic residue mod p, when p is odd, and if and only if $r \equiv 1 \pmod{\gcd(2^k, 8)}$ where $p = 2$.

Notice that this implies that exactly half of the reduced residue classes mod p^k are quadratic residues, when p is odd, and exactly one quarter when $p = 2$ and $k \geq 3$.

Using the Chinese Remainder Theorem we deduce from exercise 3.3.1 that if $(a, m) = 1$ then a is a square mod m if and only if $\left(\frac{a}{p}\right) = 1$ for every odd prime p dividing m, and $a \equiv 1 \pmod{\gcd(m, 8)}$.

Euler's criterion. $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod p$, *for all primes p and integers a.*

`Exercise 3.3.2.` Let p be a prime $\equiv 3 \pmod 4$. Show that if $\left(\frac{a}{p}\right) = 1$ and $x \equiv a^{\frac{p+1}{4}} \pmod p$ then $x^2 \equiv a \pmod p$. Can one adapt this method when $p \equiv 1 \pmod 4$?

3.4. The law of quadratic reciprocity. We have already seen that if p is an odd prime then

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4; \\ -1 & \text{if } p \equiv -1 \pmod 4. \end{cases}$$

One also has that

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \text{ or } -1 \pmod 8; \\ -1 & \text{if } p \equiv 3 \text{ or } -3 \pmod 8. \end{cases}$$

To be able to evaluate Legendre symbols we will also need the *law of quadratic reciprocity*. This states that if p and q are distinct odd primes then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \begin{cases} -1 & \text{if } p \equiv q \equiv -1 \pmod 4 \\ 1 & \text{otherwise.} \end{cases}$$

These rules, taken together, allow us to rapidly evaluate any Legendre symbol.

3.5. Sums of two squares. What primes are the sum of two squares? If we start computing we find that

$$2 = 1^2+1^2,\ 5 = 1^2+2^2,\ 13 = 2^2+3^2,\ 17 = 1^2+4^2,\ 29 = 5^2+2^2,\ 37 = 1^2+6^2,\ 41 = 5^2+4^2, \dots$$

so we might guess that the answer is 2 and any prime $\equiv 1 \pmod 4$.

Proposition 3.4. *If p is an odd prime that is the sum of two squares then $p \equiv 1 \pmod 4$.*

Proof. If $p = a^2 + b^2$ then $p \not| a$, else $p|b$ and so $p^2|a^2 + b^2 = p$ which is impossible, and similarly $p \not| b$. Now $a^2 \equiv -b^2 \pmod p$ so that

$$1 = \left(\frac{a}{p}\right)^2 = \left(\frac{-1}{p}\right)\left(\frac{b}{p}\right)^2 = \left(\frac{-1}{p}\right),$$

and therefore $p \equiv 1 \pmod 4$.

The other direction is more complicated

Theorem 3.5. *Any prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares.*

Proof. Since $p \equiv 1 \pmod 4$ we know that there exists an integer b such that $b^2 \equiv -1 \pmod p$. Consider now the set of integers

$$\{i + jb : \ 0 \le i, j \le [\sqrt{p}]\}$$

The number of pairs i, j used in the construction of this set is $([\sqrt{p}] + 1)^2 > p$, and so by the pigeonhole principle, two must be congruent mod p; say that

$$i + jb \equiv I + Jb \pmod p$$

where $0 \le i, j, I, J \le [\sqrt{p}]$ and $\{i, j\} \ne \{I, J\}$. Let $r = i - I$ and $s = J - j$ so that

$$r \equiv bs \pmod p$$

where $|r|, |s| \le [\sqrt{p}] < \sqrt{p}$, and r and s are not both 0. Now

$$r^2 + s^2 \equiv (bs)^2 + s^2 = s^2(b^2 + 1) \equiv 0 \pmod p$$

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

`Exercise 3.5.1.` Suppose that $b \pmod p$ is given, and that $R \geq 1$ and S are positive numbers such that $RS = p$. Prove that there exist integers r, s with $|r| \leq R, 0 < s \leq S$ such that $b \equiv r/s \pmod p$.

What integers can be written as the sum of two squares? Note the identity

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2. \tag{3.2}$$

`Exercise 3.5.2.` Use this to show that the product of two or more integers that are the sum of two squares is itself the sum of two squares.

We see that (3.2) is a useful identity, yet we simply gave it without indicating how one might find such an identity. Let i be a complex number for which $i^2 = -1$. Then we have $x^2 + y^2 = (x + iy)(x - iy)$, a factorization in the set $\{a + bi : \ a, b \in \mathbb{Z}\}$. Therefore

$$\begin{aligned}(a^2 + b^2)(c^2 + d^2) &= (a + bi)(a - bi)(c + di)(c - di) = (a + bi)(c + di)(a - bi)(c - di)\\ &= ((ac - bd) + (ad + bc)i)((ac - bd) - (ad + bc)i)\\ &= (ac - bd)^2 + (ad + bc)^2,\end{aligned}$$

and so we get (3.2). A different re-arrangement leads to a different identity:

$$(a^2 + b^2)(c^2 + d^2) = (a + bi)(c - di)(a - bi)(c + di) = (ac + bd)^2 + (ad - bc)^2. \tag{3.3}$$

`Exercise 3.5.3.` Prove that if prime $p = a^2 + b^2$ is coprime with $c^2 + d^2$ then $\frac{ac-bd}{ad+bc} \equiv \frac{a}{b} \pmod p$ in (3.2); and $\frac{ac+bd}{ad-bc} \equiv -\frac{a}{b} \equiv \frac{b}{a} \pmod p$ in (3.3).

In Theorem 3.5 we saw that every prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares. A few examples indicate that perhaps there is a unique such representation, up to signs and changing the order of the squares. This will now be proved:

`Exercise 3.5.4.` Suppose that p is a prime $\equiv 1 \pmod 4$ with $p = a^2 + b^2 = c^2 + d^2$ where $a, b, c, d > 0$.

(i) Prove that $(a, b) = (c, d) = 1$.
(ii) Prove that $a/b \equiv c/d$ or $-c/d \pmod p$.
(iii) Assuming that $a/b \equiv c/d \pmod p$ in (ii), use (3.3) to deduce that $p|(ac + bd)$.
(iv) Use (iii) and (3.3) to deduce that $ad = bc$, and then (i) to deduce that $a = c$ and $b = d$.
(v) Work through the case where $a/b \equiv -c/d \pmod p$ using (3.2).

Exercise 3.5.4 tells us that any prime $p \equiv 1 \pmod 4$ can be written as the sum of two squares in a unique way, thus $5 = 1^2 + 2^2$, $13 = 2^2 + 3^2$, $17 = 1^2 + 4^2$ and there are no other representations. For a composite number like 65 we can use the formulae (3.2) and (3.3) to obtain that $65 = 1^2 + 8^2 = 7^2 + 4^2$, and indeed any composite that is the product of two distinct primes $\equiv 1 \pmod 4$ can be written as the sum of two squares in exactly two ways, for examples $85 = 7^2 + 6^2 = 9^2 + 2^2$ and $221 = 13 \cdot 17 = 14^2 + 5^2 = 11^2 + 10^2$. We will discuss the number of representations further.

Theorem 3.6. (Fermat) *Positive integer n can be written as the sum of two squares of integers if and only if for every prime $p \equiv 3 \pmod 4$ which divides n, the exact power of p dividing n is even.*

Proof. exercise

`Exercise 3.5.5.` Deduce that positive integer n can be written as the sum of two squares of *rationals* if and only if n can be written as the sum of two squares of integers.

In section 3.1 we saw how to find all solutions to $x^2 + y^2 = 1$ in rationals x, y. How about all rational solutions to $x^2 + y^2 = n$? It is not difficult to do this in the case that $n = p$ prime, and this argument can be generalized to arbitrary n:

Proposition 3.7. *Suppose that prime p can be written as $a^2 + b^2$. Then all solutions in rationals x, y to $x^2 + y^2 = p$ are given by the parametrization:*

$$x = \frac{2ars + b(s^2 - r^2)}{r^2 + s^2}, \quad y = \frac{2brs + a(r^2 - s^2)}{r^2 + s^2}, \tag{3.4}$$

or the same with b replaced by $-b$.

Proof sketch. Let x, y be rationals for which $x^2+y^2 = p$. Let z be the smallest integer such that $X = xz, Y = yz$ are both integers, so that $X^2+Y^2 = pz^2$. Now $(X, Y)^2|X^2+Y^2 = pz^2$ so that $(X, Y)|z$. Therefore $Z = z/(X, Y)$ is an integer with $X/(X, Y) = xZ,\ Y/(X, Y) = yZ$ both integers implying, by the minimality of z that $Z = z$ and so $(X, Y) = 1$.

Now $X^2 + Y^2 \equiv 0 \pmod p$, and so $(X/Y)^2 \equiv -1 \pmod p$ as $(X, Y) = 1$. But then $X/Y \equiv \pm a/b \pmod p$, say '+', so that $p|(bX - aY)$. Now

$$p^2z^2 = (a^2 + b^2)(X^2 + Y^2) = (aX + bY)^2 + (aY - bX)^2$$

and so $p|(aX + bY)$. Hence $z^2 = ((aX + bY)/p)^2 + ((aY - bX)/p)^2$, and so by (3.1) there exist integers g, r, s such that

$$aX + bY = 2pgrs,\ aY - bX = pg(r^2 - s^2),\ z = g(r^2 + s^2).$$

The result follows.

3.6. The values of $x^2 + dy^2$. How about $x^2 + 2y^2$? We have the identity

$$(a^2 + 2b^2)(c^2 + 2d^2) = (ac + 2bd)^2 + 2(ad - bc)^2,$$

analogous to (3.2), so can focus on what primes are represented. Now if odd prime $p = x^2 + 2y^2$ then $(-2/p) = 1$. On the other hand if $(-2/p) = 1$ then select $b \pmod p$ such that $b^2 \equiv -2 \pmod p$. We take $R = 2^{1/4}\sqrt{p},\ S = 2^{-1/4}\sqrt{p}$ in exercise 3.5.1, so that p divides $r^2 + 2s^2$, which is $\leq 2^{3/2}p < 3p$. Hence $r^2 + 2s^2 = p$ or $2p$. In the latter case $2|2p - 2s^2 = r^2$ so that $2|r$. Writing $r = 2R$ we have $s^2 + 2R^2 = p$. Hence we have proved that p can be written as $m^2 + 2n^2$ if and only if $p = 2$ or $p \equiv 1$ or $3 \pmod 8$.

`Exercise 3.6.1.` What integers can be written as $x^2 + 2y^2$?

`Exercise 3.6.2.` Fix integer $d \geq 1$. Give an identity showing that the product of two integers of the form $a^2 + db^2$ is also of this form.

`Exercise 3.6.3.` Try to determine what primes are of the form $a^2 + 3b^2$, and $a^2 + 5b^2,\ a^2 + 6b^2$, etc.

3.7. Solutions to quadratic equations. It is easy to see that there do not exist non-zero integers a, b, c such that $a^2 + 5b^2 = 3c^2$. For if we take the smallest non-zero solution then we have

$$a^2 \equiv 3c^2 \pmod 5$$

and since $(3/5) = -1$ this implies that $a \equiv c \equiv 0 \pmod 5$ and so $b \equiv 0 \pmod 5$. Therefore $a/5, b/5, c/5$ gives a smaller solution to $x^2 + 5y^2 = 3z^2$, contradicting minimality.

Another proof stems from looking at the equation mod 4 since then $a^2 + b^2 + c^2 \equiv 0 \pmod 4$, and thus $2|a, b, c$ as 0 and 1 are the only squares mod 4, and so $a/2, b/2, c/2$ gives a smaller solution, contradicting minimality.

In general there are an even number of proofs modulo powers of different primes that a given quadratic equation has no solutions if there are none. These are not difficult to identify. On the other hand, what is remarkable, is that if there are no such "mod p^k obstructions", then there are non-zero integer solutions:

The Local-Global Principal for Quadratic Equations. *Let a, b, c be given integers. There are solutions in integers ℓ, m, n to $a\ell^2 + bm^2 + cn^2 = 0$ if and only if there are real numbers λ, μ, ν for which $a\lambda^2 + b\mu^2 + c\nu^2 = 0$, and for all positive integers r there exist residue classes $u, v, w \pmod r$, not all $\equiv 0 \pmod r$, such that $au^2 + bv^2 + cw^2 \equiv 0 \pmod r$.*

Notice the similarity with the Local-Global Principal for Linear Equations given in section 1.1. Just as there, we can restrict our attention to just one modulus r. We may also restrict the set of a, b, c without loss of generality:

`Exercise 3.7.1.` Show that we may assume a, b, c are squarefree, without loss of generality. (Hint: Suppose that $a = Ap^2$ for some prime p and establish a 1-to-1 correspondence with the solutions for A, b, c.)

`Exercise 3.7.2.` Show that we may also assume that a, b, c are pairwise coprime.

The Local-Global Principal for Quadratic Equations. (Legendre, 1785) *Suppose that squarefree non-zero integers a, b, c are pairwise coprime. Then the equation*

$$a\ell^2 + bm^2 + cn^2 = 0$$

has solutions in integers, other than $\ell = m = n = 0$ if and only if $-bc$ is a square mod a, $-ac$ is a square mod b, and $-ab$ is a square mod c, and a, b and c do not all have the same sign.

Again, we can restate this criterion, by asking only for solutions to $a\ell^2 + bm^2 + cn^2 \equiv 0 \pmod{abc}$ with $(\ell mn, abc) = 1$.

Proof $\implies$. If a, b, c all have the same sign, then so do $a\ell^2, bm^2, cn^2$ and then the only solution is $\ell = m = n = 0$. Otherwise, suppose that we have the minimal non-zero solution.

We show that $(m, a) = 1$: If not $p|(m, a)|a\ell^2 + bm^2 = -cn^2$ and so $p|n$ as $(a, c) = 1$. Moreover $p^2|bm^2 + cn^2 = -a\ell^2$ and so $p|\ell$ as a is squarefree. But then $\ell/p, m/p, n/p$ yields a smaller solution, contradicting minimality.

Now $bm^2 \equiv -cn^2 \pmod a$ and as $(m, a) = 1$ there exists r such that $rm \equiv 1 \pmod a$. Therefore $-bc \equiv -bc(rm)^2 = cr^2 \cdot (-bm^2) \equiv cr^2 \cdot cn^2 = (crn)^2 \pmod a$.

An analogous argument works mod b and mod c.

Proof $\Longleftarrow$. Interchanging a, b, c, and multiplying through by -1, as necessary, we can assume that $a, b > 0 > c$.

Suppose that α, β, γ are integers such that

$$\alpha^2 \equiv -bc \pmod{a}, \quad \beta^2 \equiv -ac \pmod{b}, \quad \gamma^2 \equiv -ab \pmod{c}.$$

Construct, using the Chinese Remainder Theorem integers u, v, w for which

$$u \equiv \begin{cases} \gamma \pmod{c} \\ c \pmod{b} \end{cases}, \quad v \equiv \begin{cases} \alpha \pmod{a} \\ a \pmod{c} \end{cases}, \quad w \equiv \begin{cases} \beta \pmod{b} \\ b \pmod{a} \end{cases}.$$

`Exercise 3.7.3.` Working mod a, b, c separately and then using the Chinese Remainder Theorem, verify that

$$au^2 + bv^2 + cw^2 \equiv 0 \pmod{abc};$$

And show that if x, y, z are integers for which $aux + bvy + cwz \equiv 0 \pmod{abc}$ then

$$ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}.$$

Now consider the set of integers

$$\{aui + bvj + cwk : \ 0 \le i \le \sqrt{|bc|}, \ 0 \le j \le \sqrt{|ac|}, \ 0 \le k \le \sqrt{|ab|}\}.$$

The number of i values is $1 + [\sqrt{|bc|}] > \sqrt{|bc|}$; and similarly the number of j and k values, so that the number of elements of the set is $> \sqrt{|bc|} \cdot \sqrt{|ac|} \cdot \sqrt{|ab|} = |abc|$. Hence two different elements of the set are congruent mod abc, say $aui + bvj + cwk \equiv auI + bvJ + cwK \pmod{abc}$. Then $x = i - I, \ y = j - J, \ z = k - K$ are not all zero, and $aux + bvy + cwz \equiv 0 \pmod{abc}$. By the previous exercise we deduce that $ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}$. Now $|x| \le \sqrt{|bc|}, \ |y| \le \sqrt{|ac|}, \ |z| \le \sqrt{|ab|}$ and so

$$-abc = 0 + 0 - abc \le ax^2 + by^2 + cz^2 \le abc + abc + 0 = 2abc.$$

Since $|bc|, |ac|, |ab|$ are squarefree integers by hypothesis, if we get equality in either inequality here then $a = b = 1$, but this case is settled by Theorem 9.3. Hence we may assume that

$$ax^2 + by^2 + cz^2 \equiv 0 \pmod{abc}, \quad \text{and} \quad -abc < ax^2 + by^2 + cz^2 < 2abc,$$

so that $ax^2 + by^2 + cz^2 = 0$ as desired or $ax^2 + by^2 + cz^2 = abc$. The first case gives us the theorem with excellent bounds on the solutions. In the second we make an unintuitive transformation to note that

$$a(xz + by)^2 + b(yz - ax)^2 + c(z^2 - ab)^2 = (z^2 - ab)(ax^2 + by^2 + cz^2 - abc) = 0$$

In 1950, Holzer showed that if there are solutions then the smallest non-zero solution satisfies

$$|a\ell^2|, \ |bm^2|, \ |cn^2| \le |abc|.$$

In 1957, Selmer showed that the Local-Global Principal does not necessarily hold for cubic equations since $3x^3 + 4y^3 + 5z^3 = 0$ has solutions in the reals, and mod r for all $r \ge 1$, yet has no integer solutions. We shall prove this later.

3.8. Pell's equation. Perhaps the most researched equation in the early history of number theory is the so-called Pell's equation: Are there integer solutions x, y to

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

We will show in Theorem 3.8 that the answer is "yes" for any non-square positive integer d. In section * we will see that solutions can always be found using the continued fraction for $\sqrt{d}$. This was evidently known to Brahmagupta in India in 628 A.D., and one can guess that it was well understood by Archimedes, judging by his "Cattle Problem" since, to resolve it, one needs to find a non-trivial solution in integers u, v to

$$u^2 - 609 \cdot 7766v^2 = 1,$$

and the smallest solution has about $2 \cdot 10^6$ digits![9]

Theorem 3.8. *Let $d \geq 2$ be a given non-square integer. There exist integers x, y for which*

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

with $y \neq 0$. If x_1, y_1 are the smallest solutions in positive integers, then all other solutions are given by the recursion $x_{n+1} = x_1x_n + dy_1y_n$ and $y_{n+1} = x_1y_n + y_1x_n$ for $n \geq 1$.

Proof. We begin by showing that there exists a solution with $y \neq 0$. By Corollary 2.1.1, there exists infinitely many pairs of integers (m_j, n_j), $j = 1, 2, \dots$ such that $|\sqrt{d} - \frac{m}{n}| \leq \frac{1}{n^2}$. Therefore

$$|m^2 - dn^2| = n^2 \left|\sqrt{d} - \frac{m}{n}\right| \cdot \left|\sqrt{d} + \frac{m}{n}\right| \leq \left|\sqrt{d} + \frac{m}{n}\right| \leq 2\sqrt{d} + \left|\sqrt{d} - \frac{m}{n}\right| \leq 2\sqrt{d} + 1.$$

Since there are only finitely many possibilities for $m^2 - dn^2$ there must be some integer r, with $|r| \leq 2\sqrt{d} + 1$ such that there are infinitely many pairs of positive integers m, n for which $m^2 - dn^2 = r$.

Since there are only r^2 pairs of residue classes ($m \bmod r$, $n \bmod r$) there must be some pair of residue classes a, b such that there are infinitely many pairs of integers m, n for which $m^2 - dn^2 = r$ with $m \equiv a \pmod r$ and $n \equiv b \pmod r$. Let m_1, n_1 be the smallest such pair, and m, n any other such pair, so that $m_1^2 - dn_1^2 = m^2 - dn^2 = r$ with $m_1 \equiv m \pmod r$ and $n_1 \equiv n \pmod r$. This implies that $r|(m_1n - n_1m)$ and

$$(m_1m - dn_1n)^2 - d(m_1n - n_1m)^2 = (m_1^2 - dn_1^2)(m^2 - dn^2) = r^2,$$

so that $r^2|(r^2 + d(m_1n - n_1m)^2) = (m_1m - dn_1n)^2$ and thus $r|(m_1m - dn_1n)$. Therefore $x = |m_1m - dn_1n|/r$ and $y = |m_1n - n_1m|/r$ are integers for which $x^2 - dy^2 = 1$.

`Exercise 3.8.1.` Show that $y \neq 0$ using the fact that $(m, n) = 1$ for each such pair m, n.

Let x_1, y_1 be the solution in positive integers with $x_1 + \sqrt{d}y_1$ minimal. Note that this is $\geq 1 + \sqrt{d} > 1$. We claim that all other such solutions take the form $(x_1 + \sqrt{d}y_1)^n$.

[9] Presumably Archimedes knew that the smallest solution was ridiculously large and did not ask this question by chance.

If not let x, y be the counterexample with $x + \sqrt{d}y$ smallest, and $X = x_1x - dy_1y$ and $Y = x_1y - y_1x$. Then $X^2 - dY^2 = (x_1^2 - dy_1^2)(x^2 - dy^2) = 1$, and

$$X + \sqrt{d}Y = (x_1 - \sqrt{d}y_1)(x + \sqrt{d}y) = \frac{x + \sqrt{d}y}{x_1 + \sqrt{d}y_1} < x + \sqrt{d}y.$$

Since x, y was the smallest counterexample, hence $X + \sqrt{d}Y = (x_1 + \sqrt{d}y_1)^n$ for some integer $n \geq 1$, and therefore $x + \sqrt{d}y = (x_1 + \sqrt{d}y_1)(X + \sqrt{d}Y) = (x_1 + \sqrt{d}y_1)^{n+1}$

`Exercise 3.8.2.` This proof is not quite complete since we have not yet shown that X and Y are both positive. Remedy this problem. (Proving that $X > 0$ is not difficult, from the fact that $x^2 = dy^2 + 1 > dy^2$. One might prove that $Y > 0$ by establishing that $x_1/y_1 - \sqrt{d} > x/y - \sqrt{d}$.)

One of the fascinating things about Pell's equation is the size of the smallest solution, as we saw in the example given by Archimedes. We will indicate in section *, that the smallest solution is $\leq e^{c\sqrt{d}}$ for some constant $c > 0$. However what is surprising is that usually the smallest solution is really this large. This is not something that has been proved; indeed understanding the distribution of sizes of the smallest solutions to Pell's equation is an outstanding open question in number theory.

Another issue is whether there is a solution to $u^2 - dv^2 = -1$. Notice, for example, that $2^2 - 5 \cdot 1^2 = -1$. Evidently if there is a solution then -1 is a square mod d, so that d has no prime factors $\equiv -1 \pmod 4$. Moreover d is not divisible by 4 else $u^2 \equiv -1 \pmod 4$ which is impossible. We saw that $x^2 - dy^2 = 1$ has solutions for every non-square $d > 1$, and one might have guessed that there would be some simple criteria to decide whether there are solutions to $u^2 - dv^2 = -1$, but there does not appear to be. Even the question of whether there are solutions for "many" d has only recently been resolved by Fouvry and Kluners.

3.9. Descent on solutions of $x^2 - dy^2 = n,\ d > 0$. Let $\epsilon_d = x_1 + y_1\sqrt{d}$, the smallest solution x_1, y_1 in positive integers to $x^2 - dy^2 = 1$.

In fact given any solution of $x^2 - dy^2 = n$ we can pick the signs of x and y so that $x, y \geq 0$. Then $\alpha := x + y\sqrt{d}$ is the largest of the four possibilities, and the four solutions give rise to $\alpha, -\alpha, n/\alpha, -n/\alpha$. We deduce that $\alpha^2 \geq (x + y\sqrt{d})(x - y\sqrt{d}) = n$ and so $\alpha \geq \sqrt{n}$, and indeed α is the unique one of the four that is $> \sqrt{n}$ (when n is not a square).

Now suppose that $\sqrt{n}\epsilon_d^k \leq \alpha < \sqrt{n}\epsilon_d^{k+1}$. We define $\beta := \alpha\epsilon_d^{-k}$, so that $\sqrt{n} \leq \beta < \sqrt{n}\epsilon_d$. Hence every such α is of the form $\beta\epsilon_d^k$, where $\beta = u + iv \in [\sqrt{n}, \sqrt{n}\epsilon_d)$ with $u, v \geq 1$ and $u^2 - dv^2 = n$. This allows to determine the full set of solutions:

Let $B := \{u + \sqrt{d}v \in [\sqrt{n}, \sqrt{n}\epsilon_d) :\ u, v \geq 1 \text{ and } u^2 - dv^2 = n\}$. Then the full set of solutions to $x^2 - dy^2 = n$ is given by $\pm\epsilon_d^{\mathbb{Z}}B$. For $n = 1$ we have $B = \{1\}$. However B can be empty and it can be large. For example, there are no solutions to $x^2 - dy^2 = n$ in integers if n is not a square mod d. In the particular case $x^2 - 5y^2 = 209$, we know that $\epsilon_5 = \left(\frac{1+\sqrt{5}}{2}\right)^6 = 9 + 4\sqrt{5}$, and then $B = \{17 + 4\sqrt{5},\ 47 + 20\sqrt{5}\}$.

4. Binary quadratic forms

4.1. Representation of integers by binary quadratic forms. We have already seen that the integers that can be represented by the binary linear form $ax + by$ are those integers divisible by $\gcd(a, b)$.

`Exercise 4.1.1.` Show that if N can be represented by $ax + by$ then there exist coprime integers m and n such that $am + bn = N$.

Now we let a, b, c be given integers, and ask what integers can be represented by the *binary quadratic form* $ax^2+bxy+cy^2$? That is, for what integers N do there exist coprime integers m, n such that

$$N = am^2 + bmn + cn^2 \quad ? \tag{4.1}$$

We may reduce to the case that $\gcd(a, b, c) = 1$ by dividing though by $\gcd(a, b, c)$. One idea is to complete the square to obtain

$$4aN = (2am + bn)^2 - dn^2$$

where the *discriminant* $d := b^2-4ac$. Hence $d \equiv 0$ or $1 \pmod 4$. When $d < 0$ the right side of the last displayed equation can only take positive values, which makes our discussion easier than when $d > 0$. For this reason we will restrict ourselves to the case $d < 0$ here, and revisit the case $d > 0$ later. In chapter 3 we already worked with a few basic examples, and we will now see how this theory develops.

`Exercise 4.1.2.` Show that if $d < 0$ then $am^2 + bmn + cn^2$ has the same sign as a, no matter what the choices of integers m and n.

We replace a, b, c by $-a, -b, -c$ if necessary, to ensure that the value of $am^2+bmn+cn^2$ is always ≥ 0, and so we call this a *positive definite* binary quadratic form.

`Exercise 4.1.3.` Show that if $ax^2 + bxy + cy^2$ is positive definite then $a, c > 0$.

The key idea stems from the observation that $x^2 + y^2$ represents the same integers as $X^2+2XY+2Y^2$. This is easy to see for if $N = m^2+n^2$ then $N = (m-n)^2+2(m-n)n+2n^2$, and similarly if $N = u^2 + 2uv + 2v^2$ then $N = (u + v)^2 + v^2$. The reason is that the substitution

$$\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} X \\ Y \end{pmatrix} \qquad \text{where } M = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

transforms $x^2 + y^2$ into $X^2 + 2XY + 2Y^2$, and the transformation is invertible, since $\det M = 1$. Much more generally define

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

`Exercise 4.1.4.` Prove that the binary quadratic form $ax^2 + bxy + cy^2$ represents the same integers as the binary quadratic form $AX^2 + BXY + CY^2$ whenever $\begin{pmatrix} x \\ y \end{pmatrix} = M \begin{pmatrix} X \\ Y \end{pmatrix}$ with $M \in \mathrm{SL}(2, \mathbb{Z})$. We say

that these two quadratic forms are *equivalent*. This yields an equivalence relation and splits the binary quadratic forms into equivalence classes.

`Exercise 4.1.5.` Show that two equivalent binary quadratic forms represent each integer in the same number of different ways.

We can write $ax^2+bxy+cy^2 = \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$ and note that the discriminant is -4 times the determinant of $\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$. We deduce that

$$AX^2+BXY+CY^2 = \begin{pmatrix} X & Y \end{pmatrix} M^{\mathrm{T}}\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} M\begin{pmatrix} X \\ Y \end{pmatrix},$$

and so $A = a\alpha^2+b\alpha\gamma+c\gamma^2$ and $C = a\beta^2+b\beta\delta+c\delta^2$ as

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} = M^{\mathrm{T}}\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} M. \tag{4.2}$$

`Exercise 4.1.6.` Use (4.2) to show that two equivalent binary quadratic forms have the same discriminant.

4.2. Equivalence classes of binary quadratic forms. Now $29X^2+82XY+58Y^2$ is equivalent to x^2+y^2 so when we are considering representations, it is surely easier to work with the latter form rather than the former. Gauss observed that every equivalence class of binary quadratic forms (with $d<0$) contains a unique reduced representative, where the quadratic form $ax^2+bxy+cy^2$ with discriminant $d<0$ is *reduced* if

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

For a reduced binary quadratic form, $|d| = 4ac-(|b|)^2 \ge 4a\cdot a - a^2 = 3a^2$ and hence

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

Therefore for a given $d<0$ there are only finitely many a, and so b (as $|b|\le a$), but then $c=(b^2-d)/4a$ is determined, and so there are only finitely many reduced binary quadratic forms of discriminant d. Hence $h(d)$, the *class number*, which is the number of equivalence classes of binary quadratic forms of discriminant d, is finite when $d<0$. In fact $h(d)\ge 1$ since we always have the *principal* form (for both positive and negative discriminants),

$$\begin{cases} x^2-(d/4)y^2 & \text{when } d\equiv 0 \pmod 4, \\ x^2+xy+\frac{(1-d)}{4}y^2 & \text{when } d\equiv 1 \pmod 4. \end{cases}$$

`Exercise 4.2.1.` Show that there are no other binary quadratic forms $x^2+bxy+cy^2$, up to equivalence.

Theorem 4.1. *Every positive definite binary quadratic form is properly equivalent to a reduced form.*

Proof. We will define a sequence of properly equivalent forms; the algorithm terminates when we reach one that is reduced. Given a form (a, b, c):[10]

i) If $c < a$ the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$, yields the form $(c, -b, a)$ which is properly equivalent to (a, b, c).

ii) If $b > a$ or $b \le -a$ then select b' to be the least residue, in absolute value, of b (mod $2a$), so that $-a < b' \le a$, say $b' = b - 2ka$. Hence the transformation matrix will be $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$. The resulting form (a, b', c') is properly equivalent to (a, b, c).

iii) If $c = a$ and $-a < b < 0$ then we use the transformation $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$ yielding the form $(a, -b, a)$.

If the resulting form is not reduced then repeat the algorithm. If none of these hypotheses holds then one can easily verify that the form is reduced. To prove that the algorithm terminates in finitely many steps we follow the leading coefficient a: a starts as a positive integer. Each transformation of type (i) reduces the size of a. It stays the same after transformations of type (ii) or (iii), but after a type (iii) transformation the algorithm terminates, and after a type (ii) transformation we either have another type (i) transformation, or else the algorithm stops after at most one more transformation. Hence the algorithm finishes in no more than $2a + 1$ steps.

Example: Applying the reduction algorithm to the form $(76, 217, 155)$ of discriminant -31, one finds the sequence of forms $(76, 65, 14), (14, -65, 76), (14, -9, 2), (2, 9, 14), (2, 1, 4)$, the sought after reduced form. Similarly the form $(11, 49, 55)$ of discriminant -19, gives the sequence of forms $(11, 5, 1), (1, -5, 11), (1, 1, 5)$.

The very precise condition in the definition of "reduced" were so chosen because every positive definite binary quadratic form is properly equivalent to a *unique* reduced form, which the enthusiastic reader will now prove:

`Exercise 4.2.2.` (i) Show that the least values taken by the reduced form $am^2 + bmn + cn^2$ with $(m, n) = 1$, are $a \le c \le a - |b| + c$, each represented twice (the last four times if $b = 0$). (Hint: One might use the inequality $am^2 + bmn + cn^2 \ge am^2 - |b|\max\{m, n\}^2 + cn^2$, to show that if the value is $am^2 + bmn + cn^2 \le a - |b| + c$ then $|m|, |n| \le 1$.)

(ii) Use this, and exercise 4.1.5, to show that if two different reduced forms are equivalent then they must be $ax^2 + bxy + cy^2$ and $ax^2 - bxy + cy^2$, and thus $a < c$ since these are both reduced.

(iii) Suppose that $M \in \mathrm{SL}(2, \mathbb{Z})$ transforms one into the other. Given that we know all the representations of a and c by $ax^2 + bxy + cy^2$, use (4.2) to deduce that $M = \pm I$.

(iv) Deduce that $b = -b$ so that $b = 0$. Therefore no two reduced forms can be equivalent.

Together with Theorem 4.1 this implies that every positive definite binary quadratic form is properly equivalent to a unique reduced form.

[10] Which we write for convenience in place of $ax^2 + bxy + cy^2$.

What restrictions are there on the values that can be taken by a binary quadratic form?

Proposition 4.2. *Suppose $d = b^2 - 4ac$ with $(a, b, c) = 1$, and p is a prime.* (i) *If $p = am^2 + bmn + cn^2$ for some integers m, n then d is a square mod $4p$.* (ii) *If d is a square mod $4p$ then there exists a binary quadratic form of discriminant d that represents p.*

Proof. (i) Note that $(m, n)^2 | am^2 + bmn + cn^2 = p$ so that $(m, n) = 1$.

Now $d = b^2 - 4ac \equiv b^2 \pmod 4$, and even mod $4p$ if $p|ac$. If $p|d$ then d is a square mod p and the result then follows unless $p = 2$. But if $2|d = b^2 - 4ac$ then b is even; therefore $d = b^2 - 4ac \equiv 0$ or $4 \pmod 8$ and hence is a square mod 8.

If $p = 2 \nmid acd$ then b is odd, and so $am^2 + bmn + cn^2 \equiv m^2 + mn + n^2 \not\equiv 0 \pmod 2$ as $(m, n) = 1$.

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

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

4.3. Class number one.

Corollary 4.3. *Suppose that $h(d) = 1$. Then p is represented by the form of discriminant d if and only if d is a square mod $4p$.*

Proof. This follows immediately from Proposition 4.2, since there is just one equivalence class of quadratic forms of discriminant d, and forms in the same equivalence class represent the same integers by exercise 4.1.4.

The proof that the number of reduced forms is finite can also by turned into an algorithm to find all the reduced binary quadratic forms of a given negative discriminant.

Example: If $d = -163$ then $|b| \le a \le \sqrt{163/3} < 8$. But b is odd so $|b| = 1, 3, 5$ or 7. Therefore $ac = (b^2 + 163)/4 = 41, 43, 47$ or 53, a prime, with $a < c$ and hence $a = 1$. However all such forms are equivalent to the principal form, by exercise 4.3.1, and therefore $h(-163) = 1$. This implies, by Corollary 4.4, that if $(-163/p) = 1$ then p can be represented by the binary quadratic form $x^2 + xy + 41y^2$.

`Exercise 4.3.1.` Determine $h(d)$ for $-20 \le d \le -1$ as well as for $d = -43$ and -67.

Typically one restricts attention to *fundamental discriminants*, which means that if $q^2|d$ then $q = 2$ and $d \equiv 8$ or $12 \pmod{16}$. It turns out that the only fundamental $d < 0$ with $h(d) = 1$ are $d = -3, -4, -7, -8, -11, -19, -43, -67, -163$. Therefore, as in the example above, if $p \nmid d$ then

p is represented by $x^2 + y^2$ if and only if $(-1/p) = 1$,
p is represented by $x^2 + 2y^2$ if and only if $(-2/p) = 1$,
p is represented by $x^2 + xy + y^2$ if and only if $(-3/p) = 1$,
p is represented by $x^2 + xy + 2y^2$ if and only if $(-7/p) = 1$,
p is represented by $x^2 + xy + 3y^2$ if and only if $(-11/p) = 1$,
p is represented by $x^2 + xy + 5y^2$ if and only if $(-19/p) = 1$,

p is represented by $x^2+xy+11y^2$ if and only if $(-43/p)=1$,
p is represented by $x^2+xy+17y^2$ if and only if $(-67/p)=1$,
p is represented by $x^2+xy+41y^2$ if and only if $(-163/p)=1$.

Euler noticed that the polynomial x^2+x+41 is prime for $x=0,1,2,\dots,39$, and similarly the other polynomials above. Rabinowiscz proved that this is an "if and only if" condition; that is

Rabinowiscz's criterion. *We have $h(1-4A)=1$ for $A\geq 2$ if and only if x^2+x+A is prime for $x=0,1,2,\dots,A-2$.*

Note that $(A-1)^2+(A-1)+A=A^2$. We will prove Rabinowiscz's criterion below

The proof that the above list gives all of the $d<0$ with $h(d)=1$ has an interesting history. By 1934 it was known that there is no more than one further such d, but that putative d could not be ruled out by the method. In 1952, Kurt Heegner, a German school teacher proposed an extraordinary proof that there are no further d; at the time his paper was ignored since it was based on a result from an old book (of Weber) whose proof was known to be incomplete. In 1966 Alan Baker gave a very different proof that was acknowledged to be correct. However, soon afterwards Stark realized that the proofs in Weber are easily corrected, so that Heegner's work had been fundamentally correct. Heegner was subsequently given credit for solving this famous problem, but sadly only after he had died. Heegner's paper contains a most extraordinary construction, widely regarded to be one of the most creative and influential in the history of number theory, that we will discuss again at the end of that part of this book on elliptic curves.

What about when the class number is not one? In the first example, $h(-20)=2$, the two reduced forms are x^2+5y^2 and $2x^2+2xy+3y^2$. By Proposition 4.2(i), p is represented by at least one of these two forms if and only if $(-5/p)=0$ or 1, that is, if $p\equiv 1,3,7$ or 9 (mod 20) or $p=2$ or 5. So can we tell which of these primes are represented by which of the two forms? Note that if $p=x^2+5y^2$ then $(p/5)=0$ or 1 and so $p=5$ or $p\equiv\pm 1$ (mod 5), and thus $p\equiv 1$ or 9 (mod 20). If $p=2x^2+2xy+3y^2$ then $2p=(2x+y)^2+5y^2$ and so $p=2$ or $(2p/5)=1$, that is $(p/5)=-1$, and hence $p\equiv 3$ or 7 (mod 20). Hence we have proved

p is represented by x^2+5y^2 if and only if $p=5$, or $p\equiv 1$ or 9 (mod 20);
p is represented by $2x^2+2xy+3y^2$ if and only if $p=2$, or $p\equiv 3$ or 7 (mod 20).

That is, we can distinguish which primes can be represented by which binary quadratic form of discriminant -20 through congruence conditions, despite the fact that the class number is not one. However we cannot always distinguish which primes are represented by which binary quadratic form of discriminant d. It is understood how to recognize those discriminants for which this is the case, indeed these *idoneal numbers* were recognized by Euler. He found 65 of them, and no more are known – it is an open conjecture as to whether Euler's list is complete. It is known that there can be at most one further idoneal number.

Proof of Rabinowiscz's criterion. We begin by showing that $f(n):=n^2+n+A$ is prime for $n=0,1,2,\dots,A-2$, if and only if $d=1-4A$ is not a square mod $4p$ for all primes $p<A$. For if n^2+n+A is composite, let p be its smallest prime factor so that $p\leq f(n)^{1/2}<f(A-1)^{1/2}=A$. Then $(2n+1)^2-d=4(n^2+n+A)\equiv 0$ (mod $4p$) so that d is a square mod $4p$. On the other hand if d is a square mod $4p$ where p is a prime

$\leq A-1$, select n to be the smallest integer ≥ 0 such that $d \equiv (2n+1)^2 \bmod 4p$. Then $0 \leq n \leq p-1 \leq A-2$, and p divides n^2+n+A with $p < A = f(0) < f(n)$ so that n^2+n+A is composite.

Now we show that $h(d)=1$ if and only if $d = 1-4A$ is not a square mod $4p$ for all primes $p < A$. If $h(d) > 1$ then there exists a reduced binary quadratic $ax^2+bxy+cy^2$ of discriminant d with $1 < a \leq \sqrt{|d|/3}$. If p is a prime factor of a then $p \leq a < A$ and $d = b^2-4ac$ is a square mod $4p$. On the other hand if d is a square mod $4p$, and $h(d)=1$ then p is represented by $x^2+xy+Ay^2$ by Proposition 4.2(ii). However the smallest values represented by this form are 1 and A, by exercise 4.2.2(i), and this gives a contradiction since $1 < p < A$. Hence $h(d) > 1$.

4.4. Ideals in quadratic fields. We are going to revisit the notion of equivalence of quadratic forms, by using ideals in quadratic fields. Before doing that, we will need to understand ideals a little better.

We saw that any ideal in $\mathbb{Z}$ may be generated by just one element. We will now prove that any ideal in a quadratic ring of integers:

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

can be generated by at most two integers. Suppose an ideal $I \subset R$ is given. Either $I \subset \mathbb{Z}$ in which case it is a principal ideal, or there exists some element $u+v\sqrt{d} \in I$ with $v \neq 0$. We may assume that $v > 0$ by replacing $u+v\sqrt{d}$ with $-(u+v\sqrt{d})$ if v was negative.

Now select that $r+s\sqrt{d} \in I$ with $s > 0$ minimal. Such an s exists since it must be a positive integer in $\{1,2,\dots|v|\}$. Note that if $u+v\sqrt{d} \in I$ then s divides v, for if not select $k,\ell \in \mathbb{Z}$ for which $ks+\ell v = g := \gcd(s,v)$ and then

$$(kr+\ell u)+g\sqrt{d} = k(r+s\sqrt{d})+\ell(u+v\sqrt{d}) \in I$$

contradicting the minimality of s.

If $u+v\sqrt{d} \in I$ then let $m = v/s$, so that $(u+v\sqrt{d})-m(r+s\sqrt{d}) = u-mr$. Therefore every element of the ideal I may be written as $m(r+s\sqrt{d})+n$ where $n \in I \cap \mathbb{Z}$, and m is an arbitrary integer. Now $I \cap \mathbb{Z}$ is an ideal in $\mathbb{Z}$ so must be principal, generated by some integer $g \geq 1$. Therefore

$$I = \{m(r+s\sqrt{d})+ng : \ m,n \in \mathbb{Z}\} = I_{\mathbb{Z}}(r+s\sqrt{d},g).$$

So we have achieved our goal, I has been shown to be generated by just two elements; and better yet we have proved that we only need to take linear combinations of those two elements with coefficients in $\mathbb{Z}$ to obtain the whole of I. However, we can simplify even more:

Since $\sqrt{d} \in R$, hence $g\sqrt{d} \in I$ and $sd+r\sqrt{d} \in I$, and so s divides both g and r. Therefore $r = sb$ and $g = sa$ for integers a and b. Finally $s(b^2-d) = (r+s\sqrt{d})(b-\sqrt{d}) \in I \cap \mathbb{Z}$ and so $s(b^2-d)$ is a multiple of $g = sa$; hence a divides b^2-d. Therefore

$$I = I_{\mathbb{Z}}(s(b+\sqrt{d}), sa) \ \text{ which we write as } \ s \cdot I_{\mathbb{Z}}(b+\sqrt{d},a),$$

for some integers s,a,b where a divides b^2-d.

Non-principal ideals. Let $R = \mathbb{Z}[\sqrt{-d}]$ with $d \geq 2$. Which ideals $I := I_R(p, r + s\sqrt{-d})$ are principal, where p is a prime in $\mathbb{Z}$ that divides $r^2 + ds^2$, but does not divide s? (This includes the example $I_R(2, 1 + \sqrt{-5})$).

Theorem 4.4.1. *Let $R = \mathbb{Z}[\sqrt{-d}]$ with $d \geq 2$. Suppose that p is a prime in $\mathbb{Z}$ which divides $r^2 + ds^2$ but not s. Then the ideal $I := I_R(p, r + s\sqrt{-d})$ is principal if and only if $p = I_R(a + b\sqrt{-d})$ where $p = a^2 + db^2$ with $a, b \in \mathbb{Z}$ (and $a/b \equiv r/s \pmod p$).*

We will use the following result:

Lemma 4.4.2. *If integer prime p equals the product of two elements of $\mathbb{Z}[\sqrt{-d}]$ then it is either as $(\pm 1) \cdot (\pm p)$, or as $p = (a + b\sqrt{-d})(a - b\sqrt{-d})$ where $p = a^2 + db^2$.*

Proof of Lemma 4.4.2. Suppose that $p = (a + b\sqrt{-d})(u + v\sqrt{-d})$ where $a, b, u, v \in \mathbb{Z}$. Now $\gcd(a, b) \cdot \gcd(u, v)$ divides p, then at least one of these gcds equals 1, say $(a, b) = 1$. Now $p = (au - dbv) + (av + bu)\sqrt{-d}$, so the coefficient of the imaginary part is $av + bu = 0$. Therefore $a|bu$ and therefore $a|u$ as $(a, b) = 1$. Writing $u = ak$ we have $v = -bk$ and therefore $p = k(a + b\sqrt{-d})(a - b\sqrt{-d}) = k(a^2 + db^2)$.

Now $a^2 + db^2$ is a positive divisor of p so must equal either 1 or p. If $b \neq 0$ then $a^2 + db^2 \geq d > 1$, and so if $a^2 + db^2 = 1$ then $a = \pm 1,\ b = 0$. Otherwise $a^2 + db^2 = p$.

Proof of Theorem 4.4.1. Suppose that $I := I_R(p, r + s\sqrt{-d})$ is principal, say, $I = I_R(g)$ where $g \in R$. Then we can write p as the product of two elements of $\mathbb{Z}[\sqrt{-d}]$, including g, and so by the lemma, $g = \pm 1$ or $\pm p$ or $a \pm b\sqrt{-d}$ where $p = a^2 + db^2$. We cannot have $g = \pm p$ for if $r + s\sqrt{-d} = \pm p(u + v\sqrt{-d})$ then we would have that p divides r and s contrary to the hypothesis.

Now let $t \equiv 1/s \pmod p$ and $m \equiv rt \pmod p$ so that $m + \sqrt{-d} \equiv t(r + s\sqrt{-d}) \pmod p$, and $r + s\sqrt{-d} \equiv s(m + \sqrt{-d}) \pmod p$ as $sm \equiv rst \equiv r \pmod p$. Moreover $m^2 + d \equiv t^2(r^2 + ds^2) \equiv 0 \pmod p$, so that $I = I_R(p, m + \sqrt{-d})$. This is presented in the form at the end of the last subsection and so $I = I_{\mathbb{Z}}(p, m + \sqrt{-d})$. Notice that ± 1 is not in this ideal (since it is not divisible by p).

We are left with the only possibility that $g = a \pm b\sqrt{-d}$. In this case $(mb)^2 - a^2 = b^2(m^2 + d) - (a^2 + db^2) \equiv 0 \pmod p$ so that p divides $(mb - a)(mb + a)$. Hence either $m \equiv a/b$ or $-a/b \pmod p$, and we choose the sign of b so that $a \equiv bm \pmod p$. Therefore $a + b\sqrt{-d} \equiv b(m + \sqrt{-d}) \pmod p$ and so $a + b\sqrt{-d} \in I$.

Example: Both $I_R(2, 1 + \sqrt{-5})$ and $I_R(3, 1 + \sqrt{-5})$ are non-principal since there do not exist integers a, b for which $a^2 + 5b^2 = 2$ or 3.

Equivalence of ideals. If $f(x, y) = ax^2 + bxy + cy^2$ then

$$af(x, y) = \left(ax + \frac{b + \sqrt{d}}{2}\ y\right)\left(ax + \frac{b - \sqrt{d}}{2}\ y\right)$$

so we see that $af(x, y)$ is the *Norm* of $\left(ax + \frac{b+\sqrt{d}}{2}\ y\right)$. So the set of possible values of $f(x, y)$ with $x, y \in \mathbb{Z}$ is in 1-to-1 correspondence with the elements of $I_{\mathbb{Z}}(a, \frac{b+\sqrt{d}}{2})$. Now

we have seen that any two equivalent binary quadratic forms can be obtained from each other by a succession of two basic transformations:

$$x \to x+y,\ y \to y \ \text{ which is equivalent to } \ I_{\mathbb{Z}}\left(a, \frac{b+\sqrt{d}}{2}\right) \to I_{\mathbb{Z}}\left(a, \frac{2a+b+\sqrt{d}}{2}\right),$$
$$x \to -y,\ y \to x \ \text{ which is equivalent to } \ I_{\mathbb{Z}}\left(a, \frac{b+\sqrt{d}}{2}\right) \to I_{\mathbb{Z}}\left(c, \frac{-b+\sqrt{d}}{2}\right).$$

It is obvious that $I_{\mathbb{Z}}\left(a, \frac{b+\sqrt{d}}{2}\right) = I_{\mathbb{Z}}\left(a, \frac{2a+b+\sqrt{d}}{2}\right)$. The connection between the two ideals in the second transformation is a little tricker to find: Note that $\frac{-b+\sqrt{d}}{2} \cdot \frac{b+\sqrt{d}}{2} = \frac{d-b^2}{4} = -ac$, and therefore

$$\frac{-b+\sqrt{d}}{2} \cdot I_{\mathbb{Z}}\left(a, \frac{b+\sqrt{d}}{2}\right) = a \cdot I_{\mathbb{Z}}\left(\frac{-b+\sqrt{d}}{2}, -c\right).$$

From this we see that the notion of equivalence of forms, in the setting of ideals, gives that for ideals I, J of $\mathbb{Q}(\sqrt{d})$, we have that $I \sim J$ if and only there exists $\alpha \in \mathbb{Q}(\sqrt{d})$, such that

$$J = \alpha I.$$

This same notion of equivalence works in any number field; moreover in any number field one has finitely many equivalence classes, and indeed bounds for the "smallest" element of any such class (just as in imaginary quadratic fields). We will discuss real quadratic fields in the next chapter.

We have seen that any ideal in the ring of integers of a quadratic field can be written as the $\mathbb{Z}$-linear combinations of a and $\frac{b+\sqrt{d}}{2}$. If $d < 0$ then we can plot these $\mathbb{Z}$-linear combinations on the complex plane and they form a *lattice*, indeed we simply write $\Lambda = \langle a, \frac{b+\sqrt{d}}{2} \rangle$; this rephrasing of the notion of ideal allows us to use ideas of geometry by working with lattices. The notion of equivalence of ideals translates too: Two lattices Λ, Λ' are *homothetic* if there exists $\alpha \in \mathbb{C}$ such that $\Lambda' = \alpha\Lambda$, and we write $\Lambda' \sim \Lambda$. With this notion, we can divide through by a, and thus every such lattice is homothetic to one of the form $\Lambda = \langle 1, \tau \rangle$ where $\tau = \pm\ \frac{b+\sqrt{d}}{2a}$, which is chosen to be in the upper half plane.

Fundamental discriminants and orders. At the very end of section 2 we sorted out the ring of algebraic integers with a given discriminant. However we did not look into the issue of what is the correct discriminant to work with in a given quadratic field. In other words, we know that $\mathbb{Q}(\sqrt{3}) = \mathbb{Q}(\sqrt{12}) = \mathbb{Q}(\sqrt{27}) = \dots$ but the ring of integers associated to each of them is different. What is the "natural" one to work with? For any such square class of integers, like $3, 12, 27, 48, \dots$ there is a minimal one, which is thus the only one that is squarefree, and that would seem like a good candidate. However, since we wish to work with the theory of binary quadratic forms (because of the connection we just discussed) we would like the discriminant to also be the discriminant of a quadratic

form, and we know that an integer is a discriminant of a binary quadratic form if and only if it is $\equiv 0$ or $1 \pmod 4$. In the first case the integers involved will not be squarefree since they are divisible by 4; nonetheless this is the correct price to pay. Thus we select the *fundamental discriminant* of a quadratic field to be the smallest element of the square class of the discriminant which is $\equiv 0$ or $1 \pmod 4$. In particular if our quadratic field of $\mathbb{Q}(\sqrt{d})$, with d a squarefree integer, then the fundamental discriminant D is given by

$$D = \begin{cases} d & \text{if } d \equiv 1 \pmod 4 \\ 4d & \text{if } d \equiv 2 \text{ or } 3 \pmod 4 \end{cases}.$$

The ring of integers is $\mathbb{Z}[\omega] = \langle 1, \omega \rangle_{\mathbb{Z}}$ where

$$\omega := \begin{cases} \frac{1+\sqrt{d}}{2} & \text{if } d \equiv 1 \pmod 4 \\ \sqrt{d} = \sqrt{D/2} & \text{if } d \equiv 2 \text{ or } 3 \pmod 4 \end{cases}.$$

An *order* is a subring of the ring of integers of an imaginary quadratic field.

`Exercise 4.4.1.`(i) Prove that every order takes the form $\mathbb{Z}[f\omega] = \{m + nf\omega : \ m, n \in \mathbb{Z}\}$ for some integer $f \geq 1$, and that these orders are all distinct.

(ii) Prove that if $\alpha \in \mathbb{Z}[f\omega]$ then $\overline{\alpha} \in \mathbb{Z}[f\omega]$.

4.5. Composition laws. We have seen in (3.2) that if we multiply together any two values of the quadratic form $x^2 + y^2$ then we get a third value of that quadratic form. Gauss observed a general rule, if f and g are two quadratic forms of discriminant d, then there exists another quadratic form h of discriminant d, such that any value of f, times any value of g, gives a value of h. Gauss's approach was simply to show this explicitly via an algebraic formula. For example, for three quadratic forms of discriminant -71 we have

$$2x^2 + xy + 9y^2 = (4a^2 + 3ab + 5b^2)(3r^2 + rs + 6s^2).$$

by taking $x = ar - 3as - 2br - 3bs$ and $y = ar + as + br - bs$. Gauss called this *composition*, and went on to show that this action stays consistent under the equivalence relation, and in fact composition allows us to find a group structure on the classes of quadratic forms of given discriminant, the *class group*. Gauss's proof is monstrously difficult, even in the hands of the master the algebra involved is so overwhelming that he does not include many details. It was his student Dirichlet who found several ways to simplify composition. The first involved finding forms that are equivalent to f and g that are easier to compose:

`Exercise 4.5.1.` Let $ax^2 + bxy + cy^2$ be a *primitive* quadratic form, that is $\gcd(a, b, c) = 1$.

(i) Prove that for any given integer A there exist non-zero integers m, n with $(am^2 + bmn + cn^2, A) = 1$.

(ii) Deduce that for any given quadratic forms f and g, with f primitive, one can find $h \sim f$ such that $h(1, 0)$ and $g(1, 0)$ are coprime non-zero integers.

(iii) Suppose that we are given quadratic forms f and g for which $f(1, 0)$ and $g(1, 0)$ are coprime non-zero integers. Show that there exists $F \sim f$ and $G \sim g$ such that $F(x, y) = ax^2 + bxy + cy^2$ and $G(x, y) = Ax^2 + bxy + Cy^2$ with $(a, A) = 1$.

(iv) In (iii) suppose that f and g have the same discriminant. Then show that there exists an integer h such that $F(x, y) = ax^2 + bxy + Ahy^2$ and $G(x, y) = Ax^2 + bxy + ahy^2$ with $(a, A) = 1$.

From exercise 4.5.1 if we are given two primitive quadratic forms of the same discriminant d, then there are forms in their equivalence classes as in (iv), where $d = b^2 - 4aAh$. Now one can verify that if $H(x, y) = aAx^2 + bxy + hy^2$ then

$$H(ux - hvy, auy + Avx + bvy) = F(u, v)G(x, y)$$

Dirichlet went on from here to note how much more easily one could define composition using ideals: To multiply two ideals one simply multiplies the elements of the two ideal together, thus $IJ = \{ij : \ i \in I, \ j \in J\}$. This can be most easily done in terms of the bases, multiplying the bases elements together. Thus in our example, the ideals $(4, \frac{3+\sqrt{-71}}{2})$ corresponds to $4a^2 + 3ab + 5b^2$, and $(3, \frac{1+\sqrt{-71}}{2})$ corresponds to $3r^2 + rs + 6s^2$. Now, the product is

$$\begin{aligned}\left(4, \frac{3+\sqrt{-71}}{2}\right)\left(3, \frac{1+\sqrt{-71}}{2}\right) &= \left(12, 2+2\sqrt{-71}, \frac{9+3\sqrt{-71}}{2}, \frac{3+\sqrt{-71}}{2} \cdot \frac{1+\sqrt{-71}}{2}\right) \\ &= \left(12, 2+2\sqrt{-71}, \frac{5-\sqrt{-71}}{2}, -17+\sqrt{-71}\right) \\ &= \left(12, 12, \frac{5-\sqrt{-71}}{2}, -12\right) = \left(12, \frac{-5+\sqrt{-71}}{2}\right),\end{aligned}$$

which corresponds to the quadratic form $12x^2 - 5xy + 2y^2$, which is also of discriminant -71, but not reduced. The reduction algorithm then yields: $(12, -5, 2) \sim (2, 5, 12) \sim (2, 1, 9)$.

For the two quadratic forms F and G of exercise 4.5.1 we multiply the ideals $\left(a, \frac{-b+\sqrt{d}}{2}\right)$ and $\left(A, \frac{-b+\sqrt{d}}{2}\right)$. The product contains both $a \cdot \frac{-b+\sqrt{d}}{2}$ and $A \cdot \frac{-b+\sqrt{d}}{2}$, and hence it also contains $(a, A) \cdot \frac{-b+\sqrt{d}}{2} = \frac{-b+\sqrt{d}}{2}$. Therefore the product of the two ideals is $\left(aA, \frac{-b+\sqrt{d}}{2}\right)$, which yields that the composition of F and G must be $aAx^2 + bxy + hy^2$.

`Exercise 4.5.2.` Suppose that I_f is the ideal corresponding to the binary quadratic form f. Show that if f, g and h are quadratic forms of the same discriminant, and $I_f I_g = I_h$, then any value of f, times any value of g, gives a value of h. (Hint: Develop the discussion given above, and use the way that we have linked ideals to forms.)

It is much more obvious that multiplication of ideals leads to a group structure than Gauss's composition idea. One finds that the identity of the ideal class group is given by the principal ideas. Moreover

$$\left(a, \frac{b+\sqrt{d}}{2}\right)\left(a, \frac{b-\sqrt{d}}{2}\right) = \left(a^2, a\,\frac{b+\sqrt{d}}{2}, a\,\frac{b-\sqrt{d}}{2}, \frac{b^2-d}{4}\right) \supseteq a\,(a, b, c) = (a),$$

if the quadratic form is *primitive* (that is $(a, b, c) = 1$), so that an ideal and its conjugate are inverses in the class group.

Recently Bhargava gave a new insight into the composition law. First note that if $IJ = K$ then $IJ\overline{K}$ is principal (where $\overline{K}$ is the conjugate of K), so we seek to understand this relationship.

4.6. Bhargava composition. We begin with a 2-by-2-by-2 cube. a, b, c, d, e, f, g, h. There are six faces, which can be split into three parallel pairs. To each such parallel pair consider the pair of 2-by-2 matrices given by taking the entries in each face, with corresponding entries corresponding to opposite corners of the cube, always starting with a. Hence we get the pairs

$$M_1(x,y) := \begin{pmatrix} a & b \\ c & d \end{pmatrix} x + \begin{pmatrix} e & f \\ g & h \end{pmatrix} y,$$
$$M_2(x,y) := \begin{pmatrix} a & c \\ e & g \end{pmatrix} x + \begin{pmatrix} b & d \\ f & h \end{pmatrix} y,$$
$$M_3(x,y) := \begin{pmatrix} a & b \\ e & f \end{pmatrix} x + \begin{pmatrix} c & d \\ g & h \end{pmatrix} y,$$

where we have, in each added the dummy variables, x, y. The determinant, $-Q_j(x,y)$, of each $M_j(x,y)$ gives rise to a quadratic form in x and y. Now we will apply an $\mathrm{SL}(2,\mathbb{Z})$ transformation in one direction. That is, if $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ then we replace the face

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ by } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha + \begin{pmatrix} e & f \\ g & h \end{pmatrix} \beta; \text{ and } \begin{pmatrix} e & f \\ g & h \end{pmatrix} \text{ by } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \gamma + \begin{pmatrix} e & f \\ g & h \end{pmatrix} \delta.$$

Then $M_1(x,y)$ gets mapped to

$$\left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha + \begin{pmatrix} e & f \\ g & h \end{pmatrix} \beta \right\} x + \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \gamma + \begin{pmatrix} e & f \\ g & h \end{pmatrix} \delta \right\} y,$$

that is $M_1(\alpha x + \gamma y, \beta x + \delta y)$. Therefore the quadratic form $Q_1(x,y)$ gets mapped to $Q_1(\alpha x + \gamma y, \beta x + \delta y)$ which is equivalent to $Q_1(x,y)$. Now $M_2(x,y)$ gets mapped to

$$\begin{pmatrix} a\alpha + e\beta & c\alpha + g\beta \\ a\gamma + e\delta & c\gamma + g\delta \end{pmatrix} x + \begin{pmatrix} b\alpha + f\beta & d\alpha + h\beta \\ b\gamma + f\delta & d\gamma + h\delta \end{pmatrix} y = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} M_2(x,y);$$

hence the determinant, $Q_2(x,y)$, is unchanged. An analogous calculation reveals that $M_3(x,y)$ gets mapped to $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} M_3(x,y)$ and the determinant, $Q_3(x,y)$, is unchanged.

Therefore we can act on our cube by such $\mathrm{SL}(2,\mathbb{Z})$-transformations, in each direction, and each of the three quadratic forms remains in the same equivalence class.

We will simplify the entries in the cube by the following reduction algorithm:

• We select the corner that is to be a so that $a \neq 0$.

• We will transform the cube to ensure that a divides b, c and e. If not, say a does not divide e, then select integers α, β so that $a\alpha + e\beta = (a,e)$, and then let $\gamma = -e/(a,e)$, $\delta = a/(a,e)$. In the transformed matrix we have $a' = (a,e)$, $e' = 0$, and $1 \leq a' \leq a-1$. It may well now be that a' does not divide b' or c', so we repeat the process. Each time we do this we reduce the value of a by at least 1; and since it remains positive this can only happen a finite number of times. At the end of the process a divides b, c and e.

• We will transform the cube to ensure that $b = c = e = 0$. We already have that $a|b, c, e$. Now select $\alpha = 1,\ \beta = 0, \gamma = -e/a,\ \delta = 1$, so that $e' = 0, b' = b, c' = c$. We repeat this in each of the three directions to ensure that $b = c = e = 0$.

Replacing a by $-a$, we have that the three matrices are:

$$M_1(x,y) := \begin{pmatrix} -a & 0 \\ 0 & d \end{pmatrix} x + \begin{pmatrix} 0 & f \\ g & h \end{pmatrix} y, \text{ so that } Q_1(x,y) = adx^2 + ahxy + fgy^2;$$

$$M_2(x,y) := \begin{pmatrix} -a & 0 \\ 0 & g \end{pmatrix} x + \begin{pmatrix} 0 & d \\ f & h \end{pmatrix} y, \text{ so that } Q_2(x,y) = agx^2 + ahxy + dfy^2;$$

$$M_3(x,y) := \begin{pmatrix} -a & 0 \\ 0 & f \end{pmatrix} x + \begin{pmatrix} 0 & d \\ g & h \end{pmatrix} y, \text{ so that } Q_3(x,y) = afx^2 + ahxy + dgy^2.$$

All three Q_j have discriminant $(ah)^2 - 4adfg$, and we observe that

$$Q_1(fy_2x_3 + gx_2y_3 + hy_2y_3, ax_2x_3 - dy_2y_3) = Q_2(x_2, y_2)Q_3(x_3, y_3)$$

where $x_1 = fy_2x_3 + gx_2y_3 + hy_2y_3$ and $y_1 = ax_2x_3 - dy_2y_3$. This is composition in the sense defined by Gauss. In fact if $a = 1$ then we are back with the composition of forms defined by Dirichlet, as in the last subsection. In particular this shows that Bhargava's construction allows us to compose any two primitive quadratic forms of the same discriminant.

4.7. The idoneal numbers. Suppose that prime $p = am^2 + bmn + cn^2$, where $ax^2 + bxy + cy^2$ is a reduced form with negative fundamental discriminant d. Then d is a square mod p by Proposition 4.2. We now prove that if $N = am^2 + bmn + cn^2$ and q is an odd prime dividing d but not N, then $(N/q) = (a/q)$, or (c/q) if $q|a$. Note first that q cannot divide both a and c else $q|d + 4ac = b^2$ so that $q|(a, b, c)$. Now $4aN = (2am + bn)^2 - dn^2 \equiv (2am + bn)^2 \pmod{d}$, so that if $q \mid a$ then $(N/q) = (a/q)$. Similarly if $q \not| c$ then $(N/q) = (c/q)$. Note that since two equivalent forms represent the same integers the values of (N/q) depend only on the equivalence class of quadratic forms, not on the form itself.

Suppose that d is odd (so $\equiv 1 \pmod 4$ and squarefree, as d is fundamental), and $d = -q_1q_2 \dots q_k$ where the q_j are distinct odd primes. If N is represented by f and is coprime to d then we just proved that the vector $\mathcal{L}(f) = ((N/q_1), (N/q_2), \dots, (N/q_k)) \in \{-1, 1\}^k$ is independent of N, but depends only on f. Moreover if $N = p$ is prime then

$$\left(\frac{p}{q_1}\right)\left(\frac{p}{q_2}\right)\cdots\left(\frac{p}{q_k}\right) = \left(\frac{p}{-d}\right) = \left(\frac{d}{p}\right) = 1$$

so we know that the product of the elements of $\mathcal{L}(f)$ equals 1. Hence $\mathcal{L}(f) \in V_k := \{(\delta_1, \dots, \delta_k) \in \{-1, 1\}^k : \ \delta_1 \cdots \delta_k = 1\}$, which contains 2^{k-1} vectors.

Now if p is a prime for which $(p/q_j) = \delta_j$ for each j, where $(\delta_1, \dots, \delta_k) \in V_k$ then by Proposition 4.2(ii) we know that p is represented by some form of discriminant d, and hence it must be a form f for which $\mathcal{L}(f) = (\delta_1, \dots, \delta_k)$. We also know by Dirichlet's Theorem on primes in arithmetic progressions[11] that for every $(\delta_1, \dots, \delta_k) \in V_k$ there exists such a prime. In other words the map $\mathcal{L} : \ \mathcal{C} \to V_k$ is surjective, where $\mathcal{C}$ is the ideal class group.

[11] Dirichlet proved, in 1839, that if $(a, q) = 1$ then there are infinitely many primes $\equiv a \pmod q$.

In the special case that there is just one binary quadratic form for each element of V_k then we call d an *idoneal* number. In this case we know exactly what primes are represented by the given quadratic form, like that example x^2+5y^2 and $2x^2+2xy+3y^2$ given in section 4.3.

Now if m is represented by f and n is represented by g then mn is represented by $f*g$ (the composition of f and g), we see that $\mathcal{L}(f*g)=\mathcal{L}(f)\mathcal{L}(g)$ multiplying component wise (as $(m/q)(n/q)=(mn/q)$). This also tells us that $\mathcal{L}(f*f)$ is the vector of all 1s, and so if $f=g*h^2$ then $\mathcal{L}(f)=\mathcal{L}(g)$. Hence the value of $\mathcal{L}(f)$ depends only on the coset that f belongs to in $\mathcal{C}/\mathcal{C}^2$, which we call the *genus* of f; and so we have that the map $\mathcal{L}:\ \mathcal{C}/\mathcal{C}^2\to V_k$ is surjective. Note that each genus contains $|\mathcal{C}^2|$ elements, and hence has just one element (which must be the identity element) if and only if $f*f=1$ for every $f\in\mathcal{C}$.

In section 2.8 we saw that $|\mathcal{C}/\mathcal{C}^2|=\#\{f\in\mathcal{C}:\ f*f=1\}$; and $f*f=1$ if and only if f is equivalent to its inverse. In section 4.5 we saw that an ideal and its conjugate are inverses in the ideal class group, and so if $f=(a,b,c)$ then a form in the ideal class that is the inverse of that of f is $f^{-1}=(a,-b,c)$. We wish to know when $(a,b,c)\sim(a,-b,c)$.

`Exercise 4.7.1.` We will prove that the only primitive, reduced forms of order dividing 2, are $(a,0,c)$ with $a\le c$ and $(a,c)=1$; (a,a,c) with $a\le c$ and $(a,c)=1$; and (a,b,a) with $1\le b<a$ and $(a,b)=1$.

i. Prove that each of these forms is equivalent to its conjugate by explicitly giving the linear transformation in each case.
ii. Show that every reduced binary quadratic form (a,b,c) with $a=c$ is equivalent to its conjugate.
iii. Use exercise 4.2.2(i) to determine the possible maps from reduced (a,b,c) to $(a,-b,c)$ when $a<c$.

`Exercise 4.7.2.` We will determine $\#\{f\in\mathcal{C}:\ f*f=1\}$.

i. Show that the forms $(a,0,c)$ with $a\le c$ and $(a,c)=1$, are in 1-to-1 correspondence with the factorizations $d/4=ac$ with $a<c$ and $(a,c)=1$.
ii. Show that the forms (a,a,c) with $a\le c$ and $(a,c)=1$, are in 1-to-1 correspondence with the factorizations $d=rs$ with $s>3r$, $r\equiv s\pmod 2$ and $(r,s)=1$.
iii. Show that the forms (a,b,a) with $1\le b<a$ and $(a,b)=1$, are in 1-to-1 correspondence with the factorizations $d=rs$ with $r<s<3r$, $r\equiv s\pmod 2$ and $(r,s)=1$.
iv. Use (ii) and (iii) to show that if d is odd then $\#\{f\in\mathcal{C}:\ f*f=1\}=2^{\omega(d)-1}$, where $\omega(d)$ denotes the number of distinct prime factors of d.
v. Use (i), (ii) and (iii) to show that if $4|d$ then $\#\{f\in\mathcal{C}:\ f*f=1\}=2^{\omega(d/4)}$.

By exercise 4.7.2(iv), if d is odd with k prime factors, then $|\mathcal{C}/\mathcal{C}^2|=2^{k-1}=V_k$ and, as $\mathcal{L}:\ \mathcal{C}/\mathcal{C}^2\to V_k$ is surjective, we deduce that it is a bijection. In other words to each genus we can associate a distinct element of V_k.

When $d=-4p_1\dots p_k$ with $-d/4\equiv 1\pmod 4$ then $\left(\frac{p}{q_1}\right)\cdots\left(\frac{p}{q_k}\right)=\left(\frac{-1}{p}\right)$, so we can re-define V_k by adding an extra component for $(-1/p)$ and the previous proof goes through using exercise 4.7.2(v).

When $d=-8p_1\dots p_k$ then $\left(\frac{p}{q_1}\right)\cdots\left(\frac{p}{q_k}\right)=\left(\frac{\pm 2}{p}\right)$, when $d\equiv\pm 8\pmod{32}$, and so we can re-define V_k by adding an extra component for $(\pm 2/p)$ and the previous proof goes through using exercise 4.7.2(v).

Euler found the 65 idoneal numbers $-d=1,2,3,4,5,6,7,8,9,10,12,13,15,16,18,\dots$, $330,345,357,385,408,462,520,760,840,1320,1365,1848$, and it is conjectured that there are no more. It is enough to verify that these are idoneal numbers — simply compute all

of the reduced primitive binary quadratic forms of discriminant $-d$ and verify that they all have one of the special forms given in exercise 4.7.1 (or that the class number is given by the number of genera, as determined in exercise 4.7.1).

One can identify idoneal numbers using rather less theory. Given one of these quadratic forms f, we wish to prove that it represents exactly those primes p for which the values of (p/q) are given by $\mathcal{L}(f)$ (and thus are exactly those primes in some union of arithmetic progressions mod $4d$). So let p be such a prime. We know that d is a square mod p so, by exercise 3.5.1, we know that there exist integers r, s with $r/s \equiv (\sqrt{d}-b)/2a \pmod{p}$ where $|r| \le (c/a)^{1/4}p^{1/2}$ and $|s| \le (a/c)^{1/4}p^{1/2}$. Then $ar^2+brs+cs^2 \equiv 0 \pmod{p}$ and $ar^2+brs+cs^2 \le (2(ac)^{1/2}+|b|)p$, so that $kp = ar^2+brs+cs^2$ with $1 \le k \le 2(ac)^{1/2}+|b|$. But since kp is represented by this form we use the argument from the first paragraph of this subsection to note that $(kp/q) = (p/q)$, so that $(k/q) = 1$ for all odd primes q dividing d, and there is an analogous condition mod 4 or 8 when d is even. We leave it as an exercise to the reader to develop this fully.

4.8. $\mathbf{SL}(2,\mathbb{Z})$-transformations. Forms-Ideals-Transformations.

Generators of $\mathbf{SL}(2,\mathbb{Z})$. We will show that $\mathrm{SL}(2,\mathbb{Z})$ is generated by the two elements $\begin{pmatrix}1&1\\0&1\end{pmatrix}, \begin{pmatrix}0&1\\-1&0\end{pmatrix}$. Given $\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}$ of determinant 1, we shall perform the Euclidean algorithm on α/γ when $\gamma \neq 0$: Select integer a so that $\gamma' := \alpha - a\gamma$ has the same sign as α and $0 \le |\gamma'| < \gamma$. If $\alpha' = -\gamma$ then $\begin{pmatrix}0&-1\\1&-a\end{pmatrix}\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \begin{pmatrix}\alpha'&\beta'\\\gamma'&\delta'\end{pmatrix}$. Other than the signs this is the same process as the Euclidean algorithm, and we reduce the size of the pair of numbers in the first column. Moreover the matrix $\begin{pmatrix}0&-1\\1&-a\end{pmatrix}$ has determinant 1, and therefore so does $\begin{pmatrix}\alpha'&\beta'\\\gamma'&\delta'\end{pmatrix}$. We repeat this process as long as we can; evidently this is impossible once $\gamma = 0$. In that case α and δ are integers for which $\alpha\delta = 1$ and therefore our matrix is $\pm I$. Hence we have that there exists integers $a_1, a_2, \dots, a_k$ such that

$$\begin{pmatrix}0&-1\\1&-a_k\end{pmatrix}\begin{pmatrix}0&-1\\1&-a_{k-1}\end{pmatrix}\cdots\begin{pmatrix}0&-1\\1&-a_1\end{pmatrix}\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \pm\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

Now $\begin{pmatrix}a&-1\\1&0\end{pmatrix}\begin{pmatrix}0&-1\\1&-a\end{pmatrix} = -I$, and so we deduce that

$$\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix} = \pm\begin{pmatrix}a_1&-1\\1&0\end{pmatrix}\begin{pmatrix}a_2&-1\\1&0\end{pmatrix}\cdots\begin{pmatrix}a_k&-1\\1&0\end{pmatrix}$$

Now $\begin{pmatrix}a&-1\\1&0\end{pmatrix} = -\begin{pmatrix}1&a\\0&1\end{pmatrix}\begin{pmatrix}0&1\\-1&0\end{pmatrix} = -\begin{pmatrix}1&1\\0&1\end{pmatrix}^a\begin{pmatrix}0&1\\-1&0\end{pmatrix}$

`Exercise 4.8.1.` Complete the proof that $\mathrm{SL}(2,\mathbb{Z})$ is generated by the two elements $\begin{pmatrix}1&1\\0&1\end{pmatrix}, \begin{pmatrix}0&1\\-1&0\end{pmatrix}$.

We consider the binary quadratic form $f(x,y) := ax^2 + bxy + cy^2$. We saw that two forms f and g are equivalent, written $f \sim g$, if there exists $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ such that $g(x,y) = f(\alpha x + \beta y, \gamma x + \delta y)$.

The root $z_f := \frac{-b+\sqrt{d}}{2a}$ of f is a point in $\mathbb{C}$, the sign of $\sqrt{d}$ chosen, when $d < 0$, to be in the upper half plane. Two points in the complex plane z, z' are said to be equivalent if there exists $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ such that $z' = u/v$ where $\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} z \\ 1 \end{pmatrix}$. Hence $z \sim z+1$ and $z \sim -1/z$.

The ideal $I_f := (2a, -b+\sqrt{d})$ corresponds to f. Note that $I_f = 2a(1, z_f)$. Two ideals I, J are said equivalent if there exists $\alpha \in \mathbb{Q}(\sqrt{d})$ such that $J = \alpha I$. Hence $I_f \sim (1, z_f)$.

The generators of $\mathrm{SL}(2,\mathbb{Z})$ correspond to two basic operations in Gauss's reduction algorithm for binary quadratic forms:

The first is $x \to x+y, y \to y$, so that

$$f(x,y) \sim g(x,y) := f(x+y,y) = ax^2 + (b+2a)xy + (a+b+c)y^2.$$

Note that $I_g = (2a, -(b+2a) + \sqrt{d}) = I_f$, and $z_g = \frac{-b-2a+\sqrt{d}}{2a} = z_f - 1$.

The second is $x \to y,\ y \to -x$ so that

$$f(x,y) \sim h(x,y) := f(y,-x) = cx^2 - bxy + ay^2.$$

Note that $I_h = (2c, b+\sqrt{d})$, and $z_h = \frac{b+\sqrt{d}}{2c}$. First observe that

$$z_f \cdot z_h = \frac{-b+\sqrt{d}}{2a} \cdot \frac{b+\sqrt{d}}{2c} = \frac{d-b^2}{4ac} = -1$$

that is $z_h = -1/z_f$. Then

$$I_h \sim (1, z_h) = (1, -1/z_f) \sim (1, -z_f) = (1, z_f) \sim I_f.$$

Since any $\mathrm{SL}(2,\mathbb{Z})$-transformation can be constructed out of the basic two transformation we deduce

Theorem 4.8.1. *$f \sim f'$ if and only if $I_f \sim I_{f'}$ if and only if $z_f \sim z_{f'}$.*

It is amazing that this fundamental, non-trivial, equivalence can be understood in three seemingly very different ways. Which is the best? That is hard to say; each has their uses, but what is good one can translate any question into the setting in which it is most natural. For example the notion of reduced binary quadratic form seems a little unnatural; however in the context of points in the upper half plane it translates to the points

$$D := \{z \in \mathbb{C} :\ \mathrm{Im}(z) > 0,\ -\frac{1}{2} \leq \mathrm{Re}(z) < \frac{1}{2},\ |z| \geq 1,\ \text{if } |z| = 1 \text{ then } \mathrm{Re}(z) \leq 0.\}$$

Be careful here; we are out by a factor of 2, and we might wish to place z on the right not the left

More on the action of SL$(2,\mathbb{Z})$ *on* $\mathbb{C}$. Henceforth let $\mathcal{H} := \{z \in \mathbb{C} : \ \mathrm{Im}(z) > 0\}$ be the "upper half plane" of $\mathbb{C}$. Let $M \in \mathrm{SL}(2,\mathbb{R})$, that is $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $ad - bc = 1$. Now M acts on $\mathbb{C} \cup \{\infty\}$ by $Mz = \frac{az+b}{cz+d}$.

`Exercise 4.8.2.` (i) Show that $\mathrm{Im}(Mz) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$. (ii) Deduce that $M : \mathcal{H} \to \mathcal{H}$. (iii) Show that $(-M)z = Mz$ for all z but, otherwise there exists τ such that $M\tau \neq N\tau$ for given $M, N \in \mathrm{SL}(2,\mathbb{R})$. Hence our group of transformations is really $\mathrm{PSL}(2,\mathbb{R}) \cong \mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$.

`Exercise 4.8.3.` Define $\mathrm{PSL}(2,\mathbb{Z}) \cong \mathrm{SL}(2,\mathbb{Z})/\{\pm 1\}$. Let $S := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $T := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. i) Prove that $S^2 = (ST)^3 = 1$. ii) Prove that the free group G of words in S and U, where $S^2 = U^3 = 1$ is isomorphic to $\mathrm{PSL}(2,\mathbb{Z})$.

We saw above that $\mathcal{H}/\mathrm{PSL}(2,\mathbb{Z}) \cong D$.

There are subgroups of $\mathrm{PSL}(2,\mathbb{Z})$ that will play a role, for example, $\Gamma_0(N)$ is the subgroup of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2,\mathbb{Z})$ for which $c \equiv 0 \pmod N$.

`Exercise 4.8.4.` Prove that $\Gamma_0(N)$ is a subgroup. Can you define any other subgroups of $\mathrm{PSL}(2,\mathbb{Z})$ that depend only on the congruence classes of the entries mod N?

4.9. The ring of integers of a quadratic field, revisited. Are there any integer solutions x, y to $x^2 + 19 = y^3$? Let us suppose that there are. Note that y is odd else $x^2 \equiv 5 \pmod 8$ which is impossible. Also 19 does not divide y else 19 divides x but then $19 \equiv x^2 + 19 = y^3 \equiv 0 \pmod{19^2}$. Hence $(y, 38) = 1$.

Now $(x+\sqrt{-19})(x-\sqrt{-19}) = y^3$ and the ideal $I(x+\sqrt{-19}, x-\sqrt{-19})$ contains their difference, which is $2\sqrt{-19}$, as well as y^3, and so contains $(y^3, 38) = 1$. Hence the ideals $(x+\sqrt{-19})$ and $(x-\sqrt{-19})$ are coprime, and their product is a cube and so they are both cubes. We proved, in section 4.3 that the ring of integers of $\mathbb{Q}[\sqrt{-19}]$ has class number one. This means that every ideal is principal, and thus we can write $x+\sqrt{-19} = u(a+b\sqrt{-19})^3$ where u is a unit. Now the only units in this ring of integers are 1 and -1 by exercise 2.9.1, so we can remove u by changing a and b, to ua and ub. Hence we can equate both sides of

$$x + \sqrt{-19} = (a + b\sqrt{-19})^3 = a(a^2 - 57b^2) + b(3a^2 - 19b^2)\sqrt{-19},$$

so that $b(3a^2 - 19b^2) = 1$. Therefore $b = \pm 1$ and so $3a^2 = 19b^2 \pm 1 = 19 \pm 1$ which is impossible. We deduce that there are no integer solutions x, y to $x^2 + 19 = y^3$.

However when we observe that $18^2 + 19 = 7^3$ we see that there is a problem with the above proof. What is the mistake??

The mistake comes in assuming that the ring of integers here is the set of numbers of the form $a + b\sqrt{-19}$ with $a, b \in \mathbb{Z}$. In fact it is the set of numbers $(a + b\sqrt{-19})/2$ with $a, b \in \mathbb{Z}$ and $a \equiv b \pmod 2$. Substituting this in to the argument above we obtain that $b(3a^2 - 19b^2) = 8$. Therefore $b = \pm 1, \pm 2, \pm 4$ or ± 8 and so $3a^2 = 19 \pm 8, 19 \cdot 4 \pm 4, 19 \cdot 16 \pm 2$ or $19 \cdot 64 \pm 1$; the only solution is $b = 1, a = \pm 3$ leading to $x = \mp 18,\ y = 7$, the only solutions.

5. Real quadratic fields

5.1. Quadratic irrationals and periodic continued fractions. Suppose that we have a solution to Pell's equation, that is $p^2-dq^2 = \pm 4$ with $p, q > 0$. Therefore $|\sqrt{d}+p/q| > \sqrt{d}$ so that

$$\left|\sqrt{d} - \frac{p}{q}\right| = \frac{|p^2 - dq^2|}{q^2(\sqrt{d} + p/q)} < \frac{4}{\sqrt{d}q^2} .$$

If $d \geq 64$ then this $< 1/2q^2$ and so p/q is a convergent for $\sqrt{d}$.

`Exercise 5.1.1.` Show that if $0 < p^2 - dq^2 \leq \sqrt{d}$ with $p, q \geq 1$ then p/q is a convergent for $\sqrt{d}$.

`Exercise 5.1.2.` Suppose that $p, q \geq 1$ and $-\sqrt{d} \leq p^2 - dq^2 < 0$. Show that $-1 < p/q - \sqrt{d} < 0$. Deduce that if $0 < -(p^2 - dq^2) \leq \sqrt{d} - \frac{1}{2}$ then p/q is a convergent for $\sqrt{d}$.

In exercise 2.3.8 saw that the continued fraction for $\frac{1+\sqrt{5}}{2}$ is just 1 repeated infinitely often. What are the values of continued fractions in which the entries are periodic? We use the notation $\alpha = [\overline{a_0,\ a_1,\ \ldots,\ a_n}]$ to mean

$$\alpha = [a_0,\ a_1, \ldots,\ a_n,\ a_0,\ a_1, \ldots,\ a_n,\ a_0,\ a_1, \ldots,\ a_n, \ldots]$$

is periodic with period $a_0,\ a_1,\ \ldots,\ a_n$. This means that $\alpha = [a_0,\ a_1,\ a_2,\ \ldots,\ a_n,\ \alpha]$; that is $\alpha_{n+1} = \alpha$ and so, as in (2.3.2),

$$\alpha = \frac{\alpha p_n + p_{n-1}}{\alpha q_n + q_{n-1}} . \tag{5.1}$$

This implies that $q_n\alpha^2 + (q_{n-1} - p_n)\alpha - p_{n-1} = 0$, that is α satisfies a quadratic equation. This equation must be irreducible, that is α is irrational, else the continued fraction would be of finite length (as we saw in section 1.2).

If $\gamma = [\overline{a_n,\ a_{n-1},\ \ldots,\ a_0}]$ then, since

$$\begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_{n-1} & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} = \left(\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}\cdots\begin{pmatrix} a_k & 1 \\ 1 & 0 \end{pmatrix}\right)^T = \begin{pmatrix} p_k & q_k \\ p_{k-1} & q_{k-1} \end{pmatrix},$$

hence $\gamma = \frac{\gamma p_n + q_n}{\gamma p_{n-1} + q_{n-1}}$ and so $p_{n-1}\gamma^2 + (q_{n-1} - p_n)\gamma - q_n = 0$, which implies that $q_n(-1/\gamma)^2 + (q_{n-1} - p_n)(-1/\gamma) - p_{n-1} = 0$; that is $-1/\gamma$ satisfies the same quadratic equation as α. However these are two distinct roots since both $\alpha > 0 > -1/\gamma$. We call $-1/\gamma$ the *conjugate* of α.

It may be that $\alpha = [a_0,\ a_1,\ a_2,\ \ldots,\ a_m, \overline{b_0,\ b_1,\ \ldots,\ b_n}]$ is eventually periodic. In that case $\beta := [\overline{b_0,\ b_1,\ \ldots,\ b_n}]$ is quadratic irrational, and hence so is $\alpha = \frac{\beta p_m + p_{m-1}}{\beta q_m + q_{m-1}}$.

Let us suppose that $\alpha = u + v\sqrt{d}$, with d squarefree, has a periodic continued fraction of period m. Then (5.1) is satisfied whenever n is a multiple of m. Hence

$$u + v\sqrt{d} = \alpha = \frac{p_n - q_{n-1} + \sqrt{(q_{n-1} - p_n)^2 + 4p_{n-1}q_n}}{2q_n}, \tag{5.2}$$

so that $(q_{n-1}+p_n)^2+4(-1)^n=(q_{n-1}-p_n)^2+4p_{n-1}q_n=d(2q_nv)^2$. Since the left side is an integer, so is the right side, and so we have infinitely many solutions to *Pell's equation*

$$x^2-dy^2=\pm 4,$$

with $x=q_{n-1}+p_n$ and $y=2q_nv$.

A continued fraction $\beta=[b_0,\ \dots,\ b_m,\overline{a_0,\ a_1,\ \dots,\ a_n}]$, for any $m\geq 0$ is called *eventually periodic*. Note that if $\alpha=[\overline{a_0,\ a_1,\ \dots,\ a_n}]$ then

$$\beta\ =\frac{\alpha p_m+p_{m-1}}{\alpha q_m+q_{m-1}}\ .$$

Theorem 5.1. *Any quadratic irrational real number has a continued fraction that is eventually periodic.*

Proof. Suppose that α has minimal polynomial $ax^2+bx+c=a(x-\alpha)(x-\beta)$, and define

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

By (5.1), $\begin{pmatrix}\alpha\\ 1\end{pmatrix}=\kappa\begin{pmatrix}p_n & p_{n-1}\\ q_n & q_{n-1}\end{pmatrix}\begin{pmatrix}\alpha_{n+1}\\ 1\end{pmatrix}$ for some $\kappa\neq 0$, and so if we define

$$\begin{pmatrix}A_n & B_n/2\\ B_n/2 & C_n\end{pmatrix}:=\begin{pmatrix}p_n & q_n\\ p_{n-1} & q_{n-1}\end{pmatrix}\begin{pmatrix}a & b/2\\ b/2 & c\end{pmatrix}\begin{pmatrix}p_n & p_{n-1}\\ q_n & q_{n-1}\end{pmatrix}$$

(so that $b^2-4ac=B_n^2-4A_nC_n$ by taking determinants of both sides) then

$$\begin{aligned}A_n\alpha_{n+1}^2+B_n\alpha_{n+1}+C_n&=\begin{pmatrix}\alpha_{n+1} & 1\end{pmatrix}\begin{pmatrix}A_n & B_n/2\\ B_n/2 & C_n\end{pmatrix}\begin{pmatrix}\alpha_{n+1}\\ 1\end{pmatrix}\\ &=\begin{pmatrix}\alpha_{n+1} & 1\end{pmatrix}\begin{pmatrix}p_n & q_n\\ p_{n-1} & q_{n-1}\end{pmatrix}\begin{pmatrix}a & b/2\\ b/2 & c\end{pmatrix}\begin{pmatrix}p_n & p_{n-1}\\ q_n & q_{n-1}\end{pmatrix}\begin{pmatrix}\alpha_{n+1}\\ 1\end{pmatrix}\\ &=\kappa^2\begin{pmatrix}\alpha & 1\end{pmatrix}\begin{pmatrix}a & b/2\\ b/2 & c\end{pmatrix}\begin{pmatrix}\alpha\\ 1\end{pmatrix}=\kappa^2 f(\alpha,1)=0.\end{aligned}$$

Therefore $f_n(x):=A_nx^2+B_nx+C_n$ has root α_{n+1}. Now $A_n=f(p_n,q_n)$ and $C_n=A_{n-1}$. By exercise 2.3.6, $\left|\alpha-\frac{p_n}{q_n}\right|<\frac{1}{q_n^2}\leq 1$, and $\left|\beta-\frac{p_n}{q_n}\right|\leq|\beta-\alpha|+\left|\alpha-\frac{p_n}{q_n}\right|\leq|\beta-\alpha|+1$, so that

$$|A_n|=|f(p_n,q_n)|=aq_n^2\left|\alpha-\frac{p_n}{q_n}\right|\left|\beta-\frac{p_n}{q_n}\right|\leq a(|\beta-\alpha|+1),$$

Since the A_n are all integers, there are only finitely many possibilities for the values of A_n and C_n. Moreover, given these values there are only two possibilities for B_n, as $B_n^2=b^2-4ac+4A_nC_n$. Hence there are only finitely many possible triples $f_n(x)$ and each corresponds to at most two roots, so one such root must repeat infinitely often. That is, there exists $m<n$ such that $\alpha_m=\alpha_n$.

`Exercise 5.1.3.` Deduce that the continued fraction for α is eventually periodic.

Proposition 5.2. *Suppose that α is a real quadratic irrational number with conjugate β. Then α has a periodic continued fraction if and only if $\alpha > 1$ and $0 > \beta > -1$.*

Proof. By Theorem 5.1 the continued fraction of α is eventually periodic. This implies that each $\alpha_n > a_n \geq 1$ for all $n \geq 1$ and we now show that if $0 > \beta > -1$ then $0 > \beta_n > -1$ for all $n \geq 1$ by induction. Since $\alpha_{n-1} = a_{n-1} + 1/\alpha_n$ by definition, we have $\beta_{n-1} = a_{n-1} + 1/\beta_n$ by taking conjugates. This means that $a_{n-1} = -1/\beta_n + \beta_{n-1}$ is an integer in $(-1/\beta_n - 1, -1/\beta_n)$ and so $a_{n-1} = [-1/\beta_n]$ and hence $-1/\beta_n > 1$ implying that $0 > \beta_n > -1$. Since the continued fraction for α is periodic, there exists $0 \leq m < n$ with $\alpha_m = \alpha_n$; select m to be the minimal integer ≥ 0 for which such an n exists. Then $m = 0$ else taking conjugates gives $\beta_m = \beta_n$, so that $a_{m-1} = [-1/\beta_m] = [-1/\beta_n] = a_{n-1}$ and hence $\alpha_{m-1} = a_{m-1} + 1/\alpha_m = a_{n-1} + 1/\alpha_n = \alpha_{n-1}$, contradicting the minimality of m.

On the other hand if the continued fraction is purely periodic of period n then, as above $f(x) := q_n x^2 + (q_{n-1} - p_n)x - p_{n-1} = 0$ for $x = \alpha$ and β. Now $f(0) = -p_{n-1} < 0$ and $f(-1) = (q_n - q_{n-1}) + (p_n - p_{n-1}) > 0$, and so f has a root in $(-1, 0)$. This root cannot be α which is $\geq a_0 = a_n \geq 1$ so must be β.

5.2. Pell's equation. Here are some examples of the continued fraction for $\sqrt{d}$:

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

These examples seem to suggest that $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n}]$ where $a_n = 2a_0 = 2[\sqrt{d}]$. Let us suppose, for now, that this is true, so that $\sqrt{d} + [\sqrt{d}]$ and $1/(\sqrt{d} - [\sqrt{d}])$ are (purely) periodic.

`Exercise 5.2.1.` Show that $\sqrt{d} + [\sqrt{d}]$ is indeed periodic.

If $\sqrt{d} = [a_0, a_1, \dots]$ then $\sqrt{d} + [\sqrt{d}] = [\overline{2a_0, a_1, \dots, a_{n-1}}]$ for some n, by exercise 5.2.1, so that $\sqrt{d} = [a_0, \overline{a_1, \dots, a_n}]$ where $a_n = 2a_0$ (as suggested by the examples). Now if P_k/Q_k are the convergents for $\sqrt{d} + [\sqrt{d}]$ then we deduce from the coefficients in (5.2) that

$$(Q_{n-2} + P_{n-1})^2 - d(2Q_{n-1})^2 = 4(-1)^n \quad \text{and} \quad P_{n-1} - Q_{n-2} = 2a_0 Q_{n-1}.$$

`Exercise 5.2.2.` Show that $P_k/Q_k = p_k/q_k + a_0$ for all k (Hint: Use matrices to evaluate the P_k, Q_k, p_k, q_k.); that is $Q_k = q_k$ and $P_k = p_k + a_0 q_k$. Deduce from this and the last displayed equation that $Q_{n-2} + P_{n-1} = 2p_{n-1}$ and so

$$p_{n-1}^2 - dq_{n-1}^2 = (-1)^n.$$

Hence we have seen, in exercise 5.1.1, that if $d \geq 64$ and $p^2 - dq^2 = \pm 4$ with $p, q \geq 1$ then p/q is a convergent to $\sqrt{d}$. Now we see that each period of the continued fraction of $\sqrt{d}$ gives rise to another solution of the Pell equation.

If one takes a slightly larger example like $\sqrt{43} = [6, \overline{1, 1, 3, 1, 5, 1, 3, 1, 1, 12}]$ one cannot help but notice that the period is symmetric, that is $a_j = a_{n-j}$ for $j = 1, 2, \dots, n-1$. To prove this is straightforward: At the beginning of section 5.1 we saw that if we

have $\gamma = [\overline{a_{n-1},\ a_{n-2},\ \dots,\ a_1,\ 2a_0}]$, then $-1/\gamma$ is the conjugate of $\sqrt{d}+[\sqrt{d}]$, that is $1/\gamma = \sqrt{d}-[\sqrt{d}]$ and therefore

$$\begin{aligned}[\overline{2a_0, a_1,\ \dots,\ a_{n-1}}] &= \sqrt{d}+[\sqrt{d}] = 2a_0 + 1/\gamma \\ &= [2a_0, \overline{a_{n-1},\ \dots,\ a_1,\ 2a_0}] = [\overline{2a_0, a_{n-1},\ \dots,\ a_1}].\end{aligned}$$

Remark: We have yet to show that the solutions to Pell's equation are precisely those that come from the period.

5.3. The size of solutions to Pell's equation. As in the proof of Theorem 5.1 but now noting that $\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{a_n q_n^2}$ since $q_{n+1} = a_n q_n + q_{n-1} > a_n q_n$, we obtain that

$$1 \le |A_n| = aq_n^2 \left|\alpha - \frac{p_n}{q_n}\right| \left|\beta - \frac{p_n}{q_n}\right| \le aq_n^2 \cdot \frac{1}{q_n q_{n+1}} \left(|\beta-\alpha|+1\right) < \frac{(a+\sqrt{d})}{a_n};$$

as $(a(\beta-\alpha))^2 = b^2 - 4ac = d$. For $\alpha = \sqrt{d}+[\sqrt{d}]$ we have $a = 1$, and so $1 \le a_n \le 2\sqrt{d}+1$, for all $n \ge 1$. This allows us to deduce upper and lower bounds on p_n and q_n:

`Exercise 5.3.` Suppose that $x_{n+1} = a_n x_n + x_{n-1}$ for all $n \ge 1$, with x_0, x_1 positive integers, not both 0.

(i) Use that each $a_n \ge 1$ to deduce that $x_n \ge F_n$ for all $n \ge 0$.

(ii) Use that each $a_n \le B$ $(= \sqrt{d}+1)$ to deduce that $x_n \le (B+1)^{n-1}(x_1+x_0)$ for all $n \ge 1$.

Hence if the continued fraction for $\sqrt{d}$ has period ℓ and this gives rise to a solution x, y to Pell's equation, then $\phi^\ell \ll \epsilon_d := x + y\sqrt{d} \ll (\sqrt{d}+1)^\ell$ where $\phi = \frac{1+\sqrt{5}}{2}$. Hence there is a direct link between the size of the smallest solution to Pell's equations and the length of the continued fraction.

5.4. More examples of Pell's equation. Here we give only the longest continued fractions and the largest fundamental solutions.

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

$$\sqrt{94} = [9,\ \overline{1,2,3,1,1,5,1,8,1,5,1,1,3,2,1,18}], \quad 2143295^2 - 94 \cdot 221064^2 = 1$$
$$\sqrt{124} = [11,\ \overline{7,2,1,1,1,3,1,4,1,3,1,1,1,2,7,22}], \quad 4620799^2 - 124 \cdot 414960^2 = 1$$
$$\sqrt{133} = [11,\ \overline{1,1,7,5,1,1,1,2,1,1,1,5,7,1,1,22}], \quad 2588599^2 - 133 \cdot 224460^2 = 1$$
$$\sqrt{139} = [11,\ \overline{1,3,1,3,7,1,1,2,11,2,1,1,7,3,1,3,1,22}], \quad 77563250^2 - 139 \cdot 6578829^2 = 1.$$

These are the champions up to 150. After that we list the continued fraction lengths and the fundamental solutions for the champions up to 1000:

$$\text{Length} = 20: \quad 1728148040^2 - 151 \cdot 140634693^2 = 1$$
$$\text{Length} = 22: \quad 1700902565^2 - 166 \cdot 132015642^2 = 1$$
$$\text{Length} = 26: \quad 278354373650^2 - 211 \cdot 19162705353^2 = 1$$
$$\text{Length} = 26: \quad 695359189925^2 - 214 \cdot 47533775646^2 = 1$$
$$\text{Length} = 26: \quad 5883392537695^2 - 301 \cdot 339113108232^2 = 1$$
$$\text{Length} = 34: \quad 2785589801443970^2 - 331 \cdot 153109862634573^2 = 1$$
$$\text{Length} = 37: \quad 44042445696821418^2 - 421 \cdot 2146497463530785^2 = -1$$
$$\text{Length} = 40: \quad 840560915469529337752 - 526 \cdot 3665019757324295532^2 = 1$$
$$\text{Length} = 42: \quad 1811243550616307861302 - 571 \cdot 7579818350628982587^2 = 1$$
$$\text{Length} = 44: \quad 5972991296311683199^2 - 604 \cdot 243037569063951720^2 = 1$$
$$\text{Length} = 48: \quad 48961575312998650035560^2 - 631 \cdot 19491295375751510364272 = 1$$
$$\text{Length} = 52: \quad 72933184667948824244189602 - 751 \cdot 2661369706772060244567932 = 1$$
$$\text{Length} = 54: \quad 77435245930576558516377652 - 886 \cdot 2601487964640241948503782 = 1$$
$$\text{Length} = 60: \quad 448160301093711945155126372^2 - 919 \cdot 14783444239653675978149958^2 = 1$$
$$\text{Length} = 60: \quad 37951640090681193063801489608^2 - 991 \cdot 12055735790331359447442538767^2 = 1$$

Notice that the length of the continued fractions here are around $2\sqrt{d}$, and the size of the fundamental solutions $10^{\sqrt{d}}$.

5.5. Binary quadratic forms with positive discriminant, and continued fractions. When $d > 0$, Gauss defined $ax^2 + bxy + cy^2$ to be *reduced* when

$$0 < \sqrt{d} - b < 2|a| < \sqrt{d} + b. \tag{5.5.1}$$

This implies that $0 < b < \sqrt{d}$ so that $|a| < \sqrt{d}$ and therefore there are only finitely many reduced forms of positive discriminant d. Note that $ax^2 + bxy + cy^2$ is reduced if and only if $cx^2 + bxy + ay^2$ is. The first inequality implies that $ac = (b^2 - d)/4 < 0$.

Let $\rho_1 := \frac{-b+\sqrt{d}}{2a}$ and $\rho_2 := \frac{-b-\sqrt{d}}{2a}$ be the two roots of $at^2 + bt + c = 0$. Then (5.5.1) holds if and only if $|\rho_1| < 1 < |\rho_2|$ and $\rho_1\rho_2 < 0$.

Forms $ax^2+bxy+cy^2$ and $cx^2+b'xy+c'y^2$ are *neighbours* (and equivalent) if they have the same discriminant and $b+b' \equiv 0 \pmod{2c}$, since they are equivalent under the transformation $\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ where $b+b'=2ck$.

The reduction algorithm proceeds as follows: Given $ax^2+bxy+cy^2$ we select a neighbour as follows: Let b_0' be the least residue in absolute value of $-b \pmod{2c}$ so that $|b_0'| \le c$.

- If $|b_0'| > \sqrt{d}$ then let $b'=b_0'$. Note that $0 < (b')^2 - d \le c^2 - d$ so that $|c'| = ((b')^2-d)/4|c| < |c|/4$.
- If $|b_0'| < \sqrt{d}$ then select $b' \equiv -b \pmod{2c}$ with b' as large as possible so that $|b'| < \sqrt{d}$. Note that $-d \le (b')^2 - d = 4cc' < 0$. If $2|c| > \sqrt{d}$ then $|c'| \le |d/4c| < |c|$.

Otherwise $\sqrt{d} \ge 2|c|$ and $\sqrt{d}-2|c| < |b'| < \sqrt{d}$, and therefore the neighbour is reduced. Thus we see that the absolute values of the coefficients a and c of the binary quadratic form are reduced at each step of the algorithm until we obtain a reduced form.

The major difference between this, the $d>0$ case, and the $d<0$ case is that there is not necessarily a unique reduced form in a given class of binary quadratic forms of positive discriminant. Rather, when we run Gauss's algorithm we eventually obtain a *cycle* of reduced forms, which must happen since every reduced form has a unique right and a unique left reduced neighbouring form, and there are only finitely many reduced forms. Given a quadratic form $a_0x^2+b_0xy+a_1y^2$ we define a sequence of forms, in the following notation:

$$a_0 \ {}^{b_0} \ a_1 \ {}^{b_1} \ a_2 \ {}^{b_2} \ a_3 \ \dots \ .$$

This represents, successively, the forms $a_0x^2+b_0xy+a_1y^2$, $a_1x^2+b_1xy+a_2y^2$, $a_2x^2+b_2xy+a_3y^2$, $\dots$, of equal discriminant, where a form is the unique reduced right neighbour of its predecessor, and then $a_{i+1} = (b_i^2-d)/4a_i$. For example, when $d=816$,

$$5 \ {}^{26} \ -7 \ {}^{16} \ 20 \ {}^{24} \ -3 \ {}^{24} \ 20 \ {}^{16} \ -7 \ {}^{26} \ 5 \ {}^{24} \ -12 \ {}^{24} \ 5 \ {}^{26} \ -7 \ \dots$$

which is a cycle of period 8.

A solution to Pell's Equation, $v^2-dw^2 = \pm 4$ yields a map $\begin{pmatrix} X \\ Y \end{pmatrix} \to \begin{pmatrix} \frac{v-bw}{2} & -cw \\ aw & \frac{v+bw}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$, for which $aX^2+bXY+cY^2 = \pm(ax^2+bxy+cy^2)$, which is an automorphism only when $v^2-dw^2=4$. Any solution to Pell's Equation yields a good rational approximation $\frac{v}{w}$ to $\sqrt{d}$, in fact with $|\frac{v}{w} - \sqrt{d}| < \frac{1}{2w^2}$ if $d \ge 19$. This implies that $\frac{v}{w}$ is a convergent for the continued fraction of $\sqrt{d}$. For $\alpha = \sqrt{d}$ let $c_n := p_n^2 - dq_n^2$, so that $c_nc_{n+1} < 0$ and $|c_n| < 2\sqrt{d}+1$, and that there is a cycle of reduced forms $c_0 \ {}^{b_0} \ c_1 \ {}^{b_1} \ c_2 \ {}^{b_2} \ c_3 \ \dots$ of discriminant d. For example $\sqrt{60} = [7, \overline{1,2,1,14}]$ gives rise to the cycle $-11 \ {}^4 \ 4 \ {}^4 \ -11 \ {}^7 \ 1 \ {}^7 \ -11 \ {}^4 \ 4$, and the first 4 corresponds to the unit $\frac{8+\sqrt{60}}{2} = 4+\sqrt{15}$. In general if $\frac{p_n}{q_n}$ is the nth convergent to $\frac{\sqrt{d}-b}{2|a|}$ then define $c_n = ap_n^2 \pm bp_nq_n + cq_n^2$ where $\pm$ represents the sign of a, and we have such a cycle. For example $\frac{\sqrt{97}-9}{8} = [0, \overline{9,2,2,1,4,4,1,2,2}]$, which gives the cycle $-1 \ {}^9 \ 4 \ {}^7 \ -3 \ {}^5 \ 6 \ {}^7 \ -2 \ {}^9 \ 2 \ {}^7 \ -6 \ {}^5 \ 3 \ {}^7 \ -4 \ {}^9 \ 1 \ {}^9 \ -4 \ {}^7 \ 3 \ {}^5 \ -6 \ {}^7 \ 2 \ {}^9 \ -2 \ {}^7 \ 6 \ {}^5 \ -3 \ {}^7 \ 4 \ {}^9 \ -1 \ {}^9 \ 4 \dots$.

The *fundamental unit* is that solution $\epsilon_d := \frac{v_0+\sqrt{d}w_0}{2}$ which is minimal and > 1. We call $\frac{v^2-dw^2}{4}$ the *norm* of ϵ_d. All other solutions of (4.5.2) take the form

$$\frac{v+\sqrt{d}w}{2} = \pm\epsilon_d^k, \tag{5.5.1}$$

for some $k \in \mathbb{Z}$ (for a proof see exercise 4.5c). We let ϵ_d^+ be the smallest unit > 1 with norm 1. One can deduce from (5.5.1) that $\epsilon_d^+ = \epsilon_d$ or ϵ_d^2, depending on whether the norm of ϵ_d is 1 or -1.

`Exercise 5.5.2.` Prove that every reduced form of positive discriminant has a unique right and a unique left reduced neighbouring form.

6. L-functions and class numbers

This is not a book about analytic methods in number theory, but we do need to touch on analytic questions in order to best understand algebraic ones. In this chapter we will develop the start of theory of Dirichlet L-functions, and connect this will class numbers of quadratic fields.

6.1. Counting solutions to Pell's equation (a heuristic). Let's study how many solutions there are to $y^2 - dx^2 = 1$ with $|x| \le N$, for N large. Now if x, y are a pair of positive integers for which $0 \le y^2 - dx^2 < 2$ then $y^2 - dx^2 = 1$ so we could guess the number of such pairs is the volume of this region. Given x it is more-or-less true that $0 \le y^2 - dx^2 < 2$ is equivalent to $0 \le y - \sqrt{d}x < 1/\sqrt{d}x$, and hence the volume with $x \le N$ is

$$\approx \int_1^N \frac{1}{\sqrt{d}} \frac{\mathrm{d}x}{x} = \frac{\log N}{\sqrt{d}} .$$

Now this heuristic pre-supposes that any x, y of the right size are going to be solutions, but we should try to take into account what we know from congruences. That is that the proportion of pairs of integers x, y for which $x^2 - dy^2 \equiv 1 \pmod p$ is not $1/p$ but rather $1/p$ times $1 - \frac{1}{p} \left(\frac{d}{p}\right)$ as we will see in section *. Hence we should multiply the above through by

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p} \left(\frac{d}{p}\right)\right).$$

What is this quantity? We work with the inverse since

$$\left(1 - \frac{1}{p} \left(\frac{d}{p}\right)\right)^{-1} = 1 + \frac{1}{p} \left(\frac{d}{p}\right) + \frac{1}{p^2} \left(\frac{d}{p^2}\right) + \dots ,$$

so if we multiply over all primes, and forget convergence issues then, by the Fundamental Theorem of Arithmetic,

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p} \left(\frac{d}{p}\right)\right)^{-1} = \sum_{n \ge 1} \frac{1}{n} \left(\frac{d}{n}\right).$$

If we replaced $1/p$ here by $1/p^s$ for any s with $\mathrm{Re}(s) > 1$ then we do not have convergence issues, so we define the Dirichlet L-function for the character $\left(\frac{d}{\cdot}\right)$ by

$$L\left(s, \left(\frac{d}{\cdot}\right)\right) := \sum_{n \ge 1} \left(\frac{d}{n}\right) \frac{1}{n^s} \quad \text{for } \mathrm{Re}(s) > 1.$$

Although this is not absolutely convergent when $\mathrm{Re}(s) \in (0, 1]$ it is convergent in this range if we sum the terms in order (i.e. treat $\sum_{n \ge 1}$ as $\lim_{N \to \infty} \sum_{1 \le n \le N}$).

Hence we might predict that the number of solutions to $y^2 - dx^2 = 1$ with $|x| \le N$, for N large is roughly

$$\frac{\log N}{\sqrt{d}\, L(1, \left(\frac{d}{\cdot}\right))} .$$

However this is not quite correct, since the quadratic form $y^2 - dx^2$ is just one of those of discriminant $4d$, and there are actually $h(4d)$ of them. Hence we might expect that the number of solutions is actually more like

$$\frac{h(4d)\log N}{\sqrt{d}\, L(1, \left(\frac{d}{\cdot}\right))}.$$

We did see earlier that all solutions to $y^2 - dx^2 = 1$ with $x, y \geq 1$ are powers of the fundamental solutions ϵ_d to Pell's equation, and hence the number of solutions is actually

$$\frac{\log N}{\log \epsilon_d}.$$

Equating the two we might thus guess that

$$h(4d)\log \epsilon_d \approx \sqrt{d}\, L\left(1, \left(\frac{d}{\cdot}\right)\right).$$

Surprisingly, given the flimsy justification for this heuristic argument, this formula is actually more-or-less true:

6.2. Dirichlet's class number formula. In 1832 Jacobi conjectured that the class number $h(-p)$, when $p \equiv 3 \pmod 4$, is given by

$$h(-p) = \frac{1}{p} \sum_{n=1}^{p-1} \left(\frac{n}{p}\right) n.$$

`Exercise 6.2.1.` Show that the right side is an integer.

`Exercise 6.2.2.` Let $S := \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right)$ and $T := \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right) n$.

(1) Show that $S = 0$ when $p \equiv 1 \pmod 4$. Henceforth assume that $p \equiv 3 \pmod 4$.
(2) Note that $\left(\frac{p-n}{p}\right)(p-n) = \left(\frac{n}{p}\right)(n-p)$. Use this to evaluate the sum $\sum_{n=1}^{p-1} \left(\frac{n}{p}\right) n$ in terms of S and T by pairing up the nth and $(p-n)$th term, for $n = 1, 2, \ldots, \frac{p-1}{2}$.
(3) Do this taking $n = 2m,\ m = 1, 2, \ldots, \frac{p-1}{2}$ to deduce that

$$h(-p) = \frac{1}{\left(\frac{2}{p}\right) - 2} \sum_{n=1}^{(p-1)/2} \left(\frac{n}{p}\right).$$

In 1838 Dirichlet gave a proof of Jacobi's conjecture and much more. His miraculous class number formula links algebra and analysis in an unforeseen way that was foretaste of many of the most important works in number theory, including Wiles' proof of Fermat's Last Theorem. We will simply state the formulae here: If $d > 0$ and d is not a square then

$$h(d)\, \log \epsilon_d = \sqrt{d}\, L\left(1, \left(\frac{d}{\cdot}\right)\right).$$

If $d<-4$ then

$$h(d) \ = \frac{1}{\pi} \ \sqrt{|d|}\ L\left(1, \left(\frac{d}{\cdot}\right)\right).$$

Note that $h(d)\geq 1$ for all d since we always have the principal form. Hence the formulae imply that $L\left(1, \left(\frac{d}{\cdot}\right)\right) > 0$ for all d, (which is needed in the proof that there are infinitely many primes in arithmetic progressions). In fact these formulae even give lower bounds; for example when $d<-4$ we have $L\left(1, \left(\frac{d}{\cdot}\right)\right) \geq \pi/\sqrt{|d|}$. Getting a significantly better lower bound for all d is a very difficult problem, though Heilbronn showed that there exists a constant $c>0$ such that $L\left(1, \left(\frac{d}{\cdot}\right)\right) \geq c/\log|d|$ (and hence $h(d) > c'\sqrt{|d|}/\log|d|$) with very few exceptions (in fact no more than one value of d in the range $D<d\leq 2D$ for any D).

The size of a fundamental unit. We know that $\sum_{n=4k|d|}^{4(k+1)|d|}\left(\frac{d}{n}\right)=0$ if d is not a square. Hence

$$\left|\sum_{n=4k|d|+1}^{4(k+1)|d|}\frac{1}{n}\left(\frac{d}{n}\right)\right| \leq \sum_{n=4k|d|+1}^{4(k+1)|d|}\frac{1}{4k|d|}\left(\frac{d}{n}\right)+\sum_{n=4k|d|+1}^{4(k+1)|d|}\left|\frac{1}{4k|d|}-\frac{1}{n}\right| \leq \frac{4|d|\cdot 4|d|}{4k|d|\cdot 4(k+1)|d|},$$

and so $|L(1, \left(\frac{d}{\cdot}\right))| \leq \sum_{n\leq 4|d|}\frac{1}{n}+1 \leq \log|d|+O(1)$. Therefore, for $d>0$, since $h(d)\geq 1$,

$$\log\epsilon_d \leq h(d)\log\epsilon_d \ = \ \sqrt{d}\ L\left(1, \left(\frac{d}{\cdot}\right)\right) \leq \sqrt{d}(\log d + O(1)).$$

Hence $\epsilon_d \leq (Cd)^{\sqrt{d}}$ for some constant $C>1$. In calculations one finds that ϵ_d is often around $e^{c\sqrt{d}}$. If that is true for d then Dirichlet's class number formula implies that $h(d) < c\log d$ for some constant $c>0$. Actually $L\left(1, \left(\frac{d}{\cdot}\right)\right)$ is much more usually close to 1; that is, it is between $\frac{1}{10}$ and 10 for more than 99% of the values of d. Hence if ϵ_d is typically around since $e^{c\sqrt{d}}$ then $h(d)$ is typically bounded.

6.3. The class number one problem in real quadratic fields. Although $h(-d)$ gets large, roughly of size $\sqrt{d}$ as d gets larger, surprisingly $h(d)$ seems to mostly remains quite small. What we do know (as we discuss in the next section) is that $h(d)\log\epsilon_d$ is roughly of size $\sqrt{d}$ as d gets larger, so that computational data suggests that ϵ_d is often around $e^{\sqrt{d}}$ whereas $h(d)$ stays small. There are exceptions; for example if $d=m^2+1$ then $\epsilon_d = m+\sqrt{d}$ and so we can prove, for such d, that $h(d)$ gets large (like $\sqrt{d}$).

Hooley, and Cohen and Lenstra, made some attempts to guess at how often $h(d)$ is small. One can show that there are distinct binary quadratic forms for each odd squarefree divisor of d and so $h(d)\geq 2^{\nu(d)}$, where $\nu(d)$ is the number of odd prime factors of d. Therefore the smallest that $h(d)$ can be is $2^{\nu(d)}$ (the idoneal numbers) and therefore if $h(d)=1$ then d must be prime. Gauss observed that $h(p)=1$ for what seemed to be a positive proportion of primes p and this is still an open problem today. Even proving that there are infinitely many primes p for which $h(p)=1$, is open.

Cohen and Lenstra made the following conjectures

$$\text{The proportion of } d \text{ for which } p \text{ divides } h(d) = 1 - \prod_{k\geq 2}\left(1 - \frac{1}{p^k}\right)$$

The proportion of primes p for which $h(p) = 1$ is $\lambda := \prod_{k\geq 2}\left(1 - \frac{1}{2^k}\right)\zeta(k) = .7544581517\ldots$ Then the proportion of primes p for which $h(p) = q$, where q is prime, equals $\lambda/q(q-1)$.

Finally if most of the class numbers are small, but the occasional one is as big as $\sqrt{p}$ then which dominates in the average? The conjecture is that for the primes $p \leq x$ with $p \equiv 1 \pmod 4$, we have that $h(p) \sim \frac{1}{8}\log x$ on average (and thus, the big class numbers are very rare).

6.4. Dirichlet L-functions. One should define these in general.

7. More about quadratic forms and lattices

7.1. Minkowski and lattices. A *lattice* Λ in $\mathbb{R}^n$ is the set of points generated by n linearly independent vectors, with basis $x_1, x_2, \dots, x_n$ say. In other words

$$\Lambda := \{a_1x_1 + a_2x_2 + \ldots + a_nx_n : \ a_1, a_2, \dots, a_n \in \mathbb{Z}\}.$$

One can see that Λ is an additive group, but it also has some geometry connected to it. The *fundamental parallellopiped* of Λ with respect to $x_1, x_2, \dots, x_n$ is the set $P = \{a_1x_1 + a_2x_2 + \cdots + a_nx_n : 0 \le a_i < 1\}$. The sets $x + P$, $x \in \Lambda$ are disjoint and their union is $\mathbb{R}^n$. The *determinant* $\det(\Lambda)$ of Λ is the volume of P; in fact $\det(\Lambda) = |\det(A)|$, where A is the matrix with column vectors $x_1, x_2, \dots, x_n$ (written as vectors in $\mathbb{R}^n$). A *convex body* K is a bounded convex open subset of $\mathbb{R}^n$.

We define $A - B$ to be the set of points that can be expressed as $a - b$. A key result is:

Blichfeldt's Lemma. *Let $K \subset \mathbb{R}^n$ be a measurable set, and Λ a lattice such that* $\mathrm{vol}(K) > \det(\Lambda)$. *Then $K - K$ contains a non-zero point of Λ.*

Proof. (By the pigeonhole principle.) Let L be the set of points $\ell \in P$ such that there exists $x \in \Lambda$ for which $\ell + x \in K$. We claim that there are two such x for at least one point in L, else $\mathrm{vol}\ K = \mathrm{vol}\ L \le \mathrm{vol}\ P = \det(\Lambda) < \mathrm{vol}(K)$, by hypothesis, a contradiction. Therefore for $k_x := \ell + x \ne k_y := \ell + y \in K$ with $x, y \in \Lambda$ we have $k_x - k_y = x - y \in \Lambda$ which is the result claimed.

`Exercise 7.1.1` Show that if $\mathrm{vol}(K) > m\ \det(\Lambda)$. Then $K - K$ contains at least m non-zero points of Λ.

We deduce:

Minkowski's First Theorem. *If K is a centrally symmetric convex body with* $\mathrm{vol}(K) > 2^n\det(\Lambda)$ *then K contains a non-zero point of Λ.*

Proof. As K is convex and centrally symmetric, $K = \frac{1}{2}\ K - \frac{1}{2}\ K$. However, $\mathrm{vol}(\frac{1}{2}\ K) > \det(\Lambda)$, so the result follows by Blichfeldt's Lemma.

Another proof of the sum of two squares theorem. Suppose that p is a prime $\equiv 1 \pmod 4$ so that there exist integers a, b such that $a^2 + b^2 \equiv 0 \pmod p$. Let Λ be the lattice in $\mathbb{Z}^2$ generated by $(a, b), (-b, a)$.

`Exercise 7.1.2.` Prove that $\det(\Lambda) = p$. Show that if $(u, v) \in \Lambda$ then $u^2 + v^2 \equiv 0 \pmod p$.

Let $K := \{(x, y : \ x^2 + y^2 < 2p\}$ so that $\mathrm{vol}(K) = 2\pi p > 2^2\det(\Lambda)$. Minkowski's First Theorem implies that there exists a non-zero $(u, v) \in K \cap \Lambda$, so that $0 < u^2 + v^2 < 2p$ and $u^2 + v^2 \equiv 0 \pmod p$, which implies that $u^2 + v^2 = p$.

Another proof of the local-glocal principle for diagonal quadratic forms. Let a, b, c be given integers such that abc is coprime and all the residue symbols work out. Let Λ be the lattice in $\mathbb{Z}^3$ generated by solutions x, y, z to $ax^2 + by^2 + cz^2 \equiv 0 \pmod{4abc}$. We will prove that $\det(\Lambda) = 4|abc|$:

The first observation is that if, say, $p|a$ then we know, by hypothesis that there exist u, v with $bu^2 + cv^2 \equiv 0 \pmod p$, etc (To be understood).

Now let $K := \{(x, y, z : \ |a|x^2 + |b|y^2 + |c|z^2 < 4|abc|\}$ so that $\text{vol}(K) = \frac{8\pi}{3} \cdot 4|abc| > 2^3\det(\Lambda)$. Hence Minkowski's First Theorem implies that there exists a non-zero $(u, v, w) \in K \cap \Lambda$, such that $au^2 + bv^2 + cw^2 \equiv 0 \pmod{4abc}$ with $|au^2 + bv^2 + cw^2| \le |a|u^2 + |b|v^2 + |c|u^2 < 4|abc|$.

Hence we have shown that there exists a non-zero integer solution to

$$ax^2 + by^2 + cz^2 = 0, \text{with } \ |a|x^2 + |b|y^2 + |c|z^2 < 4|abc|.$$

`Exercise 7.1.3.` Can you improve the 4 in the last displayed equation?

`Exercise 7.1.4.` We may assume, wlog, that $a, b, c > 0$ and we are looking for solutions to $ax^2+by^2 = cz^2$. Now try $\Lambda := \{(x, y, z) : \ ax^2 + by^2 + cz^2 \equiv 0 \pmod{2abc}\}$ with $K := \{(x, y, z : \ ax^2 + by^2, \ cz^2 < 2|abc|\}$. What do you get?

For a centrally symmetric convex body K define λ_k to be the infimum of those λ for which λK contains k linearly independent vectors of Λ. We call $\lambda_1, \lambda_2, \dots, \lambda_n$ the *successive minima* of K with respect to Λ. Let $b_1, b_2, \dots, b_n \in \mathbb{R}^n$ be linearly independent vectors with $b_k \in \lambda_k \overline{K} \cap \Lambda$ for each k. The proof of the next result, and much more, can be found in [15].

Minkowski's Second Theorem. *If $0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ are the successive minima of convex body K with respect to Λ then $\lambda_1\lambda_2 \ldots \lambda_n \text{vol}(K) \le 2^n \det(\Lambda)$.*

Let $r_1, r_2, \dots, r_k \in \mathbb{Z}/N\mathbb{Z}$ and $\delta > 0$ be given. We define the *Bohr neighbourhood*

$$B(r_1, r_2, \dots, r_k; \delta) := \{s \in \mathbb{Z}/N\mathbb{Z} : \|r_i s/N\| \le \delta \text{ for } i = 1, 2, \dots, k\};$$

that is, the least residue, in absolute value, of each $r_i s \pmod N$ is $< \delta N$ in absolute value.

7.2. The number of representations as the sum of two squares. Let $r(n)$ be the number of ways in which n can be written as the sum of two squares.

In exercise 3.5.4 we proved that there is a unique way (up to changes of sign and order) to write $p \equiv 1 \pmod 4$ as the sum of two squares, say $p = a^2 + b^2$. Then we have $p = (\pm a)^2 + (\pm b)^2 = (\pm b)^2 + (\pm a)^2$, that is $r(p) = 8$. We also have the unique factorization $p = (a + ib)(a - ib)$ so just two prime factors, and there are four units $1, -1, i, -i$. Let $R(n) = r(n)/4$, so that $R(p) = 2$, corresponding to the two possibilities $a + ib$ and $a - ib$. Now there are the three factors $(a + ib)^2, (a + ib)(a - ib), (a - ib)^2$ of p^2 so that $R(p^2) = 3$, and in general p^k has the factors $(a + ib)^j(a - ib)^{k-j}$ for $0 \le j \le k$, so that $R(p^k) = k + 1$.

Now $2 = i(1 - i)^2$ so that $R(2^k) = 1$. Finally, if $p \equiv 3 \pmod 4$ then $R(p^{\text{odd}}) = 0$ and $R(p^{\text{even}}) = 1$.

Hence $r(n) = 4R(n)$ is a multiplicative function. We saw that $r(n) \ne 0$ if and only if we can write $n = 2^k m_+ m_-^2$ where if $p|m_\pm$ then $p \equiv \pm 1 \pmod 4$. In that case $R(n) = \tau(m_+)$.

When we write $p = a^2 + b^2$ it would be nice to have an easy way to determine a and b. We will prove later that if $s_m := \sum_{1 \le n \le p} \left(\frac{n^3 - mn}{p}\right)$ then $a = s_{-1}/2$ and $b = s_r/2$ where $(r/p) = -1$.

We can also look at

$$N(x) = \#\{(a,b) \in \mathbb{Z} : \ a^2 + b^2 \le x\} = \sum_{n \le x} R(n).$$

Notice that this should be well approximated by the area of the circle πx with an error proportional to the circumference, that is bounded by a multiple of $\sqrt{x}$. Improving this error term is the *Gauss circle problem*. It is believed to of size around $x^{1/4}$.

7.3. The number of representations by arbitrary binary quadratic forms. Given a fundamental discriminant d and an integer n we are interested in how many inequivalent primitive representations of n there are by binary quadratic forms of discriminant d. Let $f(x,y)$ be a reduced form and suppose that $n = f(\alpha,\gamma)$ where $(\alpha,\gamma) = 1$. We choose integers β, δ so that $\alpha\delta - \beta\gamma = 1$ and transform f to an equivalent binary quadratic form with leading coefficient n so that our representation becomes $n = f(1,0)$. Next we transform $x \to x + ky, \ y \to y$ so that f is equivalent to a binary quadratic form $nx^2 + Bxy + Cy^2$ with $B^2 \equiv d \pmod{4n}$ and $-n < B \le n$ (and $C = (B^2 - d)/4n$).

`Exercise 7.3.1.` Prove that if $f(\alpha x + \beta y, \gamma x + \delta y) = nx^2 + Bxy + Cy^2$ then, no matter what integers β, δ we take satisfying $\alpha\delta - \beta\gamma = 1$, we get the same value of $B \pmod{2n}$.

We need to determine whether any two representations of n by f lead to the same $nx^2 + Bxy + Cy^2$. If so we have $Tf = nx^2 + Bxy + Cy^2 = Uf$ for two different (and invertible) transformations T, U and so $T^{-1}Uf = f$. So we must determine all *automorphisms* of reduced $f = ax^2 + bxy + cy^2$.

There are infinitely many automorphisms if $d > 0$ via Pell's equation, for if $k = am^2 + bmn + cn^2$ then $k = aM^2 + bMN + cN^2$ where $M = um - 2cvn - bvm, \ N = 2avm + bvn + un$ for all u, v satisfying $u^2 - dv^2 = 1$. This essentially accounts for all automorphisms, but we leave this for now to focus on the easier, case where $d < 0$.

Now, given any such automorphism, we must have $a = f(\alpha,\gamma)$ and $c = f(\beta,\delta)$.

`Exercise 7.3.2.` In this exercise we find all the automorphisms of binary quadratic forms when $d < 0$.

i. Prove that the automorphisms form a group, containing $\pm I$ (that is $(x,y) \to \pm(x,y)$).
ii. Use exercise 4.3.2(i) to show that if $0 < |b| < a < c$ then the only automorphisms are $\pm I$.
iii. Show that if $b = 0$ and $a < c$ we also have $(x,y) \to \pm(x,-y)$.
iv. Show that if $|b| < a = c$ we also have $(x,y) \to \pm(y,x)$.
v. Determine the complete set of automorphisms in all cases.

Let $R_d(n)$ be the total number of representations of n by all binary quadratic forms of discriminant d (though counting only one solution for each equivalence class via the automorphisms), and let $r_d(m)$ the number of such representations $m = f(\alpha,\gamma)$ where $(\alpha,\gamma) = 1$. Evidently these are both multiplicative functions, by the Chinese Remainder Theorem, and $R_d(n) = \sum_{k: \ k^2 | n} r_d(n/k^2)$. Now, by the above argument, we see that to determine $r_d(p^e)$ we need to count solutions to $B^2 \equiv d \pmod{p^e}$ if p is odd, and $\pmod{2^{e+2}}$ if $p = 2$. Hence $r_d(p^e) = 1 + \left(\frac{d}{p}\right)$ if p is odd, and $= 0$ if $p = 2$, unless $d \equiv 1 \pmod 8$ in which case it equals 4. Hence if n is odd then

$$r_d(n) = \prod_{p^e \| n} r_d(p^e) = \prod_{p^e \| n} \left\{1 + \left(\frac{d}{p}\right)\right\},$$

from which we deduce that

$$(7.3.1) \qquad R_d(n) = \prod_{p^e \| n} \left\{ 1 + \left(\frac{d}{p}\right) + \left(\frac{d}{p^2}\right) + \ldots + \left(\frac{d}{p^e}\right) \right\} = \sum_{m|n} \left(\frac{d}{m}\right) = \sum_{\ell m = n} \left(\frac{d}{m}\right).$$

Hence if $d < 0$ then the total number of representations is $w(d)R_d(n)$ where, as we have seen in exercise 7.3.2, $w(d) = 6$ if $d = -3$, $w(d) = 4$ if $d = -4$ and $w(d) = 2$ for all fundamental discriminants $d < -4$. Evidently there are infinitely many representations when $d > 0$.

It is instructive to derive the formula for $R_d(n)$ via ideals. If we think through the equivalence between representations by binary quadratic forms and ideals then we find that $R_d(n)$ can be interpreted as the number of distinct factorizations of the ideal (n) into two conjugate ideals of the ring of integers of $\mathbb{Q}(\sqrt{d})$ (where d is a fundamental discriminant). Breaking n into prime powers (p^e), we must first understand how (p) splits. If (p) is a prime ideal then $(p^e) = (p)^e$ cannot split into two conjugate ideals unless e is even, and then it can only factor as $(p)^{e/2} \cdot (p)^{e/2}$. Hence $R_d(p^e) = 1$ if e is even, and 0 if e is odd and, in this case, $(d/p) = -1$, so we verify the formula given earlier. If $(p) = P^2$ where P is a prime ideal (in fact this happens if and only if $p|d$, so that $(d/p) = 0$), then the only way to factor $(p)^e$ is as $P^e \cdot P^e$, and so $R_d(p^e) = 1$, again verifying the above formula. Finally if $(d/p) = 1$ then we can write $(p) = P\overline{P}$ for some prime ideal P, and then we have the factorizations $(p)^e = A\overline{A}$ where $A = P^j\overline{P}^{e-j}$, for $j = 0, 1, 2, \ldots, e$. Hence $R_d(p^e) = e + 1$, as given above.

Let $f(x, y)$ be a binary quadratic form and let $R_f(N)$ be the number of representations of N by f; that is, the number of pairs of integers m, n for which $f(m, n) = N$. If $f_1, f_2, \ldots, f_h$ are the reduced form of discriminant d then

$$(7.3.2) \qquad R_{f_1}(n) + R_{f_2}(n) + \ldots + R_{f_h}(n) = w(d)R_d(n), \text{ for all } n,$$

by definition. We sum this up over $n = 1, 2, \ldots N$. Now

$$\sum_{n=1}^{N} R_f(n) = \#\{(u, v) \in \mathbb{Z} : \ f(u, v) \le N\} = \frac{2\pi}{\sqrt{|d|}}\, N + O_f(\sqrt{N}).$$

`Exercise` 7.3.3. Prove this last estimate by approximating the number of lattice points with $f(u, v) \le N$ by the volume of this region.

On the other hand

$$\begin{aligned} \sum_{n=1}^{N} R_d(n) &= \sum_{n \le N} \sum_{\ell m = n} \left(\frac{d}{m}\right) = \sum_{\ell m \le N} \left(\frac{d}{m}\right) \\ &= \sum_{m \le \sqrt{N}} \left(\frac{d}{m}\right) \sum_{\ell \le N/m} 1 + \sum_{\ell \le \sqrt{N}} \sum_{\sqrt{N} < m \le N/\ell} \left(\frac{d}{m}\right). \end{aligned}$$

For the first term the inner sum is N/m with an error of at most 1, so in total we obtain

$$N \sum_{m \le \sqrt{N}} \left(\frac{d}{m}\right) \cdot \frac{1}{m}$$

plus an error of $O(\sqrt{N})$. For the second term we note that if m runs through $4d$ consecutive integers then the sum of (d/m) equals 0. Hence the sum of (d/m) on any intervals is at most $4d$. Therefore the second term is at most $4d\sqrt{N}$.

So comparing the two side of our sum we obtain

$$Nw(d) \sum_{m \le \sqrt{N}} \left(\frac{d}{m}\right) \cdot \frac{1}{m} = \; h(d) \cdot \frac{2\pi}{\sqrt{|d|}} \, N + O_d(\sqrt{N})$$

where the constant in the error term depends only on d. Dividing through by n and taking the limit as $N \to \infty$ we obtain

$$w(d) \sum_{m \ge 1} \left(\frac{d}{m}\right) \cdot \frac{1}{m} = \; h(d) \cdot \frac{2\pi}{\sqrt{|d|}} \, .$$

Notice that the right side here is $L\left(1, \left(\frac{d}{\cdot}\right)\right)$, and so we have proved Dirichlet's class number formula, given in section 6.2.

7.4. The number of representations as the sum of three squares. We have seen which integers are representable as the sum of two squares. How about three?

• The only squares mod 4 are 0 and 1. Therefore if n is divisible by 4 and is the sum of three squares then all three squares must be even. Hence if $n = 4m$ then to obtain every representation of n as the sum of three squares, we just take every representation of m as the sum of three squares, and double the number that are being squared.

• The only squares mod 8 are $0, 1$ and 4. Therefore no integer $\equiv 7 \pmod 8$ can be written as the sum of three squares (of integers). By the previous remark no integer of the form $4^k(8m + 7)$ can be written as the sum of three squares.

Legendre's Theorem. (1798) *A positive integer n can be written as the sum of three squares of integers if and only if it is not of the form $4^k(8m + 7)$.*

We will not prove this as all known proofs are too complicated for a first course.

One might ask how many ways are there to write an integer as the sum of three squares? Gauss proved the following remarkable theorem (for which there is still no easy proof): Suppose that n is squarefree.[12] If $n \equiv 3 \pmod 8$ then there are $8h(-4n)$ ways in which n can be written as the sum of three squares; if $n \equiv 1$ or $2 \pmod 4$, $n > 1$ there are $12h(-4n)$ ways. In fact Gauss worked with the genera of the quadratic forms of discriminant $d = -4n$, and showed that the number of elements in each genus is given by the number of representations of n as the sum of three squares, divided by 8 or 12 times the number of genera (which he determined much as we did in exercise 4.7.2).

[12] That is $p^2 \not| n$ for all primes p.

This also leads to an even easier way to determine the idoneal numbers (or at least those that are not of the form $4^k(8m+7)$), since these are exactly those discriminants for which the number of genera equals the class number. Hence we must have that d not of the form $4^k(8m+7)$ is an idoneal number if and only if the number of representations of n (suitably defined) as the sum of three squares equals 8 or 12 times the number of genera.

7.5. The number of representations as the sum of four squares.

Lagrange's Theorem. *Every positive integer is the sum of four squares*

Proof. We start from the identity

$$(a^2+b^2+c^2+d^2)(u^2+v^2+w^2+x^2) = (au+bv+cw+dx)^2 + (av-bu-cx+dw)^2$$
$$+ (aw+bx-cu-dv)^2 + (ax-bw+cv-du)^2, \tag{7.5.1}$$

which is much like what we saw for the sum of two squares. Hence it suffices to show that every prime is the sum of four squares, and we can show any product of primes is the sum of four squares using the above identity. Now $2 = 1^2+1^2+0^2+0^2$ so we focus on odd primes p: We know that there exist non-zero integers a,b,c,d such that $a^2+b^2+c^2+d^2 \equiv 0 \pmod p$; select them so that m is minimal, where $mp = a^2+b^2+c^2+d^2$. Our goal is to show that $m=1$.

`Exercise 7.5.1.` Prove that $|a|,|b|,|c|,|d| < p/2$, so that $m<p$.

`Exercise 7.5.2.` Show that m is odd: Show that if m is even then we can reorder a,b,c,d so that $a-b$ and $c-d$ are both even. But then $\frac{m}{2}\cdot p = (\frac{a-b}{2})^2+(\frac{a+b}{2})^2+(\frac{c-d}{2})^2+(\frac{c+d}{2})^2$ contradicting the minimality of m.

Let u,v,w,x be the least residues, in absolute value, of $a,b,c,d \pmod m$, respectively. Therefore $u^2+v^2+w^2+x^2 \equiv a^2+b^2+c^2+d^2 \equiv 0 \pmod m$. Moreover $|u|,|v|,|w|,|x| < m/2$ (since m is odd), and so $u^2+v^2+w^2+x^2 < 4(m/2)^2 = m^2$. Hence we can write $u^2+v^2+w^2+x^2 = mn$ for some integer $n<m$.

`Exercise 7.5.3.` Let $A := au+bv+cw+dx$; $B := -av+bu-cx+dw$; $C := -aw+bx+cu-dv$; $D := -ax-bw+cv+du$. Prove that $A \equiv B \equiv C \equiv D \equiv 0 \pmod m$.

Now $A/m, B/m, C/m, D/m$ are integers by the last exercise, and so

$$(A/m)^2+(B/m)^2+(C/m)^2+(D/m)^2 = \frac{(a^2+b^2+c^2+d^2)}{m}\cdot\frac{(u^2+v^2+w^2+x^2)}{m} = np.$$

This contradicts the minimality of m unless $n=0$ in which case $u=v=w=x=0$ so that $a\equiv b\equiv c\equiv d\equiv 0 \pmod m$ and so $m^2 | a^2+b^2+c^2+d^2 = mp$. Therefore $m=1$ as $m<p$, which is what we wished to prove.

Quaternions. The identity $(a^2+b^2)(u^2+v^2) = (x^2+y^2)$ where $x=au-bv$ and $y=av+bu$, which was used in the result about representations by the sum of two squares, is perhaps most naturally obtained from taking norms in the product $(a+ib)(u+iv) = x+iy$ where $i^2=-1$. Above we have an identity for sums of four squares which comes from

multiplying$(a^2+b^2+c^2+d^2)(u^2+v^2+w^2+x^2)$. To obtain this from taking norms in an algebraic product we need to work in the *quaternions* which is a non-commutative ring; in this case there are three square roots of -1 which do not commute with one another: We let i, j, k be "imaginary numbers" such that

$$i^2 = j^2 = k^2 = -1$$
$$ij = k = -ji, \quad jk = i = -kj, \quad ki = j = -ik.$$

The *quaternions* are the ring $\{a + bi + cj + dk : \ a, b, c, d \in \mathbb{Z}\}$. Now

$$\text{Norm}(a + bi + cj + dk) := (a + bi + cj + dk)(a - bi - cj - dk) = a^2 + b^2 + c^2 + d^2$$

since $ij + ji = jk + kj = ki + ik = 0$. Moreover

$$(a + bi + cj + dk)(u - vi - wj - xk) = A + Bi + Cj + Dk$$

with A, B, C, D defined as in exercise 7.5.3. The above proof of representations can be translated into the language of quaternions.

The number of representations. In 1834 Jacobi showed that there are $8\sigma(n)$ representations of n as a sum of four squares if n is odd, and $24\sigma(m)$ representations if $n = 2^k m, \ k \geq 1$ is even.

`Exercise 7.5.4.` Prove that this can re-written as follows:

$$\left(\sum_{n\in\mathbb{Z}} x^{n^2}\right)^4 = 8 \sum_{\substack{d\geq 1\\ 4\nmid d}} \frac{dx^d}{1-x^d} \ .$$

7.6. Universality of quadratic forms. Once one knows that every positive integer can be represented by the sum of four squares, but not as the sum of three squares, one might ask for further positive definite quadratic forms with this property.

It turns out that no quadratic or ternary quadratic form can represent all integers.

In 1916 Ramanujan asserted that the quaternary quadratic forms with the following coefficients represent all integers: $\{1, 1, 1, k\}$, $\{1, 2, 2, k\}, : 1 \leq k \leq 7$; $\{1, 1, 2, k\}$, $\{1, 2, 4, k\}$: $1 \leq k \leq 14$; $\{1, 1, 3, k\} : 1 \leq k \leq 6$; $\{1, 2, 3, k\}, \{1, 2, 5, k\} : 1 \leq k \leq 10$; though this is not quite true for $\{1, 2, 5, 5\}$ since it represent every positive integer except 15. We deduce

The Fifteen criterion, I. *Suppose that f is a positive definite diagonal quadratic form. Then f represents all positive integers if and only if f represents all positive integers ≤ 15.*

Proof. Suppose that $f = a_1x_1^2 + a_2x_2^2 + \ldots + a_dx_d^2$, with $1 \leq a_1 \leq a_2 \leq \ldots \leq a_d$ represents all positive integers. Since f represents 1 we must have $a_1 = 1$. Since f represents 2 we must have $a_2 = 1$ or 2. If $a_1 = a_2 = 1$ then, since f represents 3 we must have $a_3 = 1, 2$

or 3. If $a_1 = 1, a_2 = 2$ then, since f represents 5 we must have $a_3 = 2, 3, 4$ or 5. Now

$$\begin{aligned}
x_1^2 + x_2^2 + x_3^2 &\text{ represents } m,\ 1 \le m \le 6, \text{but not } 7,\ \text{and so } 1 \le a_4 \le 7;\\
x_1^2 + x_2^2 + 2x_3^2 &\text{ represents } m,\ 1 \le m \le 13, \text{but not } 14,\ \text{and so } 1 \le a_4 \le 14;\\
x_1^2 + x_2^2 + 3x_3^2 &\text{ represents } m,\ 1 \le m \le 5, \text{but not } 6,\ \text{and so } 1 \le a_4 \le 6;\\
x_1^2 + 2x_2^2 + 2x_3^2 &\text{ represents } m,\ 1 \le m \le 6, \text{but not } 7,\ \text{and so } 1 \le a_4 \le 7;\\
x_1^2 + 2x_2^2 + 3x_3^2 &\text{ represents } m,\ 1 \le m \le 9, \text{but not } 10,\ \text{and so } 1 \le a_4 \le 10;\\
x_1^2 + 2x_2^2 + 4x_3^2 &\text{ represents } m,\ 1 \le m \le 13, \text{but not } 14,\ \text{and so } 1 \le a_4 \le 14;\\
x_1^2 + 2x_2^2 + 5x_3^2 &\text{ represents } m,\ 1 \le m \le 9, \text{but not } 10,\ \text{and so } 1 \le a_4 \le 10.
\end{aligned}$$

Ramanujan's result implies that $a_1x_1^2 + a_2x_2^2 + a_3x_3^2 + a_4x_4^2$ then represents every positive integer except perhaps 15, and the result follows.

We could look to represent only the nine integers $1, 2, 3, 5, 6, 7, 10, 14, 15$ rather than all integers ≤ 15.

By the 1940s researchers had come up with complicated criteria to decide whether a quadratic form represented all integers, but it took the genius of John Conway to come up with the following simply checked criterion:

The Fifteen criterion, II. *Suppose that f is a positive definite quadratic form, which is diagonal mod 2. Then f represents all positive integers if and only if f represents all positive integers ≤ 15.*

Notice that this is sharp since $x^2 + 2y^2 + 5z^2 + 5w^2$ represents every positive integer other than 15.

This was extended to all quadratic forms by Bhargava and Hanke:

The 290 criterion. *Suppose that f is a positive definite quadratic form. Then f represents all positive integers if and only if f represents all positive integers ≤ 290.*

Notice that this is sharp since $x^2+xy+2y^2+xz+4z^2+29(a^2+ab+b^2)$ represents every positive integer other than 290. In fact one need only verify that f represents the 29 integers $1, 2, 3, 5, 6, 7, 10, 13, 14, 15, 17, 19, 21, 22, 23, 26, 29, 30, 31, 34, 35, 37, 42, 58, 93, 110, 145, 203$ and 290.

The Fifteen criterion has been generalized to representations of any set S of positive integers by Bhargava: There exists a finite subset T of S such that a positive definite quadratic form f, which is diagonal mod 2, represents every integer in S if and only if it represents every integer in T. Hence Bhargava has reduced any such classification problem to a finite problem. For example, such an f represents all primes if and only if f represents the 15 primes up to 47 as well as 67 and 73

7.7. Representation by positive definite quadratic forms. Let us suppose that f is any positive definite quadratic form in three or more variables. One might ask which integers can be represented by f. In all of the examples we have seen f represents all integers, or (like $x^2+y^2+z^2$) all integers in certain residue classes. Is this true in general? We saw even in the quadratic form case that if an integer satisfies certain obvious congruence

conditions that it is represented by some form of the given discriminant, and this result easily generalizes; however we are interesting in representation by a specific form.

In 1929 Tratowsky showed that for any positive definite quadratic form f of discriminant $D > 0$ in five or more variables, if n is sufficiently large then n is represented by f if and only if n is represented by $f \pmod D$.

For three or four variables it usually makes sense (and is easier) to restrict our attention to the representation of squarefree integers n. In 1926 Kloosterman introduced an analytic method which implies that if f of discriminant $D > 0$ has four variables, and if n is a sufficiently large squarefree integer then n is represented by f if and only if n is represented by $f \pmod D$. It was only in 1990, that Duke and Schulze-Pillot extended this to positive definite quadratic forms f of discriminant $D > 0$ in three variables: If n is a sufficiently large squarefree integer then n is represented by f if and only if n is represented by $f \pmod{D^2}$; or more explicitly:

Theorem 7.7. *There exists an absolute constant $c > 0$ such that if n is a squarefree integer with $n > cD^{337}$ then n can be represented by f if and only if there exist integers a, b, c such that $f(a, b, c) \equiv n \pmod{8D^3}$.*

7.8. Descent and the quadratics.. A famous problem asks to prove that if a and b are positive integers for which $ab + 1$ divided by $a^2 + b^2$ is an integer then the quotient is a square. One can approach this as follows: Suppose that $a \geq b \geq 1$ and $a^2 + b^2 = k(ab + 1)$ for some positive integer k. Then a is the root of the quadratic $x^2 - kbx + (b^2 - k)$. If c is the other root then $a + c = kb$ so that c is also an integer for which $b^2 + c^2 = k(bc + 1)$. We shall now prove that this is a "descent": If $c = 0$ then $k = b^2$ and $a = b^3$ and we have a solution. Otherwise $c \geq 1$ else $bc + 1 \leq 0$ and so $b^2 + c^2 \leq 0$ which is impossible. But then $b^2 - k = ac > 0$ and so $c = (b^2 - k)/a < b^2/b = b$. Hence (b, c) gives a smaller pair of solutions than (a, b). We deduce that all solutions can be obtained by iterating the map

$$(b, c) \to (kb - c, b)$$

starting from initial solutions $(d, 0)$ with $k = d^2$.

Other quadratics have a similar property. Perhaps the most famous is the Markov equation: Find positive integers x, y, z for which

$$x^2 + y^2 + z^2 = 3xyz.$$

One finds many solutions: $(1, 1, 1), (1, 1, 2), (1, 2, 5), (1, 5, 13), (2, 5, 29), (1, 13, 34), (1, 34, 89), (2, 29, 169), (5, 13, 194), (1, 89, 233), (5, 29, 433), (89, 233, 610)$. Given one solution (x, y, z) one has that x is a root of a quadratic, the other root being $3yz - x$, and so we obtain a new solution $(3yz - x, y, z)$. (And one can do the same procedure with y or z. If we fix one co-ordinate we see that if there is one solution there are infinitely many. For example, taking $z = 1$ yields the equation $x^2 + y^2 = 3xy - 1$.

`Exercise 7.8.1.` Determine what solutions are obtained from (1,1,1) by using the maps $(x, y) \to (3y - x, y)$ and $(x, y) \to (x, 3x - y)$.

One open question is to determine all of the integers that appear in a Markov triple. The first few are $1, 2, 5, 13, 29, 34, 89, 169, 194, 233, 433, 610, 985, 1325, \ldots$; it is believed that they are quite sparse.

Arguably the most beautiful such problem is the Apollonian circle packing problem.[13] Take three circles that touch each other (for example, take three coins and push them together). In between the circles one has a crescent type shape (*a hyperbolic triangle*), and one can inscribe a (unique) circle that touches all three of the original circles. What is the relationship between the radius of the new circle and the radii of the original circles? If we define the *curvature* of the circles to be $1/r$ (where r is the radius) then Descartes observed, in 1643, that the four curvatures satisfy the equation

$$2(a^2+b^2+c^2+d^2) = (a+b+c+d)^2, \text{ that is } a^2+b^2+c^2+d^2-2(ab+bc+cd+da+ac+bd) = 0.$$

We see that given b, c, d there are two possibilities for a, since this is a quadratic equation, the other is the circle that contains the three original circles and touches them all. We scale up the first three curvatures so that they are integers (with $\gcd(b, c, d) = 1$). We will focus on the case that a is also an integer, for example if we start with $b = c = 2$ and $d = 3$ we have $a^2 - 14a - 15 = 0$ so that $a = -1$ or $a = 15$. Evidently $a = -1$ corresponds to the outer circle (the negative sign comes from the fact that the circle contains the original circles), and $a = 15$ the inner one. In general if we have a solution (a, b, c, d) then we also have a solution (A, b, c, d) with $A = 2(b + c + d) - a$. Yet again we can iterate this (perhaps using the variables b, c or d) and obtain infinitely many Apollonian circles. But there is another interpretation of this, since each time we have a crescent in-between three existing circles we fill part of it in with a new circle, and we are eventually *tiling* the while of the original circle (see the enclosed pictures). There are many questions that can be asked: What integers appear as curvatures in a given packing? There are some integers that cannot appear because of congruence restrictions. For example if a, b, c, d are all odd, then all integers that arise as curvatures in this packing will be odd. The conjecture is that all sufficiently large integers that satisfy these trivial congruence constraints (mod 24) will appear as curvatures in the given packing. Although this is an open question, we do know that a positive proportion of integers appear in any such packing, that the total number of circles in packing with curvature $\leq x$ is $\sim cT^\alpha$ where $\alpha = 1.30568\ldots$, and that the Apollonian twin prime conjecture holds: that there are infinitely many pairs of touching circles in the packing whose curvatures are both primes.

This last question is accessible because we see that any given solution (a, b, c, d) is mapped to another solution by any permutation of the four elements, as well as the matrix $\begin{pmatrix} -1 & 2 & 2 & 2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. These (linear) transformations generate a subgroup of $\mathrm{SL}(4, \mathbb{Z})$, and one can proceed by considering orbits under the actions of this subgroup.

[13] Apollonius lived in Perga, 262-190 BC.

8. More useful mathematics

8.1. Finite fields. Congruences in the rings of integer of an extension of $\mathbb{Q}$ are a little subtle. For example, what are the set of residue classes $m+ni \pmod 3$ with $m,n \in \mathbb{Z}$? It is evident that any such number is congruent to some $a+bi \pmod 3$ with $a,b \in \{0,1,2\}$; are any of these nine residue classes congruent? If two are congruent then take their difference to obtain $u \equiv iv \pmod 3$ for some integers u,v, not both 0, and each ≤ 2 in absolute value. But then $u-iv = 3(r-is)$ for some integers r,s and hence $3|u,v$ and hence $u=v=0$, a contradiction

`Exercise 8.1.1.` Generalize this argument to show that there are exactly p^2 distinct residue classes amongst the integers $a+bi \pmod p$ for any prime p.

When we work modulo a prime in $\mathbb{Z}$ we have all of the usual rules of addition, multiplication and even division. But if we work modulo a composite number then division becomes more perilous (for example $6 \equiv 18 \pmod{12}$ but $2 \not\equiv 6 \pmod{12}$). The issue is that the group $(\mathbb{Z}/n\mathbb{Z})^*$ has zero divisors; that is $ab \equiv 0 \pmod n$ with neither a nor b are $\equiv 0 \pmod n$.

What about mod p when working in $\mathbb{Z}[i]$? If (rational) prime $p \equiv 1 \pmod 4$ we can write $p = a^2+b^2$ and so $(a+bi)(a-bi) = p \equiv 0 \pmod p$ yet neither $a+bi$ nor $a-bi$ are $\equiv 0 \pmod p$. That is the primes $\equiv 1 \pmod 4$ *are composite numbers in* $\mathbb{Z}[i]$. For example $5 = (2+i)(2-i)$ so that $a+bi \equiv a-2b \pmod{2+i}$.

`Exercise 8.1.2.` Generalize this argument to show that if rational prime $p = a^2+b^2$ then there are exactly p distinct residue classes amongst the integers $u+vi \pmod{a+bi}$.

On the other hand the primes $p \equiv 3 \pmod 4$ are primes in $\mathbb{Z}[i]$, so that if $(a+bi)(c+di) \equiv 0 \pmod p$ then p divides $(a+bi)(c+di)(a-bi)(c-di) = (a^2+b^2)(c^2+d^2)$. Hence p divides one of a^2+b^2 and c^2+d^2, say the first, so that p divides both a and b. But then $a+bi \equiv 0 \pmod p$, and so we have proved that there are no zero divisors mod p.

A set of numbers in which all the usual rules of addition, subtraction, multiplication and division hold is called a *field*. Its definition is that it is a set F, where F has an additive group and $F \setminus \{0\}$ has a multiplicative group, both commutative, their identity elements, denoted 0 and 1 respectively, are distinct, and that $a \cdot (b+c) = a \cdot b + a \cdot c$. Since $F \setminus \{0\}$ is a multiplicative group hence F has no zero divisors.

We have seen many example of infinite fields $(\mathbb{Q}, \mathbb{Q}(\sqrt{d}), \mathbb{R}, \dots)$, so we now suppose that F is a finite field.

`Exercise 8.1.3.` By Lagrange's Theorem we know that $|F| \cdot 1 = 0$.

i. Show that if prime q divides $|F|$ then either $q \cdot 1 = 0$ or $|F|/q \cdot 1 = 0$. Use an induction hypothesis to show that there exists a prime p such that $p \cdot 1 = 0$ in F.

ii. Now show that this prime p is unique.

iii. Begin with a non-zero element $a_1 \in F$. Let $P = \{1, 2, \dots, p\}$. If $a_2 \notin I_P(a_1)$ then show that $I_P(a_1, a_2)$ has p^2 distinct elements. Hence by induction show that there are p^r distinct elements of F given by $I_P(a_1, a_2, \dots, a_r)$.

Hence we deduce that finite fields must have $q = p^r$ elements for some prime p and integer $r \geq 1$. It can be shown that, up to isomorphism, there is just one such field for each prime power. We denote this field as $\mathbb{F}_q$.

The easiest way to construct a finite field of p^r elements is to use a root α of a polynomial $f(x)$ of degree r which is irreducible in $\mathbb{F}_p$. (Roughly 1 in r polynomials of degree r are irreducible). Then we can represent the elements of the finite field as $a_0 + a_1\alpha + \ldots + a_{r-1}\alpha^{r-1}$ where we take the $a_i \in \mathbb{F}_p$.

`Exercise 8.1.4.` Verify that this indeed gives the field on p^r elements.

`Exercise 8.1.5.` Show that the multiplicative group is indeed cyclic. We call any generator of this group, a *primitive root*.

The multiplicative group of $\mathbb{F}_{p^r}$ has $p^r - 1$ elements so that $a^{p^r-1} = 1$ for all $a \in F^*$ by Lagrange's Theorem. Therefore $a^{p^r} = a$ for all $a \in F$. Hence the map $x \to x^p$ partitions the field into orbits of size $\leq r$ of the form $a, a^p, a^{p^2}, a^{p^3}, \ldots, a^{p^{r-1}}$. In particular, since $p = 0$ we can use the multinomial theorem to establish that $f(x^p) = f(x)^p$. Hence if a is a root of an irreducible polynomial over $\mathbb{F}_p$ of degree r then $a, a^p, a^{p^2}, a^{p^3}, \ldots, a^{p^{r-1}}$ are the r distinct roots of the polynomial. This implies, for instance, that if g is a primitive root then g is the root of an irreducible polynomial of degree r over $\mathbb{F}_p$.

The integers mod p are not only a field but isomorphic to $\mathbb{F}_p$.

`Exercise 8.1.6.` Show that the finite field on p^2 elements is quite different from the integers mod p^2.

8.2. Affine vs. Projective. When we discussed the pythagorean equation $x^2 + y^2 = z^2$ we saw a correspondence between the integer solutions with $z \neq 0$ and $\gcd(x, y, z) = 1$ and the rational points on $u^2 + v^2 = 1$. Let us look at this a little more closely.

To deal with the uninteresting fact that we get infinitely many solutions by scaling a given solution of $x^2 + y^2 = z^2$ through by a constant, we usually impose a condition like $\gcd(x, y, z) = 1$, and stick with integer solutions or, when $z \neq 0$, divide out by z and get a rational solution. The first is arguably unsatisfactory since we select one of an infinite class of solutions somewhat arbitrarily; moreover we haven't really decided between (x, y, z) and $(-x, -y, -z)$, and when we ask the same question say in $\mathbb{Z}[\sqrt{5}]$ then there will infinitely many such equivalent solutions (i.e. we can multiply through by the unit $(2 + \sqrt{5})^k$ for any k). One can overcome these issues by treating solutions as the same if the ratios $x : y : z$ are the same. This equivalence class of solutions is called a *projective* solution to the Diophantine equation. This is only possible if the different monomials in the equation all have the same degree.

This is almost the same thing as dividing out by the z-value. This reduces the number of variables in the equation by 1, and the different monomials do not all have the same degree. The solutions here are *affine* solutions. Often it is more convenient to work with rational solutions to an affine equation, but it does have the disadvantage that we have "lost" the solutions where $z = 0$. One way to deal with this is to ask oneself what was so special about z? Why not divide through by y and get a different affine equation, and recover all the solutions except those with $y = 0$? Or do the same with x. It seems like a bit of overkill for what turns out to be just one or two solutions, but this discussion does make the point that there can be several different affine models for a given projective equation. What is typically done is to work with one affine model, say our first, and treat the lost solutions separately, calling them *the points at infinity* as if we divided through by

$z = 0$. It is good to keep track of them since then all affine models of the same equation, have the same solutions!

Affine equations in two variables are curves, and so projective equations in three variables are also known as curves.

There are, moreover, other ways to transform equations. One can often reduce the number of monomials in a problem by suitable linear transformations of the variables (like completing the square to solve a quadratic equation). In this case the rational solutions are be mapped 1-to-1, though the integer solutions are not necessarily. One special case is the projective equation $x^2 - dy^2 = z^2$. If we divide through by z we get the Pell equation $u^2 - dv^2 = 1$ and we will see (in section 9) that this has $p - (d/p)$ solutions mod p. When $z = 0$ we are asking for solutions to $x^2 = dy^2$, and this evidently has $1 + (d/p)$ solutions mod p. Hence the total number of solutions, counting those at infinity, is $p + 1$. More on points at infinity later.

The linear transformations on binary quadratic forms (that we worked with in previous chapters) kept the number of variables the same, which is different from the above transformations. For example solutions to $z^2 = x^2 + 2xy + 2y^2$ and in 1-1 correspondence with solutions to $z^2 = v^2 + y^2$ taking $v = x + y$. However other linear transformations can be a little confusing. For example if we are looking at solutions to $y^4 = x^4 + x$ we might take the change of variables $y = v/u$ and $x = 1/u$ to obtain the equation $v^4 = u^3 + 1$. At first sight this new model appears to be of lower degree than the original equation; so to understand rational points we probably need more subtle invariants than degree. The *genus* of the curve handles this for us, though it's definition involves more algebraic geometry than we want to discuss in this book. For now it suffices to note that linear and quadratic equations have genus zero and, if solvable, will have parameterized families of solutions, like the Pythagorean equation. Equations of degree 3 like $y^2 = x^3 + ax + b$ and $x^3 + y^3 = 1$ have genus one, and if solvable non-trivially, typically have infinitely many solutions which can be determined from the first few – we will discuss this in some detail in chapter 10 (and beyond). Higher degree curves usually have genus > 1 and, as we shall see, typically have only finitely many rational solutions, though this is a very deep result.

8.3. Lifting solutions. Gauss discovered that if an equation has solutions mod p then one can often use those solutions to determine solutions to the same equation mod p^k:

Proposition 8.3.1. *Suppose that p does not divide a and that $u^n \equiv a \pmod p$. If p does not divide n then, for each integer $k \geq 2$, there exists a unique congruence class $b \pmod{p^k}$ such that $b^n \equiv a \pmod{p^k}$ and $b \equiv u \pmod p$.*

Proof. We prove this by induction on $k \geq 2$. We may assume that there exists a unique congruence class $b \pmod{p^{k-1}}$ such that $b^n \equiv a \pmod{p^{k-1}}$ and $b \equiv u \pmod p$. Therefore if $B^n \equiv a \pmod{p^k}$ and $B \equiv u \pmod p$ then $B^n \equiv a \pmod{p^{k-1}}$ and so $B \equiv b \pmod{p^{k-1}}$. Writing $B = b + mp^{k-1}$ we have

$$B^n = (b + mp^{k-1})^n \equiv b^n + nmp^{k-1}b^{n-1} \pmod{p^k}$$

which is $\equiv a \pmod{p^k}$ if and only if

$$m \equiv \frac{a - b^n}{np^{k-1}b^{n-1}} \equiv \frac{u}{an} \cdot \frac{a - b^n}{p^{k-1}} \pmod p,$$

as $ub^{n-1} \equiv u^n \equiv a \pmod p$.

Exercise 8.3.1. Show that if prime $p \not| an$ then the number of solutions $x \pmod{p^k}$ to $x^n \equiv a \pmod{p^k}$ does not depend on k.

Starting with the root $b_1 = u \pmod p$ to $x^n \equiv a \pmod p$, Proposition 8.3.1 gives us a root b_k to $x^n \equiv a \pmod{p^k}$, where $b_i \equiv b_j \pmod{p^k}$ for all $i, j \geq k$. We can define the p-adic norm of $p^k r$ where $p \not| r$ as $|r|_p := p^{-k}$. With this norm we have that $|b_i - b_j|_p \leq p^{-k}$ whenever $i, j \geq k$, so that $\lim_{k\to\infty} b_k$ exists if we complete the space. This completion is called the *p-adic integers* and can be written in the form

$$a_0 + a_1 p + a_2 p^2 + \dots \text{ with each } \ 0 \leq a_i \leq p-1.$$

Thus Proposition 8.3.1 implies that the roots of $x^n = a$ in the p-adics are in 1-to-1 correspondence with the solutions to $x^n \equiv a \pmod p$.

Exercise 8.3.2. If prime $p \not| a$, show that the sequence $a_n = a^{p^n}$ converges in the p-adics. Show that $\alpha := \lim_{n\to\infty} a_n$ is a $(p-1)$st root of unity, and that all solutions to $x^{p-1} - 1$ in $\mathbb{Q}_p$ can be obtained in this way. Conclude that $i := \lim_{n\to\infty} 2^{5^n}$ is a square root of -1 in $\mathbb{Q}_5$.

Exercise 8.3.3. Suppose that prime $p \not| a$ and that m is the order of $a \pmod p$. Suppose that $a^m - 1 = p^k n$ where $k \geq 1$ and $p \not| n$. Prove that mp^ℓ is the order of $a \pmod{p^{k+\ell}}$ for each $\ell \geq 0$.

We can find p-adic roots of most equations.

Theorem 8.3.2. *Suppose that $f(x) \in \mathbb{Z}[x]$ and that p is an odd prime. If a is an integer for which $f(a) \equiv 0 \pmod p$ and $f'(a) \not\equiv 0 \pmod p$ then there is a unique p-adic root α to $f(\alpha) = 0$ with $\alpha \equiv a \pmod p$. On the other hand if α is a p-adic root of $f(\alpha) = 0$ with $|f'(\alpha)|_p = 1$ then $f(a) \equiv 0 \pmod p$ where a is an integer for which $a \equiv \alpha \pmod p$.*

This follows immediately from the following result:

Proposition 8.3.3. *Suppose that $f(x) \in \mathbb{Z}[x]$ and that p is an odd prime. If $f(a) \equiv 0 \pmod p$ and $f'(a) \not\equiv 0 \pmod p$ then for each integer k there exists a unique residue class $a_k \pmod{p^k}$ with $a_k \equiv a \pmod p$ for which $f(a_k) \equiv 0 \pmod{p^k}$.*

Proof. The Taylor expansion of polynomial $f(x)$ at a is simply the expansion of f as a polynomial in $x - a$. In fact

$$f(x) = f(a) + f'(a)(x-a) + f^{(2)}(a)\,\frac{(x-a)^2}{2!} + \dots + f^{(k)}(a)\,\frac{(x-a)^k}{k!}\,.$$

Now, proceeding by induction on $k \geq 2$ we see that if $f(A) \equiv 0 \pmod{p^{k-1}}$ we can write $A = a_{k-1} + rp^{k-1}$ for some integer r. Using the Taylor expansion we deduce that $0 \equiv f(A) \equiv f(a_{k-1} + rp^{k-1}) \equiv f(a_{k-1}) + f'(a_{k-1})rp^{k-1} \pmod{p^k}$, as p is odd. Hence r is uniquely determined to be $\equiv -f(a_{k-1})/f'(a_{k-1})p^{k-1} \equiv -(f(a_{k-1})/p^{k-1})/f'(a) \pmod p$.

Exercise 8.3.4. Show that if $f(x) \in \mathbb{Z}[x]$ has no repeated roots then there are only finitely many primes p for which there exists an integer a_p with $f(a_p) \equiv f'(a_p) \equiv 0 \pmod p$. In fact prove that p divides the discriminant of f.

9. Counting points mod p

Given a polynomial $f(x_1, \dots, x_n) \in \mathbb{Z}[x_1, \dots, x_n]$, it is natural to ask for the number of solutions mod q; that is

$$N_f(q) := \#\{(m_1, \dots, m_n) \in (\mathbb{Z}/q\mathbb{Z})^n : \ f(m_1, \dots, m_n) \equiv 0 \pmod q\}.$$

We divide f through by the gcd of the coefficients, so henceforth we assume that the gcd of the coefficients of f is 1.

Now each $f(m)$ takes some value mod q, and we might guess that those values are roughly equi-distributed (as long as f is not a constant), so that the "probability" that a random f-value is 0 (mod q) is about $1/q$. Hence we might expect that $N_f(q)$ is roughly q^{n-1} since there are q^n vectors $m \in (\mathbb{Z}/q\mathbb{Z})^n$ and each has probability $1/q$ of equalling 0.

`Exercise 9.1.1.` Use the Chinese Remainder Theorem to show that $N_f(q)$ is a multiplicative function in q, for any given f.

This implies that we can (and should) focus on q equals a prime power.

9.1. Linear and quadratic equations mod p. The simplest example is $f(x) = x$ in which case the answer is $N_f(q) = 1 = q^{1-1}$.

`Exercise 9.1.2.` (i) Suppose that $f(x) = ax + b$ where $(a, b) = 1$. Prove that $N_f(p) = 1$ if $(p, a) = 1$, and $N_f(p) = 0$ if $p|a$ where p is prime. Deduce that $N_f(q) = 1$ if $(q, a) = 1$, and otherwise $N_f(q) = 0$.

(ii) Now suppose that $f(x, y) = ax + by$ with $(a, b) = 1$. Prove that $N_f(p^e) = p^e$ for all primes p^e. Deduce that $N_f(q) = q$ for all integers q.

Moving straight on to quadratics, the simplest examples are where $f(x) \in \mathbb{Z}[x]$ has no repeated roots with degree 1 or 2, and we are asking for solutions to $y^2 \equiv f(x) \pmod p$. We use the observation:

`Exercise 9.1.3.` $\#\{m \pmod p : \ m^2 \equiv n \pmod p\} = 1 + \left(\frac{n}{p}\right)$, where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol.

Writing $g(x, y) = y^2 - f(x)$ so that $g(x, y) \equiv 0 \pmod p$ if and only if $y^2 \equiv f(x) \pmod p$ we deduce that

$$N_g(p) = \sum_{x \pmod p} 1 + \left(\frac{f(x)}{p}\right) = p + \sum_{n \pmod p} \left(\frac{f(n)}{p}\right).$$

`Exercise 9.1.4.` Prove that if $f(n) = an + b$ where $p \nmid a$ then $\sum_{1 \le n \le p} (f(n)/p) = 0$.

When $f(x)$ is a quadratic polynomial, things are a bit more complicated.

`Exercise 9.1.5.` Show that if $p \nmid 2a$ then

$$\#\{x \pmod p : \ m \equiv x^2 \pmod p\} - 1 = \left(\frac{a}{p}\right)(\#\{x \pmod p : \ m \equiv ax^2 \pmod p\} - 1).$$

Taking $m = y^2 - b$ deduce that

$$\sum_{1 \le n \le p} \left(\frac{an^2 + b}{p}\right) = \left(\frac{a}{p}\right) \sum_{1 \le n \le p} \left(\frac{n^2 + b}{p}\right).$$

Show that the solutions x, y mod p to $y^2 \equiv x^2 + b \pmod p$ are in 1-to-1 correspondence with the solutions r, s mod p to $rs \equiv b \pmod p$. Deduce that $\sum_{1\le n\le p} \left(\frac{n^2+b}{p}\right) = -1$, and $\sum_{1\le n\le p} \left(\frac{an^2+b}{p}\right) = -\left(\frac{a}{p}\right)$.

`Exercise 9.1.6.` Deduce that for $g(x, y) = y^2 - (ax^2 + bx + c)$ and $\Delta := b^2 - 4ac$, with p an odd prime that does not divide (a, b, c), that we have

$$N_g(p) = \begin{cases} p - \left(\frac{a}{p}\right) & \text{if } p\not| a\Delta; \\ p + (p-1)\left(\frac{a}{p}\right) & \text{if } p\not| a,\ p|\Delta; \\ p & \text{if } p|a,\ p\not| b; \\ 1 + \left(\frac{c}{p}\right) & \text{if } p|a,\ p|b. \end{cases}$$

`Exercise 9.1.7.` Prove that there is always a solution to $ax^2 + by^2 \equiv c \pmod p$. (Hint: How many distinct residues are there of the form $ax^2 \pmod p$?)

General quadratics in two variables. We will show that $ax^2+bxy+cy^2+dx+ey+f = 0$ can always be put in one of the above forms mod p, by a suitable (and invertible) linear change of variables, when p is an odd prime. We may assume that $p\not|(a, b, c)$ else the above equation is linear mod p. We may also assume that $p\not| c$ else if $p|c$ and $p\not| a$ we interchange x and y, and if $p|c$ and $p|a$ then we replace x by $x + y$. But now we replace y by $y - (bx + e)/2c \pmod p$ (as $2c$ is invertible mod p), and then we obtain an equation of the form $y^2 = Ax^2 + Bx + C$ as desired.

`Exercise 9.1.8.` Use much the same idea to show that if p is an odd prime, then any ternary quadratic $ax^2 + bxy + cy^2 + dxz + eyz + fz^2 = 0$ where $p\not|(a, b, c, d, e, f)$, may be turned into a diagonal quadratic form, after a suitable (and invertible) linear change of variables. Hence we may assume that our form is the diagonal quadratic $ax^2 + by^2 + cz^2 \pmod p$.

Suppose that $p\not| abc$. We then can divide through by $-c \pmod p$, and hence assume that $c = -1$. Now if we have a solution with $y \not\equiv 0 \pmod p$, $x \equiv ny \pmod p$, $z \equiv vy$ then $an^2 + b \equiv v^2 \pmod p$; and vice-versa, if $an^2 + b \equiv v^2 \pmod p$ then we have $p-1$ solutions to $ax^2 + by^2 \equiv z^2 \pmod p$ by taking $x \equiv ny \pmod p$, $z \equiv vy$. The total number of such solutions is $(p-1)\left(p - \left(\frac{a}{p}\right)\right)$ by the first part of exercise 9.1.6. Now if $y \equiv 0 \pmod p$ we have the solutions to $ax^2 \equiv z^2 \pmod p$, and there are $p + (p-1)\left(\frac{a}{p}\right)$ such solutions by the second part of exercise 9.1.6. Hence the total number of solutions to $ax^2 + by^2 \equiv z^2 \pmod p$, is

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

This is a much simpler formula than the four cases of exercise 9.1.6! It is also not difficult to deduce the non-homogenous cases from this. The case with $p = 2$ is easy to work on directly, particularly after using Fermat's Little Theorem. We summarize:

Proposition 9.1. *Suppose that p is a prime that does not divide integers a, b, c. Then*
(i) If $f(x, y, z) = ax + by + cz$ then $N_f(p) = p^2$.
(ii) If $f(x, y, z) = ax^2 + by^2 + cz^2$ then $N_f(p) = p^2$.

Alternate proof of Proposition 9.1(ii). We will being by studying the solutions $x, y \pmod p$ to $ax^2+by^2+c \equiv 0 \pmod p$. By exercise 9.1.7 we know that there exists a solution (x_0, y_0). For any other solution we can write $x = x_0 + u$ and $y = y_0 + tu$, except if $x = x_0$ whence $y \equiv -y_0 \pmod p$. Substituting in we obtain $u \equiv -2(ax_0 + by_0t)/(a + bt^2) \pmod p$, and therefore

$$a\left(\frac{bx_0t^2 - 2by_0t - ax_0}{a+bt^2}\right)^2 + b\left(\frac{ay_0 - 2ax_0t - by_0t^2}{a+bt^2}\right)^2 + c \equiv 0 \pmod p.$$

Writing $t = n/m$ and multiplying though by $(am^2 + bn^2)^2$ we obtain

$$a\left(x_0(bn^2 - am^2) - 2by_0mn\right)^2 + b\left(y_0(am^2 - bn^2) - 2ax_0mn\right)^2 + c(am^2+bn^2)^2 \equiv 0 \pmod p.$$

These are distinct except that $(m, n) \sim (-m, -n)$. So if $(m, n) \not\equiv (0, 0)$ then we multiply inside the squares by δ which equals 1 or r where $(r/p) = -1$.[14]

Let $N_f^*(p)$ be the number of solutions to $ax^2 + by^2 + cz^2 \equiv 0 \pmod p$ with $p \nmid xyz$. The solutions with $p|xyz$ are $(0, 0, 0)$, together with $(0, y, ty)$ where $y \not\equiv 0 \pmod p$ and $t^2 \equiv -b/c \pmod p$, and analogous solutions when $y = 0$ and when $z = 0$. Hence there are $\le 1 + 6(p - 1)$ such solutions, and so $N_f^*(p) \ge p^2 - 1 - 6(p - 1) > 0$ if $p > 7$. Otherwise we find that there are solutions with $p \nmid xyz$ except $N_f(5) = 0$ when $(a/5) = (b/5) = (c/5)$, and $N_f(3) = 0$ when a, b, c are not all congruent mod 3.

9.2. The diagonal cubic equation mod p. Following Proposition 9.1 one might hazard the guess that $N_f(p)$ equals p^2 for $f(x, y, z) = ax^3 + by^3 + cz^3$, when $p \nmid 6abc$. We can test this theory with $p = 7$. A calculation reveals that $N_f(7)$ always equals one of $19, 55$ or 73, so our guess is wrong.

`Exercise` 9.2.1. Prove that $N_f(p) = p^2$ if $p = 3$ or if $p \equiv 2 \pmod 3$. (Hint: In the latter case consider what is the set of cubes mod p.)

Note that if we multiply a, b, c through by the same non-zero constant, or we re-order them, then we do not effect the number of solutions. We also note that the set of values $\{ax^3 : x \pmod p\}$ and the frequency with which each value appears, depends only on which coset of $(\mathbb{Z}/p\mathbb{Z})/(\mathbb{Z}/p\mathbb{Z})^3$ that a belongs to. Let g be a primitive root mod p. If we combine these observations then letting $T_{i,j,k}(p)$ denote the number of solutions to $g^ix^3 + g^jy^3 + g^kz^3 \equiv 0 \pmod p$, we find that every possibility is equal to one of $T_{0,0,0}(p)$, $T_{0,0,1}(p)$, $T_{0,0,2}(p)$ and $T_{0,1,2}(p)$.

`Exercise` 9.2.3. Use the definition to show that $T_{0,0,0}(p) + T_{0,0,1}(p) + T_{0,0,2}(p) = 3p^2$, by studying the number of solutions, for fixed x, y, of $x^3 + y^3 + g^kz^3 \equiv 0 \pmod p$ as k varies over $0, 1, 2$, and z varies mod p.

`Exercise` 9.2.4. Similarly establish that $T_{0,1,0}(p) + T_{0,1,1}(p) + T_{0,1,2}(p) = 3p^2$. Deduce that $T_{0,1,2}(p) = T_{0,0,0}(p)$.

[14] This needs fixing up nicely, perhaps with reference to a previous parametrization. Also comparison to the degree one case.

Hence we see that if $p \equiv 1 \pmod 3$ then there are at most three different possible values of $N_f(p) - p^2$, and their sum is 0. We have seen that we need only restrict our attention to

$$x^3 + y^3 + dz^3 \equiv 0 \pmod p.$$

Let $u = (x+y)/72d$ and if $u \not\equiv 0 \pmod p$ then $v = 36d - y/u,\ w = -z/6u$, to obtain

$$v^2 = w^3 - 432d^2.$$

If $u \equiv 0 \pmod p$ then $p|(x+y)|(x^3+y^3)$ so that $z \equiv 0 \pmod p$, so we have the p solutions $(t, -t, 0) \pmod p$. On the other hand given any solution v, w we can take any $u \not\equiv 0 \pmod p$ and $x = u(36d+v), y = u(36d-v), z = -6uw$. Hence,

$$T_{0,0,k}(p) = p + (p-1)N_{y^2-(x^3-432g^{2k})}(p).$$

Hence we wish to count the solutions to $y^2 \equiv x^3 + b \pmod p$. In fact counting the solutions to y^2 equals a cubic mod p is much more difficult than y^2 equals a linear or a quadratic, and we will be only able to attack several special cases in this chapter. In general if $p > 3$ we can change variables to the problem of determining

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 + ax + b \pmod p\}.$$

In the next two subsections we focus on the two special cases $b = 0$ and $a = 0$.

9.3. The equation $y^2 = x^3 + ax$. Let $N_p(a)$ be the number of pairs $(x, y) \pmod p$ such that $y^2 \equiv x^3 + ax \pmod p$. We know that $N_p(a) = p + S_a$ where we define

$$S_a := \sum_{1 \le n \le p} \left(\frac{n^3 + an}{p} \right).$$

Note that $(-n)^3 + a(-n) = -(n^3 + an)$ so that if $p \equiv 3 \pmod 4$ then $\left(\frac{n^3+an}{p}\right) + \left(\frac{(p-n)^3+a(p-n)}{p}\right) = 0$ and therefore $S_a = 0$, so that $N_p(a) = p$. So henceforth assume that $p \equiv 1 \pmod 4$. Calculations when $a = -1$ (that is for the curve $y^2 = x^3 - x \pmod p$) reveal that

$$N_5 = 7,\ N_{13} = 7,\ N_{17} = 15,\ N_{29} = 39,\ N_{37} = 39,\ N_{41} = 31,\ N_{53} = 39,\ N_{61} = 71,\ N_{73} = 79.$$

These are all odd and close to p, so we compute $E_p = E_p(-1) := (N_p - p)/2$:

$$E_5 = 1,\ E_{13} = -3,\ E_{17} = -1,\ E_{29} = 5,\ E_{37} = 1,\ E_{41} = -5,\ E_{53} = -7,\ E_{61} = 5,\ E_{73} = 3, \dots$$

Now we have all odd numbers... any guesses as to what these odd numbers are? Bear in mind that we are only dealing with primes $\equiv 1 \pmod 4$, so one might expect these odd numbers to reflect a property that primes $\equiv 1 \pmod 4$ have, but not primes $\equiv 3 \pmod 4$.

If you can predict $|E_p|$ then try to predict the sign of E_p. The answer to these questions will be revealed but it is instructive to play with the data for a while.

`Exercise 9.3.1.` By making the substitution $n \equiv m/b \pmod p$ show that $S_{ab^2} = \left(\frac{b}{p}\right) S_a$. Show that $S_0 = 0$.

Let $T_+ = S_{-1}$ and $T_- = S_a$ where $(a/p) = -1$. By exercise 9.3.1 we have

$$\begin{aligned}\frac{p-1}{2}\,(T_+^2 + T_-^2) &= \sum_{a \pmod p} S_a^2 = \sum_{a \pmod p} \sum_{m,n \pmod p} \left(\frac{m^3+am}{p}\right)\left(\frac{n^3+an}{p}\right)\\ &= \sum_{m,n \pmod p} \left(\frac{mn}{p}\right) \sum_{a \pmod p} \left(\frac{(a+m^2)(a+n^2)}{p}\right)\\ &= p \sum_{\substack{m,n \pmod p\\ m^2 \equiv n^2 \pmod p}} \left(\frac{mn}{p}\right) - \sum_{m,n \pmod p} \left(\frac{mn}{p}\right)\end{aligned}$$

by exercise 9.1.6. The second sum here is clearly 0. In the first sum we get 0 when $n = 0$. For all other n we have $m \equiv \pm n \pmod p$ so that $\left(\frac{mn}{p}\right) = \left(\frac{\pm n^2}{p}\right) = 1$, so the sum equals $(p-1)\cdot 2$. Therefore we have

$$T_+^2 + T_-^2 = 4p.$$

Now $\left(1+\left(\frac{n-1}{p}\right)\right)\left(1+\left(\frac{n}{p}\right)\right)\left(1+\left(\frac{n+1}{p}\right)\right) = 0$ or 8 unless $n \equiv -1, 0$ or $1 \pmod p$. For $n = \pm 1$ we get $2\left(1+\left(\frac{2}{p}\right)\right)$ which sum to 0 or 8, and for $n = 0$ we get 4. Hence

$$\sum_{n \pmod p} \left(1+\left(\frac{n-1}{p}\right)\right)\left(1+\left(\frac{n}{p}\right)\right)\left(1+\left(\frac{n+1}{p}\right)\right) \equiv 4 \pmod 8.$$

But each $\sum_{n \pmod p}\left(\frac{n+i}{p}\right) = 0$ and each $\sum_{n \pmod p}\left(\frac{n+i}{p}\right)\left(\frac{n+j}{p}\right) = -1$, so the above expands out to being $p-3+S_{-1} = p-3+T_+$. Hence $T_+ \equiv -(p+1) \pmod 8$. We deduce that $T_+ = 2a,\ T_- = 2b$ where a is odd; and $a^2+b^2 = p$. We choose the sign of a so that $a \equiv -(\frac{p+1}{2}) \pmod 4$. Summarizing:

Proposition 9.3. *Let p be a prime $\equiv 1 \pmod 4$, and a and b be those unique integers (up to sign) for which $p = a^2 + b^2$ with a odd and b even. Then*

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 - x \pmod p\} = p - 2(-1)^{\frac{a+b+1}{2}} a.$$

We also have

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 - k^2 x \pmod p\} = p - 2(-1)^{\frac{a+b+1}{2}} a \left(\frac{k}{p}\right)$$

for any $k \not\equiv 0 \pmod p$; and

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 - rk^2 x \pmod p\} = p - 2b\left(\frac{k}{p}\right)$$

if $(r/p) = -1$.

`Exercise` 9.3.2. Show that $S_1 = \sum_{1\le n\le p-1} \left(\frac{n+1/n}{p}\right)$. Determine the number of solutions $a, b, c \pmod p$ to $a + b \equiv c^2 \pmod p$ where $ab \equiv 1 \pmod p$.

9.4. The equation $y^2 = x^3 + b$. Since the map $x \to x^3$ is an automorphism if $p \equiv 2 \pmod 3$, we then have that

$$S_b := \sum_{1\le n\le p} \left(\frac{n^3+b}{p}\right)$$

equals 0. Hence we may assume $p \equiv 1 \pmod 3$. The map $n \to n/d$ yields that $S_b = S_{bd^3}\left(\frac{d}{p}\right)$. So define $T_i = S_{g^i}\left(\frac{g^i}{p}\right)$ for $i = 0, 1, 2$, where g is a primitive root mod p. As before we have $S_0 = 0$ and so

$$\begin{aligned}\frac{p-1}{3}\,(T_0^2+T_1^2+T_2^2) &= \sum_{b \pmod p} S_b^2 = \sum_{m,n \pmod p} \sum_{b \pmod p} \left(\frac{(b+m^3)(b+n^3)}{p}\right)\\ &= p \sum_{\substack{m,n \pmod p\\ m^3\equiv n^3 \pmod p}} 1 - \sum_{m,n \pmod p} 1\\ &= p(1+3(p-1)) - p^2 = 2p(p-1),\end{aligned}$$

yielding that $T_0^2 + T_1^2 + T_2^2 = 6p$. Note also that

$$\begin{aligned}\frac{p-1}{3}\,(T_0+T_1+T_2) &= \sum_{b \pmod p} \left(\frac{b}{p}\right) S_b = \sum_{1\le n\le p} \sum_{b \pmod p} \left(\frac{b(b+n^3)}{p}\right)\\ &= (p-1) + \sum_{1\le n\le p-1} (-1) = 0.\end{aligned}$$

Now $((n^3+b)/p) \equiv 1 \pmod 2$ unless $n^3 \equiv -b \pmod p$. There are three solutions to this for $b = 1$ and none for $b = g$ or g^2, and so T_0 is even and T_1, T_2 odd. We also have that if $m \not\equiv 0 \pmod p$ and $n^3 \equiv m \pmod p$ then there are 3 such solutions. Therefore $S_b \equiv (b/p) \pmod 3$, and so each $T_i \equiv 1 \pmod 3$. Hence if we write $T_0 = 2a$ then $a + T_1$ is divisible by 3, so call is $3b$. So $T_1 = -a - 3b$ and therefore $T_2 = -a + 3b$ as $T_0 + T_1 + T_2 = 0$. But then $T_0^2 + T_1^2 + T_2^2 = 6(a^2 + 3b^2)$ and therefore

$$p = a^2 + 3b^2.$$

Finding the sign of a is easy since $a \equiv 2 \pmod 3$.

Proposition 9.4. *Let p be a prime $\equiv 1 \pmod 3$, and a and b be those unique integers (up to sign) for which $p = a^2 + 3b^2$ with $a \equiv 2 \pmod 3$. Then, for any $k \not\equiv 0 \pmod p$,*

$$\#\{x, y \pmod p :\ y^2 \equiv x^3 + k^3 \pmod p\} = p - 2a\left(\frac{k}{p}\right).$$

Moreover, for $i = 1$ or 2, we have

$$\#\{x, y \pmod p : \ y^2 \equiv x^3 + g^i k^3 \pmod p\} = p + (a + 3(-1)^i b) \left(\frac{k}{p}\right).$$

Now we can write $4p$ as $x^2 + 3y^2$ in three different ways:

$$4p = (2a)^2 + 3(2b)^2 = (a + 3b)^2 + 3(a - b)^2 = (a - 3b)^2 + 3(a + b)^2,$$

so that $N_{y^2-(x^3+g^i)}(p) = p + A_i$ where $4p = A_i^2 + 3B_i^2$. Hence

$$T_{0,0,k}(p) = p + (p - 1)N_{y^2-(x^3-2\cdot 6^3\cdot g^{2k})}(p) = p^2 + (p - 1)A_i.$$

This implies that the number of solutions to $ax^3 + by^3 + cz^3 \equiv 0 \pmod p$ with $p \nmid xyz$ is $\geq p^2 - 2(p - 1)\sqrt{p} - (1 + 9(p - 1)) > 0$ for $p \geq 17$. We check for $p = 7$ and $p = 13$. We find that there are non-zero solutions except for $T^*_{0,0,0}(13) = T^*_{0,0,0}(7) = T^*_{0,0,1}(7) = 0$.
Example: The curve $3x^3 + 4y^3 + 5z^3 = 0$ obviously has solutions in the reals, and we now see that it has solutions mod p^r for all $r \geq 1$ for all primes p, except perhaps for $p = 2, 3, 5, 7, 13$. We check that there are 108 coprime solutions mod 13, 16 coprime solutions mod 5. We can take $x = 0, y = 1$ and $z^3 \equiv -4/5 \pmod{3^r}$ and there are always three solutions. We can take $x = 1, z = 0$ and $y^3 \equiv -3/4 \pmod{7^r}$ and there are always three solutions. Finally we can take $x = 1, y = 0$ and $z^3 \equiv -3/5 \pmod{2^r}$ and always find a solution. Hence there are solutions locally everywhere to $3x^3 + 4y^3 + 5z^3 = 0$.

It is worth emphasizing the important similarities in the last two subsections. We found that the number of solutions $(x, y) \pmod p$ to $y^2 = x^3 + ax \pmod p$, and also to $y^2 = x^3 + b \pmod p$ is $p - \alpha - \overline{\alpha}$ where α is an algebraic integer of degree two, with $|\alpha| = \sqrt{p}$. More precisely in the first case we have $\alpha = u + iv$ for integers u and v satisfying $u^2 + v^2 = p$, so that α has minimal polynomial $X^2 - 2uX + p$. In the second case we have $\alpha = \frac{r+s\sqrt{3}}{2}$ where r and s are integers satisfying $r^2 + 3s^2 = 4p$, so that α has minimal polynomial $X^2 - rX + p$. We also remark that as we run through the values of a and b mod p, we also run through all such possibilities for α. For a general (non-degenerate) cubic we will also find that the number of points mod p is $p - \alpha - \overline{\alpha}$ where α is an algebraic integer of degree two, with $|\alpha| = \sqrt{p}$, though we must develop quite a bit of algebra to prove this. It is useful to count the points in $\mathbb{F}_p$, rather than mod p, in which case it is natural to add the point at infinity, and thus the number of points is $\#E(\mathbb{F}_p) = p + 1 - \alpha - \overline{\alpha}$.

The claims in the last paragraph imply Hasse's amazing theorem that

$$|\#\{x, y \pmod p : \ y^2 \equiv x^3 + ax + b \pmod p\} - p| \leq 2\sqrt{p},$$

which means we can identify the precise number. This was generalized by Weil, so that if $f(x)$ is a polynomial of degree d that has no repeated factors mod p then

$$|\#\{x, y \pmod p : \ y^2 \equiv f(x) \pmod p\} - p| \leq (d - 1)\sqrt{p}.$$

9.5. Counting solutions mod p. We are going to count, mod p, the number of solutions to $f(x) \equiv 0 \pmod p$. To do so we need:

`Exercise 9.5.1.` Reprove Gauss's result that $\sum_{a=0}^{p-1} a^k \equiv 0 \pmod p$ unless $p-1$ divides k, and $k > 0$ in which case the sum is $\equiv -1 \pmod p$.

Let $f(x_1, x_2, \dots, x_n) \in \mathbb{Z}[x_1, x_2, \dots, x_n]$ be of degree d.[15] The number of solutions to $f \equiv 0 \pmod p$ is congruent to

$$\sum_{m_1,\dots,m_n \pmod p} 1 - f(m_1, \dots, m_n)^{p-1} \pmod p,$$

since

$$1 - f(m)^{p-1} \equiv \begin{cases} 1 \pmod p & \text{if } f(m) \equiv 0 \pmod p; \\ 0 \pmod p & \text{if } f(m) \not\equiv 0 \pmod p, \end{cases}$$

by Fermat's Little Theorem. The first term evidently sums to $p^n \equiv 0 \pmod p$. When we expand the second term, we get a sum of terms, each of total degree $\le d(p-1)$. For the sum, over the m_i, of the term to be non-zero, the degree in each variable must be $\ge p-1$, and so the total degree of the term must be $\ge n(p-1)$. This implies that $n(p-1) \le d(p-1)$ and hence $d \ge n$. Hence we have proved:

The Chevalley-Warning theorem. *If we suppose that f has degree $d < n$ then*

$$\#\{m_1, \dots, m_n \pmod p : \ f(m_1, \dots, m_n) \equiv 0 \pmod p\} \equiv 0 \pmod p.$$

Therefore if $f(0, 0, \dots, 0) = 0$, that is f has a zero constant term, then there are $\ge p-1$ distinct non-zero solutions to $f(m_1, \dots, m_n) \equiv 0 \pmod p$.

One example is the equation $ax^2 + by^2 + cz^2 \equiv 0 \pmod p$.

`Exercise 9.5.2.` Show that the number of solutions $x, y \pmod p$ to $y^2 = f(x) \bmod p$ is congruent to

$$\sum_{n \pmod p} f(n)^{\frac{p-1}{2}} \pmod p.$$

Now if $f(x) = ax^2 + bx + c$ with $p \not| (a, b)$ then we have, by the multinomial theorem,

$$(an^2 + bn + c)^{\frac{p-1}{2}} = \sum_{i+j+k=\frac{p-1}{2}} \frac{\frac{p-1}{2}!}{i!j!k!} (an^2)^i (bn)^j c^k.$$

The exponent on n is $2i + j$, and for the sum over $n \pmod p$ to be non-zero this must be divisible by $p-1$ and > 0. Since $i + j + k = \frac{p-1}{2}$ the only possibility arises with $i = \frac{p-1}{2}$, $j = k = 0$, and so

$$\sum_{n \pmod p} f(n)^{\frac{p-1}{2}} \equiv -a^{\frac{p-1}{2}} \pmod p.$$

[15] The degree of $x_1^{e_1} x_2^{e_2} \dots x_n^{e_n}$ is $e_1 + e_2 + \dots + e_n$.

Hence the number of solutions to $y^2 = ax^2+bx+c \bmod p$ is congruent to $-a^{\frac{p-1}{2}} \equiv -(a/p) \pmod p$, which is confirmed by exercise 9.1.6.

What about the number $N = N_{a,b}(p)$ of solutions to $y^2 = x^3 + ax + b \bmod p$? By Hasse's Theorem, we know that $|N-p| < 2\sqrt{p}$, so we can identify N by knowing the value of $N \pmod p$. Rather than proceeding directly as above we simply the question a little first:

Exercise 9.5.3. Suppose that $a, b \not\equiv 0 \pmod p$.

i. Show that if $b^2/a^3 \equiv t \pmod p$ where $t \not\equiv 0 \pmod p$ then there exists $m \pmod p$ such that $a \equiv m^2 t,\ b \equiv m^3 t^2 \pmod p$. (The ratio b^2/a^3 is associated to the "j-invariant" that we discuss later.)

ii. Substitute $y = mY,\ x = mX,\ a = m^2 t,\ b = m^3 t^2$ in the equation $y^2 = x^3 + ax + b \bmod p$ to show that $N_{a,b}(p) - p = \left(\frac{b}{p}\right)(N_{t,t^2}(p) - p)$.

By the trinomial theorem, we know that $N_{a,b}(p)$ is congruent to

$$\sum_{\substack{i+j+k=\frac{p-1}{2}\\ i,j,k\geq 0}} \sum_{n \pmod p} \frac{\frac{p-1}{2}!}{i!j!k!}\,(n^3)^i (an)^j b^k \equiv - \sum_{\substack{i+j+k=\frac{p-1}{2}\\ 3i+j=p-1}} \frac{\frac{p-1}{2}!}{i!j!k!}\,a^j b^k.$$

If $a = 0$ then the only non-zero term has $j = 0,\ i = \frac{p-1}{3}$, so that $p \equiv 1 \pmod 3$ and

$$N_{0,b}(p) \equiv -\binom{\frac{p-1}{2}}{\frac{p-1}{6}} b^{\frac{p-1}{6}} \pmod p.$$

If $b = 0$ then the only non-zero term has $k = 0,\ i = \frac{p-1}{4}$, so that $p \equiv 1 \pmod 4$ and

$$N_{a,0}(p) \equiv -\binom{\frac{p-1}{2}}{\frac{p-1}{4}} a^{\frac{p-1}{4}} \pmod p.$$

Exercise 9.5.4. Compare these results to Propositions 9.3 and 9.4 to deduce some surprising congruences for $\binom{\frac{p-1}{2}}{\frac{p-1}{6}}$ and $\binom{\frac{p-1}{2}}{\frac{p-1}{4}} \bmod p$.

Finally we have the case where $a = t, b = t^2$, and so

$$N_{t,t^2}(p) \equiv \sum_{\frac{p-1}{4} \leq i \leq \frac{p-1}{3}} c_i t^i \quad \text{where } c_i := -\frac{\frac{p-1}{2}!}{i!(p-1-3i)!(2i-\frac{p-1}{2})!}\ .$$

There is another way to parametrize a subclass of elliptic curves mod p with interesting results: We study the equations $y^2 \equiv f(x) \pmod p$ where f is a cubic that splits into linear factors $\pmod p$.

Exercise 9.5.5. Show that we make a linear change of variable to put f in the form $f(x) \equiv ax(x-1)(x-\lambda) \pmod p$, for some non-zero residues $a, \lambda \pmod p$. Use the first part of exercise 9.1.5 to show that $\#\{(x,y) :\ y^2 \equiv ax(x-1)(x-\lambda) \pmod p\} - p = \left(\frac{a}{p}\right)(\#\{(x,y) :\ y^2 \equiv x(x-1)(x-\lambda) \pmod p\} - p)$.

Now

$$\#\{(x,y):\ y^2\equiv x(x-1)(x-\lambda)\pmod p\}\equiv \sum_{x\pmod p}(x(x-1)(x-\lambda))^{\frac{p-1}{2}}$$
$$\equiv \sum_{0\le i,j\le\frac{p-1}{2}}\binom{\frac{p-1}{2}}{i}\binom{\frac{p-1}{2}}{j}(-1)^{\frac{p-1}{2}-i}(-\lambda)^j\sum_{x\pmod p}x^{p-1+i-j}$$
$$\equiv -(-1)^{\frac{p-1}{2}}\sum_{0\le i\le\frac{p-1}{2}}\binom{\frac{p-1}{2}}{i}^2\lambda^i\pmod p.$$

Exercise 9.5.6. Allow the abuse of notation $\binom{\frac{-1}{2}}{i}$ and prove that $\binom{\frac{-1}{2}}{i}\equiv\binom{\frac{p-1}{2}}{i}\pmod p$ for $0\le i\le p-1$. Re-write the last formula as $-\left(\frac{-1}{p}\right)\sum_{i=0}^{p-1}\binom{\frac{-1}{2}}{i}^2\lambda^i\pmod p$. Also prove that $\binom{\frac{p-1}{2}}{i}\equiv(-1)^i\binom{\frac{p-1}{2}+i}{i}\pmod p$, and so the last formula can also be written as $-(-1)^{\frac{p-1}{2}}\sum_{0\le i\le\frac{p-1}{2}}\binom{\frac{p-1}{2}+i}{i}\binom{\frac{p-1}{2}}{i}(-\lambda)^i\pmod p$

9.6. Degenerate elliptic curves mod p with $p\ge 5$. We will see in chapter 11 that there is a standard form for an elliptic curve mod p, for $p>3$, namely $y^2=x^3+ax+b$. In sections 9.3 and 9.4 we studied how many points there are on the curve mod p, in the special cases where $b=0$ and $a=0$, respectively. Now we study the case where the equation degenerates, that is the cubic x^3+ax+b does not have distinct roots mod p (though it may do over $\mathbb{Q}$).

Exercise 9.6.1. Give an example of a cubic polynomial $x^3+ax+b\in\mathbb{Z}[x]$ which has distinct roots in $\mathbb{C}$ but does not have distinct roots mod p, for a given prime p.

If x^3+ax+b has (at least) two roots in common mod p, we can re-write our curve, after translating x by the double root, as $y^2=x^2(x+c)$. We have one rational point with $x=0$. Otherwise we may let $t\equiv y/x\pmod p$ so that $t^2\equiv x+c\pmod p$. Hence we can parametrize solutions, other than $(0,0)$ as $(t^2-c,t(t^2-c))$.

Exercise 9.6.2. Deduce that there are $p-(\frac{c}{p})$ *distinct* solutions $x,y\pmod p$ to $y^2\equiv x^2(x+c)\pmod p$. Show that in the finite field $\mathbb{F}_{p^k}$ there are exactly $p^k+(-(\frac{c}{p}))^k$ points $x,y\in\mathbb{F}_{p^k}$ for which $y^2=x^2(x+c)$. Express $(\frac{c}{p})$ in terms of $(\frac{b}{p})$.

9.7. Lifting solutions. Suppose that we are given a solution (x,y,z) to $ax^q+by^q+cz^q\equiv 0\pmod p$ where $p\nmid abcq$. We look for a solution mod p^2, in the same congruence classes mod p; that is we want $a(x+ip)^q+b(y+jp)^q+c(z+kp)^q\equiv 0\pmod{p^2}$. Expanding this, reducing mod p^2, we have a solution if and only if

$$q(aix^{q-1}+bjy^{q-1}+ckz^{q-1})\equiv -\frac{ax^q+by^q+cz^q}{p}\pmod p.$$

Now if $p\nmid ax$ then we get a unique value of $i\pmod p$ for each pair j,k, and hence there are p^2 such triples $(i,j,k)\pmod p$. Similarly we can proceed if $p\nmid by$ or $p\nmid cz$; that is unless $p|(x,y,z)$. We can continue lifting in a similar way modulo higher powers of p, by the analogous proof. Hence if $M_{a,b,c;q}(p^k)$ is the number of triples $x,y,z\pmod{p^k}$ for which $ax^q+by^q+cz^q\equiv 0\pmod{p^k}$ and $p\nmid(x,y,z)$, then $M_{a,b,c;q}(p^k)=p^{2(k-1)}M_{a,b,c;q}(p)$.

9.8. Simultaneous Pell equations. Given integers a, b, c, d for which $p \not| abcd(ad-bc)$ we define $P_{a,b,c,d}(p)$ to be the number of solutions $x, y, u, v \pmod p$ to

$$ax^2 + by^2 = u^2$$
$$cx^2 + dy^2 = v^2.$$

Therefore $P_{a,b,c,d}(p) - p^2$

$$= \sum_{x,y \pmod p} \left(1 + \left(\frac{ax^2+by^2}{p}\right)\right)\left(1 + \left(\frac{cx^2+dy^2}{p}\right)\right) - p^2$$
$$= \sum_{x,y \pmod p} \left(\frac{ax^2+by^2}{p}\right)\left(\frac{cx^2+dy^2}{p}\right)$$
$$= \sum_{u,v \pmod p} \left(1 + \left(\frac{u}{p}\right)\right)\left(1 + \left(\frac{v}{p}\right)\right)\left(\frac{au+bv}{p}\right)\left(\frac{cu+dv}{p}\right).$$

Now, splitting the sum into the parts where $v = 0$, and $v \neq 0$ so we may take $t \equiv u/v \pmod p$, we obtain

$$\sum_{u,v \pmod p} \left(\frac{u(au+bv)(cu+dv)}{p}\right) = \sum_{u \pmod p} \left(\frac{acu^3}{p}\right) + \sum_{v,t \pmod p} \left(\frac{t(at+b)(ct+d)v}{p}\right) = 0.$$

With this argument we thus obtain that

$$P_{a,b,c,d}(p) - p^2 = \sum_{u,v \pmod p} \left(\frac{uv(au+bv)(cu+dv)}{p}\right) = (p-1) \sum_{u,v \pmod p} \left(\frac{t(at+b)(ct+d)}{p}\right)$$

writing $u \equiv tv \pmod p$ when $v \not\equiv 0 \pmod p$. We deduce that

$$P_{a,b,c,d}(p) - p^2 = (p-1)(\#E(\mathbb{F}_p) - p - 1),$$

where $E: \ y^2 = x(ax+b)(cx+d)$.

10. Combinatorics, partitions, and an example of modularity

10.1. Partitions. Let $p(n)$ denote the number of ways of partitioning n into smaller integers. For example $p(7) = 15$ since

$$\begin{aligned} 7 &= 6+1 = 5+2 = 5+1+1 = 4+3 = 4+2+1 = 4+1+1+1 = 3+3+1 \\ &= 3+2+2 = 3+2+1+1 = 3+1+1+1+1 = 2+2+2+1 \\ &= 2+2+1+1+1 = 2+1+1+1+1+1 = 1+1+1+1+1+1+1. \end{aligned}$$

Euler observed that there is a beautiful generating function for $p(n)$: In the generating function $p(n)$ is the coefficient of t^n, and for each partition $n = a_1 + \ldots + a_k$ we can think of t^n as $t^{a_1} \ldots t^{a_k}$. Splitting this product up into the values of the a_i, but taking $(t^a)^j$ if there are j of the a_i's that equal a, we see that

$$\sum_{n\geq 0} p(n)t^n = \prod_{a\geq 1}\left(\sum_{j\geq 0}(t^a)^j\right) = \frac{1}{(1-t)(1-t^2)(1-t^3)\ldots}.$$

Similarly the generating function for the number of partitions into odd parts is $1/(1-t)(1-t^3)(1-t^5)\ldots$, for the number of partitions with no repeated parts is $(1+t)(1+t^2)(1+t^3)\ldots$, etc.

`Exercise 10.1.` Deduce that the number of partitions of n into odd parts is equals to the number of partitions of n with no repeated parts.

Partitions can be represented by rows and columns of dots in a *Ferrers diagram*; for example $27 = 11 + 7 + 3 + 3 + 2 + 1$ is represented by

o o o o o o o o o o o
o o o o o o o
o o o
o o o
o o
o

the first row having 11 dots, the second 7, etc. Now, reading the diagram in the other direction yields the partition $27 = 6+5+4+2+2+2+1+1+1+1$. This bijection between partitions is at the heart of many beautiful theorems about partitions. For example if a partition has m parts then its "conjugate" has largest part m. Using generating functions, we therefore find that the number of partitions with $\leq m$ parts, equals the number of partitions with all parts $\leq m$, which has generating function

$$\frac{1}{(1-t)(1-t^2)(1-t^3)\ldots(1-t^m)}.$$

Using the Ferrers diagram the partitions come in pairs, other than those that are *self-conjugate*; that is the conjugate partition is the same as the original partition. This

implies a symmetry about the diagonal axis of the diagram. Hence a self-conjugate Ferrers diagram looks like

$o\ o\ o\ o\ o\ o$	1 1 1 1 1 1 1
$o\ o\ o\ o\ o$	1 2 2 2 2
$o\ o\ o$	1 2 3
$o\ o$	1 2
$o\ o$	1 2
o	1

yielding $19 = 6+5+3+2+2+1$. We have constructed another partition of 19, using the same Ferrers diagram. The first element is obtained by peeling off the top row and top column (which have a total of 11 entries), then what's left of the second row and second column (7 remaining entries), and finally the single element left (1 entry), yielding the partition $19 = 11 + 7 + 1$.

`Exercise 10.2.` Prove that there is a bijection between self-conjugate partitions and partitions where all the entries are odd and distinct. Give an elegant form for the generating function for the number of self-conjugate partitions.

10.2. The numerology of partitions.

The sequence $p(n)$ begins $p(1) = 1$, 2, 3, 5, 7, 11, 15, 22, 30, $p(10) = 42$, 56, 77, 101, 135, 176, 231, 297, 385, 490, $p(20) = 627, \dots$ with $p(100) = 190,569,292$ and $p(1000) \approx 2.4 \times 10^{31}$. Ramanujan was intrigued by these numbers, both their growth (which seems to get quite fast) and their congruence conditions. For $n = 1000$ we see that there are roughly $10^{\sqrt{n}}$ partitions, which is an unusual function in mathematics. Hardy and Ramanujan proved the extraordinary asymptotic

$$p(n) \sim \frac{1}{4n\sqrt{3}}\, e^{\pi\sqrt{2n/3}}, \tag{10.1}$$

and Rademacher developed their idea to give an exact formula. This is too far afield for this book, but suffice to say this this one proof gave birth to the circle method, still one of the most important techniques in number theory.

Ramanujan also noted several congruences for the $p(n)$:

$$p(5k+4) \equiv 0 \pmod 5, \quad p(7k+5) \equiv 0 \pmod 7, \quad p(11k+6) \equiv 0 \pmod{11},$$

for all k. Notice that these are all of the form $p(n) \equiv 0 \pmod q$ whenever $q|24n-1$, and these seem to be the only such congruences. However Ono has found many more such congruences, only a little more complicated: For any prime $q \geq 5$ there exist primes ℓ such that $p(n) \equiv 0 \pmod q$ whenever $q\ell^3|24n-1$.

10.3. Generating functions for binary quadratic forms, and L-functions. Back in section 7.3 we studied representations on integers by a given binary quadratic form f –

it is always easier to restrict attention to when $d < 0$, which we will continue to do here. If we create the generating function for representations by f then we evidently obtain

$$\sum_{N\geq 0} R_f(N)q^N = \sum_{m,n\in\mathbb{Z}} q^{am^2+bmn+cn^2}.$$

It also pays to study the Dirichlet series with the $R_f(n)$ as coefficients, that is

$$\sum_{N\geq 1} \frac{R_f(N)}{N^s} = \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{1}{(am^2+bmn+cn^2)^s}\ .$$

We also studied there $w(d)R_d(N)$ the total number of representations of N by binary quadratic forms of discriminant d. This evidently equals the sum of the $R_f(n)$, summed over representatives f of each equivalence class of quadratic forms (as in (7.3.2)). Using (7.3.1) to create a Dirichlet series, we obtain

$$\sum_{N\geq 1} \frac{R_d(N)}{N^s} = \sum_{\ell,m\geq 1} \frac{(d/m)}{(\ell m)^s} = \sum_{\ell\geq 1} \frac{1}{\ell^s} \sum_{m\geq 1} \frac{(d/m)}{m^s} = \zeta(s)L\left(s, \left(\frac{d}{\cdot}\right)\right).$$

On the other hand, (7.3.1) also readily implies that

$$\sum_{N\geq 1} R_d(N)q^N = \sum_{N\geq 1}\sum_{\ell m=N} \left(\frac{d}{m}\right) q^N = \sum_{m\geq 1} \left(\frac{d}{m}\right) \sum_{\ell\geq 1} q^{\ell m} = \sum_{m\geq 1} \left(\frac{d}{m}\right) \frac{q^m}{1-q^m}$$

When we have class number one we can directly compare the formulae. For example if $d = -4$ then let $f = x^2 + y^2$ so we have

$$\sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} q^{m^2+n^2} = \frac{q}{1-q} - \frac{q^3}{1-q^3} + \frac{q^5}{1-q^5} - \frac{q^7}{1-q^7} + \dots,$$

and

$$\zeta(s)L\left(s, \left(\frac{-4}{\cdot}\right)\right) = \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{1}{(m^2+n^2)^s}\ .$$

`Exercise` 10.3. Do something similar for $d = -163$.

10.4. Poisson's summation formula and Jacobi's theta function. Poisson's summation formula is an extremely useful tool in analytic number theory, which can easily be deduced from basic results about Fourier series. This is not the place to discuss the theory at length (but see chapter 2 of [Da]), so we simply quote that if $f(x)$ is "well-behaved" then

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{m\in\mathbb{Z}} \hat{f}(m)$$

where the *Fourier transform* of f is defined by

$$\hat{f}(m) := \int_{-\infty}^{\infty} f(t)e^{-2i\pi mt}dt.$$

Proving this is easy if $f(x)$ is "well-behaved", that is when we swap orders of summation and integration we can understand convergence (for example, if everything we work with is absolutely convergent). One such example is the case where $f(n) = e^{-n^2x}$.

`Exercise` 10.4. Prove that if $f(n) = e^{-\pi n^2 x}$ then $\hat{f}(m) = x^{-1/2}e^{-\pi m^2/x}$.

Define Jacobi's theta function

$$\theta(s) := \sum_{n\in\mathbb{Z}} e^{i\pi n^2 s} \quad \text{for any} \quad s \in \mathbb{C} \text{ with } \mathrm{Im}(s) > 0.$$

(so that $|e^{i\pi n^2 s}| < 1$). Using Poisson's summation formula and the last exercise, we deduce the remarkable identity,

$$\theta(-1/s) \;=\; (s/i)^{1/2}\theta(s). \tag{10.2}$$

We note that we also have $\theta(s+2) = \theta(s)$, and also

$$\theta(s) + \theta(s+1) \;= 2 \sum_{n \text{ even}} e^{i\pi n^2 s} \;=\; 2\theta(4s).$$

10.5. Jacobi's powerful triple product identity. This states that if $|x| < 1$ then

$$\prod_{n\geq 1}(1-x^{2n})(1+x^{2n-1}z)(1+x^{2n-1}z^{-1}) = \sum_{m\in\mathbb{Z}} x^{m^2}z^m.$$

`Exercise` 10.5. (Proof of Jacobi's triple product identity.) Define

$$P_N(x,z) := \prod_{k=1}^{N}(1-x^{2k})(1+x^{2k-1}z)(1+x^{2k-1}z^{-1}) = \sum_{n\in\mathbb{Z}} c_{N,n}(x)\ z^n.$$

i. Prove that $c_{N,n}(x) \in \mathbb{Z}[x]$; $c_{N,n}(x) = 0$ for $n > N$, and $c_{N,-n}(x) = c_{N,n}(x)$.
ii. Prove that $P_{N+1}(x,z) = (1-x^{2N+2})(1+x^{2N+1}z)(1+x^{2N+1}z^{-1})P_N(x,z)$.
iii. Deduce that $c_{N+1,n}(x) = (1-x^{2N+2})(x^{2N+1}c_{N,n-1}(x) + (1+x^{4N+2})c_{N,n}(x) + x^{2N+1}c_{N,n+1}(x))$.
iv. Prove that $c_{N,n}(x) = x^{n^2}\prod_{m=N-n+1}^{N}(1-x^{2m})\cdot\prod_{m=N+n+1}^{2N}(1-x^{2m})$ for $0 \leq n \leq N$, for all $N \geq 1$. (Hint: Proceed by induction, using the recursion in (iii).)
v. Show that if $|x| < 1$ then $\lim_{N\to\infty} c_{N,n}(x) = x^{n^2}$. Deduce Jacobi's triple product identity.

We shall determine some useful consequences of it:
Letting $x = t^a,\ z = t^b$ and $n = k+1$ in Jacobi's triple product identity we obtain

$$\prod_{k\geq 0}(1-t^{2ak+2a})(1+t^{2ak+a+b})(1+t^{2ak+a-b}) = \sum_{m\in\mathbb{Z}} t^{am^2+bm}.$$

Some special cases include $a = 1,\ b = 0$, yielding

$$\prod_{n\geq 1}(1-t^{2n})(1+t^{2n-1})^2 = \sum_{m\in\mathbb{Z}} t^{m^2};$$

and $a = b = \pm\,\frac{1}{2}$, yielding

$$\prod_{k\geq 0}(1+t^k)(1-t^{2k+2}) = \sum_{m\in\mathbb{Z}} t^{\frac{m^2+m}{2}}.$$

`Exercise 10.6.` By writing $1+t^k$ as $(1-t^{2k})/(1-t^k)$ or otherwise, deduce that

$$\sum_{m\geq 1} t^{\frac{m^2+m}{2}} = \frac{(1-t^2)(1-t^4)(1-t^6)\dots}{(1-t)(1-t^3)(1-t^5)\dots}$$

Letting $x = t^a,\ z = -t^b$ and $n = k+1$ in Jacobi's triple product identity we obtain

$$\prod_{k\geq 0}(1-t^{2ak+2a})(1-t^{2ak+a+b})(1-t^{2ak+a-b}) = \sum_{m\in\mathbb{Z}}(-1)^m t^{am^2+bm}.$$

Some special cases include $a = 1,\ b = 0$, yielding

$$\prod_{n\geq 1}(1-t^n)(1-t^{2n-1}) = \sum_{m\in\mathbb{Z}}(-1)^m t^{m^2};$$

and $a = \frac{3}{2},\ b = \frac{1}{2}$, yielding Euler's identity,

$$\prod_{n\geq 1}(1-t^n) = \sum_{m\in\mathbb{Z}}(-1)^m t^{\frac{3m^2+m}{2}}.$$

`Exercise 10.7.` Interpret this combinatorially, in terms of the number of partitions of m into unequal parts.

`Exercise 10.8.` If $(12/.)$ is the Jacobi symbol, show that

$$t^{1/24}\prod_{n\geq 1}(1-t^n) = \sum_{m\geq 1}\left(\frac{12}{m}\right) t^{\frac{m^2}{24}}\ .$$

10.6. An example of modularity: $y^2 = x^3 - x$. Letting $x = t^4$ and $z = -1$ in Jacobi's triple product identity we obtain

$$\prod_{n\geq 1}(1-t^{8n})(1-t^{8n-4})^2 = \sum_{\substack{b\in\mathbb{Z}\\ b \text{ even}}}(-1)^{b/2}t^{b^2}.$$

Now letting $x = t^4$ and $z = -t^4u$ in Jacobi's triple product identity we obtain

$$\prod_{n\ge 1}(1-t^{8n})(1-\alpha t^{8n}u)(1-t^{8n-8}/u) = \sum_{m\in\mathbb{Z}} t^{4m^2+4m}(-u)^m.$$

In the product we change the exponent of the last term taking n in place of $n-1$, and hence we have a left over term of $1-1/u$. On the right side we pair up the term for m with the term for $-m-1$. Dividing through by $1-1/u$ then yields

$$t\prod_{n\ge 1}(1-t^{8n})(1-t^{8n}u)(1-t^{8n}/u) = \sum_{m\ge 0} t^{4m^2+4m+1}(-1)^m(u^m+u^{m-1}+\ldots+u^{-m}).$$

Taking $u=1$ we obtain

$$t\prod_{n\ge 1}(1-t^{8n})^3 = \sum_{\substack{a\ge 1\\ a \text{ odd}}} (-1)^{\frac{a-1}{2}} a t^{a^2}.$$

Multiplying these together yields

$$t\prod_{n\ge 1}(1-t^{4n})^2(1-t^{8n})^2 = \sum_{\substack{a,b\in\mathbb{Z},\ a\ge 1,\\ a \text{ odd},\ b \text{ even}}} (-1)^{\frac{a+b-1}{2}} a t^{a^2+b^2} = \sum_{n\ge 1} a_n t^n, \tag{10.3}$$

for certain integers a_n. In particular if p is a prime $\equiv 3 \pmod 4$ then $a_p = 0$; if p is a prime $\equiv 3 \pmod 4$ then writing $p = a^2+b^2$ with a odd, we have $a_p = 2(-1)^{\frac{a+b-1}{2}} a$.

Multiplicativity. We will show that the a_n are multiplicative: If odd $n = rs$ with $(r,s) = 1$ then the representations of n as the sum of two squares are given by the representations of $r = a^2+b^2$ and $s = A^2+B^2$ where a, A are odd. This pair yields the two representations

$$rs = (aA+bB)^2+(aB-Ab)^2 = (aA-bB)^2+(aB+Ab)^2.$$

These two representations contribute the following to the sum for a_{rs}:

$$\begin{aligned} 2(-1)^{\frac{aA+bB+aB-Ab-1}{2}}(aA+bB) + 2(-1)^{\frac{aA-bB+aB+Ab-1}{2}}(aA-bB) &= 4(-1)^{\frac{aA+B+b-1}{2}} aA \\ &= 2(-1)^{\frac{a+b-1}{2}} a \cdot 2(-1)^{\frac{A+B-1}{2}} A, \end{aligned}$$

since $\frac{aA+bB+aB-Ab-1}{2} \equiv \frac{aA-bB+aB+Ab-1}{2} \equiv \frac{aA+B+b-1}{2} \equiv A\cdot\frac{a+b-1}{2} + \frac{A+B-1}{2} \equiv \frac{a+b-1}{2} + \frac{A+B-1}{2} \pmod 2$. Summing these up we do deduce that

$$a_{rs} = a_r a_s \text{ if } (r,s) = 1.$$

At the prime powers. If $p \equiv 1 \pmod 4$ then we can factor $p = \pi\overline{\pi}$ in $\mathbb{Z}[i]$ for some prime $\pi = a + bi$. The possible factorizations of p^k are $\tau\overline{\tau}$ where $\tau = \pi^j\overline{\pi}^{k-j}$, each factorization counted twice. Now each $\pi^j\overline{\pi}^{k-j}$ is counted twice by the sets $\pi^i\overline{\pi}^{k-1-i} \times \pi$ and $\pi^i\overline{\pi}^{k-1-i} \times \overline{\pi}$, except π^k and $\overline{\pi}^k$ which are each counted once. Also each $\pi^j\overline{\pi}^{k-j}$ is counted twice by $\pi^i\overline{\pi}^{k-2-i} \times p$. Using the calculation from the previous subsection, we deduce that

$$a_{p^k} = a_p a_{p^{k-1}} - p a_{p^{k-2}}.$$

This implies that

$$\left(1 - \frac{a_p}{p^s} + \frac{p}{p^{2s}}\right) \sum_{k\geq 0} \frac{a_{p^k}}{p^{ks}} = 1 + \frac{a_p}{p^s} - \frac{a_p}{p^s} + \sum_{k\geq 2} \frac{a_{p^k} - a_p a_{p^{k-1}} + p a_{p^{k-2}}}{p^{ks}} = 1.$$

If $p \equiv 3 \pmod 4$ then the only representations of p^k as a sum of two squares are $(p^{k/2})^2 + 0^2$ when k is even. Now $(p^{k/2} - 1)/2 \equiv k/2 \pmod 2$ and so $a_{p^k} = 0$ if k is odd, and $a_{p^k} = (-p)^{k/2}$ is k is even. This implies, taking $k = 2j$, that

$$\left(1 + \frac{p}{p^{2s}}\right) \sum_{k\geq 0} \frac{a_{p^k}}{p^{ks}} = 1 + \sum_{j\geq 1} \frac{(-p)^j + (-p)^{j-1}}{p^{2js}} = 1.$$

Finally, for $p = 2$, we have no representation of $2^k = a^2 + b^2$ with a odd and b even.

Modularity. Combining the information in the last two subsections, since a_n is multiplicative and we know the Euler factor at each prime, we deduce that

$$\sum_{n\geq 1} \frac{a_n}{n^s} = \prod_{\substack{p \text{ prime} \\ p \text{ odd}}} \left(1 - \frac{a_p}{p^s} + \frac{p}{p^{2s}}\right)^{-1}$$

where a_p is as above (though note that $a_p = 0$ if $p \equiv 3 \pmod 4$). We see that the power series can be created by working only with the a_p. Notice that in the Euler product we have the usual Euler product term $\left(1 - \frac{a_p}{p^s} + \frac{p}{p^{2s}}\right)^{-1}$, except for the exceptional prime 2.

Earlier in the course we proved that for the elliptic curve $E: \ y^2 = x^3 - x$ if we let $\#E(\mathbb{F}_p) = p + 1 - a_p$ then $a_p = 0$ if $p \equiv 3 \pmod 4$ and, if $p \equiv 1 \pmod 4$, we write $p = a^2 + b^2$ where a is odd and then

$$a_p = 2(-1)^{\frac{a+b-1}{2}} a.$$

That is we get exactly the same values as above. Notice also that there is one exceptional prime, 2, which is the only prime that divides the discriminant of $y^2 = x^3 - x$ (and the discriminant of the quadratic form $x^2 + y^2$). Could this all be a wild co-incidence?

11. Degree three curves — why the curve $y^2 = x^3 + ax + b$?

11.1. Parametric families of rational points on curves. We have seen that if a linear curve has infinitely many rational points then they can be parametrized (that is, they are given by a formula which is a linear polynomial in some unknown t). Indeed starting from one given solution to $ax_0 + by_0 = c$ we then have the parametric family $x = x_0 + bt,\ y = y_0 - at$ for any integer t.

We saw that we can proceed similarly for quadratics, given one solution we can parametrize infinitely many others: To see this start by suppose that a, b, c are given non-zero integers for which $a + b + c = 0$. One can check that if r, s, t are integers for which $r + s + t = 0$ then

$$a(bs^2 + ct^2)^2 + b(as^2 + cr^2)^2 + c(at^2 + br^2)^2 = 0$$

This can be applied to show that any quadratic equation $Ax^2 + By^2 = Cz^2$ with one non-zero integral point has infinitely many given as polynomials in r, s, t. To do so simply select $a = Ax^2,\ b = By^2,\ c = -Cz^2$ above and we obtain $AX^2 + BY^2 = CZ^2$ with

$$X = x(B(ys)^2 - C(zt)^2), \quad Y = y(A(xs)^2 - C(zr)^2), \quad Z = z(A(xt)^2 + B(yr)^2).$$

For example, for the equation $x^2 + y^2 = 2z^2$ starting from the solution $(1, 1, 1)$ we obtain the parametrization $X = s^2 - 2t^2,\ Y = s^2 - 2r^2,\ Z = t^2 + r^2$, which we can re-write, taking $t = -s - r$, as $-X = 2r^2 + 4rs + s^2,\ Y = s^2 - 2r^2,\ Z = 2r^2 + 2rs + s^2$. Hence we see that any quadratic with one solution has infinitely many given parametrically.

How about cubics or higher degree? We will show that there are no coprime, non-constant polynomials $x(t), y(t), z(t)$ for which $ax(t)^3 + by(t)^3 + cz(t)^3 = 0$, as a corollary of a much more general result:

The *abc* Theorem for Polynomials. *If $a(t), b(t), c(t) \in \mathbb{C}[t]$ do not have any common roots and provide a genuine polynomial solution to $a(t) + b(t) = c(t)$, then the maximum of the degrees of $a(t), b(t), c(t)$ is less than the number of distinct roots of $a(t)b(t)c(t) = 0$.*

This is a "best possible" result in that we can find infinitely many examples where there is exactly one more zero of $a(t)b(t)c(t) = 0$ than the largest of the degrees. For example the familiar identity $(2t)^2 + (t^2 - 1)^2 = (t^2 + 1)^2$; or the rather less interesting $t^n + 1 = (t^n + 1)$.

Corollary 11.1. *There do not exist coprime, non-constant polynomials $x(t), y(t), z(t)$ for which $ax(t)^3 + by(t)^3 + cz(t)^3 = 0$.*

Proof. Suppose that $ax(t)^3 + by(t)^3 + cz(t)^3 = 0$ with $a, b, c \in \mathbb{C}$ where $x(t), y(t), z(t)$ have no common roots and let $a(t) = ax(t)^3,\ b(t) = by(t)^3,\ c(t) = cz(t)^3$. Let d be the maximum of the degrees of $x(t), y(t), z(t)$. By the *abc* Theorem for Polynomials we know that $3d$ is less than the number of distinct roots of $x(t)y(t)z(t) = 0$, which is $\leq 3d$, and hence we have a contradiction.

Proof of the abc Theorem for Polynomials. We differentiate $a + b = c$ to obtain $a' + b' = c'$. Now, define

$$\Delta(t) := \begin{vmatrix} a(t) & b(t) \\ a'(t) & b'(t) \end{vmatrix}.$$

Note that $\Delta(t) \neq 0$, else $ab' - a'b = 0$ so that $a'/a = b'/b$ and integrating implies that b is a scalar multiple of a, which contradicts hypothesis.

Let α be a root of $a(t)$, and suppose $(t-\alpha)^e$ is the highest power of $(t-\alpha)$ which divides $a(t)$. Evidently $(t-\alpha)^{e-1}$ is the highest power of $(t-\alpha)$ which divides $a'(t)$, and thus it is the highest power of $(t-\alpha)$ which divides $\Delta(t) = a(t)b'(t) - a'(t)b(t)$, since α is not a root of $b(t)$. Therefore $(t-\alpha)^e$ divides $\Delta(t)(t-\alpha)$. Multiplying all such $(t-\alpha)^e$ together we obtain

$$a(t) \text{ divides } \Delta(t) \prod_{a(\alpha)=0} (t-\alpha).$$

An analogous argument works for $b(t)$, and for $c(t)$ since by adding the first column to the second in the definition of $\Delta(t)$, we obtain

$$\Delta(t) = \begin{vmatrix} a(t) & c(t) \\ a'(t) & c'(t) \end{vmatrix}.$$

Hence, multiplying these together we obtain that

$$a(t)b(t)c(t) \quad \text{divides} \quad \Delta(t) \prod_{(abc)(\alpha)=0} (t-\alpha).$$

The result follows by taking the degrees of both sides and noting that, by definition,

$$\deg(\Delta) \leq \min\{\deg(a) + \deg(b),\ \deg(a) + \deg(c),\ \deg(c) + \deg(b)\} - 1.$$

11.2. Constructing new rational points on cubic curves from old ones.

Let us suppose that $f(x,y) \in \mathbb{Q}[x,y]$ of degree 3.[16] We wish to study rational points on the curve

$$f(x,y) = 0.$$

If $P = (u,v)$ is a rational point on $f(x,y) = 0$ we mean that $f(u,v) = 0$ and we sometimes write $f(P) = 0$. We also let $x(P) := u$ and $y(P) := v$, the x- and y-coordinates of P. We suppose that we are given two rational points P_1, P_2 on $f(x,y) = 0$. The equation of the line between them has rational coefficients, say $y = \lambda x + \tau$.

Now if any point is on both the line and the curve then its coordinates satisfy

$$f(x,y) = 0 \text{ and } y = \lambda x + \tau,$$

so that $F(x) := f(x, \lambda x + \tau) = 0$. This is typically a polynomial of degree ≤ 3 with rational coefficients, in which case we can write $F(x) = ax^3 + bx^2 + cx + d$ where $a, b, c, d \in \mathbb{Q}$. We know that this has roots $x_1 = x(P_1)$ and $x_2 = x(P_2)$. Moreover the sum of the roots of a polynomial F of degree 3 is given by $-b/a$ and so the third root of F equals $-b/a - x_1 - x_2$,

[16] That is, f is a polynomial of degree at most 3 in x and y (that is each term in the polynomial takes the form $a_{i,j}x^iy^j$ with $i+j \leq 3$ and $= 3$ for some non-zero $a_{i,j}$, and the coefficients are rational numbers.

which we will call x_3, another rational number. But then the point $P_3 = (x_3, y_3)$, where $y_3 = \lambda x_3 + \tau$, is also on the curve $f(x, y) = 0$ and the line $y = \lambda x + \tau$.

This is a beautiful way to construct new rational points on $f(x, y) = 0$ from given ones, and if we start with three non-collinear rational points on $f(x, y) = 0$ then we can typically generate infinitely many in this fashion. However it would be preferable to start this process with fewer points.

If we study the proof above then the key idea is that we need two rational points on a line with rational coefficients to generate a third point. The trick (spotted by Sir Isaac Newton) is to select the tangent line at a given rational point, in which case the tangent line has rational coefficients, and hence the third point at which the tangent line meets the curve is also a rational point. Let's study this more precisely: If the original rational point is at (u, v) make the change of variable $x \to x + u,\ y \to y + v$ so that we have $f(P) = 0$ with $P = (0, 0)$. Then

$$\frac{\partial f}{\partial y} \cdot \frac{dy}{dx} + \frac{\partial f}{\partial x} = \frac{df(x, y)}{dx} = 0$$

and so if $r = \frac{\partial f}{\partial y}\ (P)$ and $s = \frac{\partial f}{\partial x}\ (P)$ (which are rational numbers by substitution) then the tangent line at P has equation $ry + sx = 0$. The tangent line is well-defined except if $r = s = 0$.[17] We therefore have a third rational point of intersection (of the curve and the tangent line), which we denote by P_2. Using this point we create $P_3, P_4, \ldots$. With a bit of luck these will all be distinct and there will therefore be infinitely many rational points on the original cubic curve.

So if we are given a cubic curve with one rational point then we wish to study whether there are infinitely many rational points on the curve. It is difficult to attack this question directly because there are **10** coefficients of any arbitrary degree three polynomial in two variables. As with quadratics we seek to reduce that number, since there are many "models" for any given curve, which can be obtained through rational transformations of the co-ordinates. In the next section we will see, given that there is a rational point on the curve, we can reduce the number of variables in our general model to **two**.

11.3. Cubic curves into Weierstrass form. Suppose that we are given a curve $C : f(x, y) = 0$ of degree 3 with a rational point P, that is not a point of inflexion.[18] Our goal is to show that either the curve can be parametrized,[19] or the curve can be "re-written" as

$$y^2 = x^3 + ax + b, \tag{11.1}$$

with a rational point in the xy-plane, where a and b are integers with $\Delta := 4a^3 + 27b^2 \neq 0$, and there does not exist a prime p for which $p^4 | a$ and $p^6 | b$. By "re-written" we mean that there is, more-or-less, a 1-1 correspondence between rational points on both curves, and we can easily classify where there is not.

[17] In which case all of the monomials in f have degree 3. If we now change variables $x \to x/z,\ y \to 1/z$ and multiply through by z^3, then we have an equation that is linear in z, and hence easily solved in rationals.

[18] P is a point of inflexion for C if the tangent at P only intersects C at P; that is there is no other point of intersection.

[19] That is we can find a solution for x, y as functions of single variable.

Now draw the tangent to C at the point $P_1 := P = (x_1, y_1)$ and find a second rational point $P_2 = (x_2, y_2)$ on C. These points are distinct, since P is not a point of inflexion. We make the change of co-ordinates $x \to (x_2 - x_1)x + x_1,\ y \to (y_1 - y_2)y + y_2$ to get a new equation for the curve with rational points at $P_1 = (0, 1),\ P_2 = (1, 0)$. If we now write the curve in homogenous form and map $z \to z + x + y$ then we have a curve $z^3 + f_1(x, y)z^2 + f_2(x, y)z + f_3(x, y) = 0$ (with each f_j homogenous of degree j) and rational points at $P_1 = (0, 1, 0)$ and $P_2 = (1, 0, 0)$ so that $f_3(1, 0) = f_3(0, 1) = 0$. Moreover the tangent at P_1 is the line between the points, namely $z = 0$. We can also do a calculation to find the slope of the tangent at P_1 in, say, the (x, z)-plane: we differentiate to obtain

$$(3z^2 + 2f_1(x, 1)z + f_2(x, 1))\ \frac{dz}{dx} + z^3 + f_1'(x, 1)z^2 + f_2'(x, 1)z + f_3'(x, 1) = 0$$

We know that at P_1 we have $x = z = \frac{dz}{dx} = 0$ and hence $f_3'(0, 1) = 0$. Therefore $f_3(x, y) = gx^2y$ for some rational number g. We may take $g = 1$ else if $g = 0$ we have a quadratic in x, y, and if $g \neq 0$ we divide the whole equation through by g. Therefore our curve is of the form

$$x^2y + ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Now let $y \to y - a$ to obtain the form

$$x^2y + bxy + cy^2 + dx + ey + f = 0.$$

Next multiply through by y and let $X = xy$ to obtain

$$X^2 + bXy + cy^3 + dX + ey^2 + fy = 0.$$

Mapping $X \to y$ and $y \to x$ this is an equation of the form

$$ax^3 + by^2 + cxy + dx^2 + ex + fy + h = 0 \tag{11.2}$$

with $h = 0$. That is there is a rational point at $(0, 0)$.

We now show that the general equation (11.2) may be reduced to the form (11.1) or something simpler. We may assume that $b \neq 0$ else (11.2) is linear in y. If $b \neq 0$ then $by^2 + cxy + fy = bY^2 - b((c/2b)x + f/2b)^2$ where $Y = y + (c/2b)x + f/2b$ so we reduce to an equation of the form

$$Y^2 = ax^3 + bx^2 + cx + k;$$

and if $h = 0$ the rational point is now at $(0, f/2b)$. Letting $X = x + b/3a$ to reduce to the form $Y^2 = aX^3 + cX + d$; if $h = 0$ the rational point is now at $(-b/3a, f/2b)$. Now we multiply through by a^2 and let $v = aY, u = ax$ to obtain $v^2 = u^3 + Cu + D$. Now we multiply through by q^6 and make the change of variable $v = q^3y, u = q^2x$ to obtain $y^2 = x^3 + q^4Cx + q^6D$. We select q to be that rational number for which q^4C and q^6D are integers, and minimally so, and we obtain (11.1). We select q so that these integers are minimal in order that when we reduce (11.1) mod p it is non-degenerate if possible – note

that it is degenerate, that is two roots of the cubic are congruent mod p, if and only if p divides the discriminant.[20]

Moreover if we began with a rational point then we had $h = 0$ and the rational point $(0, 0)$ satisfying (11.2); and now the rational point $(-bq^2/3, aq^3f/2b)$ satisfying (11.1).

We have just seen that we may change co-ordinates on (11.1) for which $a \to q^4a,\ b \to q^6b$ for some rational q, and there is a 1-to-1 correspondence of rational points. We wish to determine invariants of the elliptic curve, that is quantities that do not change under such changes of variable. We see that the ratio $a^3 : b^2$ is fixed under such a transformation, and hence any rational function of that ratio. For reasons that will become more apparent later, we choose the following (linear) transformation of that ratio for our invariant: The *j-invariant* of the elliptic curve $E :\ y^2 = x^3 + ax + b$ over $\mathbb{C}$, is defined as

$$j(E) = 1728 \cdot \frac{a^3}{4a^3 + 27b^2}\ .$$

`Exercise 11.3.1.` Prove that two elliptic curves with the same j-invariant are isomorphic over $\mathbb{C}$, and in fact, at most a degree 2 field extension of the field defined by the coefficients of the two elliptic curves.

Now $4a^3(x^3+ax+b) = (2ax+3b)^2(ax-3b) + \Delta(ax+b) = 0$, where $\Delta := 4a^3 + 27b^2$. Therefore if $\Delta = 0$, then either our curve is $y^2 = x^3$, or $y^2 = (x + 3b/2a)^2(x - 3b/a) = X^2(X - c)$ for $c = 9b/2a$. In the first case the solutions are parametrized by (t^2, t^3); in the second let $z = y/X$ to obtain the curve $z^2 = X - c$ so our parametrization is $(z^2 + c, z(z^2 + c))$.

`Exercise 11.3.2.` Assume that there is a point of inflexion of C and that the curve cannot be parametrized. Follow a similar procedure to above, to first move the point to $(0, 0)$, then make the tangent line $y = 0$, and then show that the equation of the curve becomes $ey^3 + (ax + b)y^2 + (cx + d)y + x^3 = 0$, for some rationals a, b, c, d, e, after a simple change of variable.

`Exercise 11.3.3.` Show that $x^3 = y(y - 1)(y - 2)$ has three inflexion points, at $(0, 0), (0, 1)$ and $(0, 2)$.

`Exercise 11.3.4.` Prove that the point at infinity on (11.1) is an inflexion point.

11.4. Diagonal cubic curves. If one has a rational point (x, y, z) on the diagonal cubic

$$ax^3 + by^3 + cz^3 = 0, \tag{11.3}$$

then one obtains a rational point on another diagonal cubic:

$$X^3 + Y^3 + dZ^3 = 0. \tag{11.4}$$

[20] We seek the *minimal (integer) discriminant* over all isomorphic models of the elliptic curve. Thence p divides this discriminant if and only if the curve is isomorphic mod p to some curve of lower degree (as in section 9.6). Our techniques here work for all primes $p > 3$. However for $p = 2$ and 3 one has to take into account different possible models of the elliptic curve (because, for example, we cannot "complete the square" when working mod 2, to remove certain monomials).

where $X = C^3+6AC^2+3A^2C-A^3,\ Y = A^3+6CA^2+3C^2A-C^3,\ Z = 3xyz(A^2+AC+C^2)$, with $A = ax^3, B = by^3, C = cz^3$ and $d = abc$.[21] Therefore we can focus on the case where $a = b = 1,\ c = d$ in (11.3). Note that this is not an invertible map but certainly if there are no solutions to $X^3 + Y^3 + abcZ^3 = 0$ then there are none to $ax^3 + by^3 + cz^3 = 0$.

Now, given a solution to (11.4) let $X = u+v, Y = u-v$ to obtain $2u^3+6uv^2+dZ^3 = 0$. De-homogenize by dividing through by u^3 and letting $z = -Z/u,\ V = v/u$ to obtain $2 + 6V^2 = dz^3$, and obtain (11.2), and hence (11.1), specifically

$$y^2 = x^3 - 432d^2 \tag{11.5}$$

with $y = 36dV,\ x = 6dz$.

11.5. y^2 equals a quartic with a rational point. We have already seen that solutions to quartics are "equivalent" to solutions to cubics: We saw that

If $x^4 + y^4 = z^2$ then $v^2 = u^3 + u$ with $(u, v) = (x^2/y^2, xz/y^3)$.

In the other direction one has that

If $v^2 = u^3 + u$ then $x^4 + y^4 = z^2$ with $(x, y, z) = (u^2 - 1, 2v, u^4 + 6u^2 + 1)$.

One can generalize this further:

If $x^4 + bx^2y^2 + cy^4 = z^2$ then $v^2 = u^3 + bu^2 + cu$ with $(u, v) = (x^2/y^2, xz/y^3)$, and in other direction we take $(x, y, z) = (u^2 - c, 2v, u^4 + 2bu^3 + 6cu^2 + 2bcu + c^2)$.

In general we will suppose that we have any curve of the form

$$y^2 = f(x) \ \text{ where } f \text{ has degree 4, with a rational point } (k, a), \tag{11.6}$$

and try to covert it into the form (11.1). If we make a change of variable $x \to x - k$, that is letting $g(x) = f(x + k)$ then the rational point on $y^2 = g(x)$ where g has degree 4 is at $(0, a)$, and so $g(0) = a^2$, that is $g(x) = ex^4 + dx^3 + cx^2 + bx + a^2$. Now divide $y^2 = g(x)$ through by $1/x^4$ letting $v = y/x^2,\ u = 1/x$ (which has inverse $x = 1/u,\ y = v/u^2$), so that

$$v^2 = a^2u^4 + bu^3 + cu^2 + du + e. \tag{11.7}$$

If $a = 0$ we have a cubic and have reduced to (11.2). If $a \neq 0$ then write $h(u) = au^2 + (b/2a)u + (c/2a - b^2/4a^3)$ (which we write as $au^2 + Bu + C$) so that $h(u)^2 = a^2u^4 + bu^3 + cu^2 + \dots$ Therefore changing variable $v = w + h(u)$ we obtain the equation

$$w^2 + 2w(au^2 + Bu + C) = \ell u + m.$$

Now we multiply through by w and let $t = uw$ to obtain

$$w^3 + 2at^2 + 2Btw + 2Cw^2 = \ell t + mw,$$

and we have reduced to (11.2).

---

[21] The transformation from (11.3) to (11.4) can be obtained from the identity

$$(1 - 6t + 3t^2 + t^3)^3 - (1 + 3t - 6t^2 + t^3)^3 + t(t - 1)(3(t^2 - t + 1))^3 = 0;$$

by substituting in $t = -A/C$, and multiplying through by C^9.

11.6. The intersection of two quadratic polynomials in three variables (that is, the intersection of two quadratic surfaces). Here we have two quadratic polynomials f, g in three variables, with no common factors, and for which

$$f(x,y,z) = g(x,y,z) = 0 \text{ has the common rational point } (x_0, y_0, z_0). \tag{11.8}$$

By change of variable $x \to x + x_0, y \to y + y_0, z \to z + z_0$ we may assume that the rational point is at $(0,0,0)$. We write $f = f_2 + f_1 + f_0$ where f_j is the sum of the monomials of f of degree j. Then $f_0 = f(0,0,0) = 0$. Hence we are studying the solutions to $f_2 + f_1 = g_2 + g_1 = 0$. Therefore $f_1 g_2 - g_1 f_2 = 0$ is a homogenous cubic on which the rational points lie; de-homogenizing yields a cubic in two variables, and thus we are in the situation above which we know can be reduced to (11.1), unless $f_1 g_2 - g_1 f_2$ is identically zero. But then either $f_1 = cg_1$ and $f_2 = cg_2$ so that $f = cg$; or $f_1 | f_2$ and $g_1 | g_2$, so that $f = f_1 h$ and $g = g_1 h$ where h is a linear polynomial. Either way f and g have a common factor contradicting the hypothesis.

11.7. Doubling a point on a diagonal cubic. Doubling a point on a diagonal cubic involves several transformations, but it works out more elegantly to give the formula directly: If we have a solution to (11.3) then we take $t = -A/C$ in the identity,

$$t(t-2)^3 + (1-t)(1+t)^3 + (2t-1)^3 = 0,$$

multiply through by C^9, where $A = ax^3, B = by^3, C = cz^3$ and take $X = x(B-C),\ Y = y(C-A),\ Z = z(A-B)$ to obtain another solution to (11.3):

$$aX^3 + bY^3 + cZ^3 = A(B-C)^3 + B(C-A)^3 + C(A-B)^3 = 0.$$

12. The rational points on an elliptic curve

12.1. The group of rational points on an elliptic curve. We have just seen how a cubic with a rational point can be transformed to an equation of the affine form

$$E:\ y^2 = x^3 + ax + b$$

with $a, b \in \mathbb{Z}$ by linear maps with rational coefficients. This is called an *elliptic curve*. We have seen that two rational points on the unit circle gave rise to a line with rational coefficients and vice-versa; this allowed us to find all the rational points on the unit circle. We extend that idea to elliptic curves. Let $E(\mathbb{Q})$ denote all of the rational points on E (that is (x, y) on E with $x, y \in \mathbb{Q}$).

`Exercise 12.1.1.` Show that if $(x, y) \in E(\mathbb{Q})$ then there exist integers ℓ, m, n such that $x = m/n^2,\ y = \ell/n^3$ with $(\ell m, n) = 1$. Moreover the point at ∞ corresponds to $n = 0,\ \ell = m = 1$.

`Exercise 12.1.2.` Let $\Delta = 4a^3 + 27b^2$. Show that if $a > 0$ or if $\Delta > 0$ then $x^3 + ax + b = 0$ has just one real root. Show that if $a, \Delta < 0$ then $x^3 + ax + b = 0$ has three real roots. Sketch the shape of the curve $y^2 = x^3 + ax + b$ in the two cases.

Suppose that we are given two points $(x_1, y_1), (x_2, y_2) \in E(\mathbb{Q})$. The line between them, $y = mx + \nu$ has $m, \nu \in \mathbb{Q}$.[22] These two points are both intersections of the line $y = mx + \nu$ with the elliptic curve $y^2 = x^3 + ax + b$, that is x_1, x_2 satisfy

$$(mx + \nu)^2 = y^2 = x^3 + ax + b;$$

in other words x_1 and x_2 are two of the three roots of the cubic polynomial

$$x^3 - m^2x^2 + (a - 2m\nu)x + (b - \nu^2) = 0.$$

If the third root is x_3 then $x_3 = m^2 - x_1 - x_2 \in \mathbb{Q}$ and if we let $y_3 = mx_3 + \nu$ we obtain the third intersection of the line with E, and $(x_3, y_3) \in E(\mathbb{Q})$. This method of generating a third rational point from two given ones goes back to Fermat.

Actually one can do even better and generate a second point from a given one: If $(x_1, y_1) \in E(\mathbb{Q})$ let $y = mx + \nu$ be the equation of the tangent line to $y^2 = x^3 + ax + b$ at (x_1, y_1). To calculate this simply differentiate to obtain $2y_1m = 3x_1^2 + a$ and then $\nu = y_1 - mx_1$. Now our cubic polynomial has a double root at $x = x_1$ and we again compute a third point by taking $x_3 = m^2 - 2x_1,\ y_3 = mx_3 + \nu$ so that $(x_3, y_3) \in E(\mathbb{Q})$.

In these constructions we missed the case when the line is vertical (in the first case $x_1 = x_2$ which implies that $y_2 = -y_1$; in the second case $y = 0$). Where is the third point of intersection? One cannot see another point of intersection on the graph (that is on the real plane), but the line stretches to infinity, and indeed the third point is, rather surprisingly, the point at infinity, which we denote 0. Remember from section C11, in projective co-ordinates the elliptic curve is $y^2z = x^3 + axz^2 + bz^3$ so the point at infinity is $(0, 1, 0)$.

[22] Or is of the form $x = x_1 = x_2$, a situation we will deal with a little later.

Exercise 12.1.3. Prove that there cannot be four points of $E(\mathbb{Q})$ on the same line.

Poincaré made an extraordinary observation: If we take any three points P, Q, R of E on the same line then we can define a group by taking $P + Q + R = 0$. The line at infinity tells us that the point at infinity is indeed the 0 of this group. Moreover we have seen that $(x, y) + (x, -y) = 0$. Note that this implies that, in the notation above, $(x_3, -y_3) = (x_1, y_1) + (x_2, y_2)$

It is clear that the operation is closed under addition (and, most interestingly, closed in the subgroup $E(\mathbb{Q})$). The one thing that is complicated to justify is that Poincaré's operation (of addition) is indeed associative, and that hence we do indeed have a group.

Exercise 12.1.4. Show that the addition law given here is indeed associative. There are several proofs of this fact. One method is to prove it purely geometrically, which is not easy (such proofs can be found in most textbooks). Another proof involves find formulae for the co-ordinates of $(P+Q)+R$ and $P+(Q+R)$, much as in 12.1.6, and verify that these are the same. This is tedious to implement, even using symbolic algebra software. We will give a third, analytic, proof, later in section *.*. The advantage is that this proof is quite elegant and follows from a body of essential theory.

It is also obvious that the addition law is commutative. The question then becomes to identify the structure of the group of rational points, $E(\mathbb{Q})$.

Is $E(\mathbb{Q})$ finite or infinite? Suppose that we have a rational point P. Take the tangent, find the third point of intersection of the tangent line with E to obtain $-2P$, and then reflect in the x-axis to obtain $2P$. Fermat suggested that if we repeat this process over and over again, then we are unlikely to come back again to the same point. If we never return to the same point then we say that P has *infinite order*; otherwise P has finite order, the order being the minimum positive integer n for which $nP = 0$ (points of finite order are known as *torsion points*).

Exercise 12.1.5. Prove that the torsion points form a subgroup.

Exercise 12.1.6. Prove that if $P = (x, y)$ with $y \neq 0$ then $2P = (X, Y)$ where

$$X = \frac{(x^2 - a)^2 - 8bx}{4y^2}, \quad Y = \frac{x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3}{8y^3}.$$

Notice that the numerator and denominator of X are polynomials of degree four in x implying (roughly) that the co-ordinates of $2P$ are about four times the length of the co-ordinates of P, unless there is an enormous amount of cancelation between numerator and denominator. To express this better it is convenient to define the *height* of P, $H(P) := \max\{|m|, n^2\}$. Our observation is that $H(2P) \approx H(P)^4$.

Exercise 12.1.7. Show that (x, y) has order 2 if and only if $y = 0$. Deduce that the number of points of order 1 or 2 is one plus the number of integer roots of $x^3 + ax + b$; and therefore equals $1, 2$ or 4.

12.2. The group law on the circle, as an elliptic curve. The points on the circle $x^2 + y^2 = 1$ may be parametrized by $P_\theta := (\cos\theta, \sin\theta)$, where θ is taken mod 2π. Addition of points may be defined by $P_\theta + P_\phi = P_{\theta+\phi}$, which yields an abelian group with $P_0 = (1, 0)$ as the identity. To show that this is a special case of the addition laws above we begin by homogenizing to make this into the cubic curve

$$C: \ x^2z + y^2z = z^3,$$

which has the affine points $x^2+y^2=1$, and all possible points $(x,y,0)$ at infinity. Any line $ax+by=c$ meets the circle $x^2+y^2=1$ in two (complex) points (including multiplicities), and also meets C at infinity at the point $(-b,a,0)$. (In other words these are the projective points on the intersection of $ax+by=cz$ and $x^2+y^2=z^2$.) The line joining this to $(1,0,1)$ is parallel to the original line and has equation $ax+by=az$; and so meets $x^2+y^2=z^2$ again at $(a^2-b^2,2ab,a^2+b^2)$.

We need to figure out how to add two points on this curve:

Doubling a point $P_\theta := (\cos\theta,\sin\theta)$ on $x^2+y^2=1$, we note that $2y\,\frac{dy}{dx}+2x=0$ so that the slope of the tangent line $t_\theta := \frac{dy}{dx} = -\frac{x}{y} = -\frac{\cos\theta}{\sin\theta}$. As in the paragraph above, this line meets C again at $(\cos\theta,-\sin\theta,0) = -2P_\theta$; and then drawing the line from $-2P_\theta$ to P_0 the third point of intersection is

$$2P_\theta = (\cos^2\theta-\sin^2\theta, 2\sin\theta\cos\theta, \sin^2\theta+\cos^2\theta) = (\cos 2\theta,\sin 2\theta,1) = P_{2\theta},$$

as claimed.

Adding two distinct points: The line between P_θ and P_ϕ has slope $\frac{\sin\theta-\sin\phi}{\cos\theta-\cos\phi} = -\frac{\cos(\frac{\theta+\phi}{2})}{\sin(\frac{\theta+\phi}{2})}$; so proceeding as in the paragraph this meets C again at

$$-P_\theta - P_\phi = \left(\cos(\frac{\theta+\phi}{2}), -\sin(\frac{\theta+\phi}{2}), 0\right) = -2P_{\frac{\theta+\phi}{2}} = -P_{\theta+\phi}$$

and the result follows.

Since any conic can be transformed into the circle via line transforming transformations, the same generalization of the elliptic curve group law holds true on *any* conic.

12.3. No non-trivial rational points by descent. There are some elliptic curves for which we can show that there are no rational points by an easy descent: Suppose that $x,y\in\mathbb{Q}$ such that $y^2=x^3+x$, so that $x=m/n^2, y=\ell/n^3$ with $(\ell m,n)=1$ and

$$\ell^2 = m(m^2+n^4).$$

Now $(m,m^2+n^4)=(m,n^4)=1$ so that both m and m^2+n^4 are squares, say $m=u^2$ and $m^2+n^4=w^2$. Therefore

$$n^4+u^4 = m^2+n^4 = w^2,$$

and we showed, in section 6.4, that this has no non-trivial solutions. The trivial solutions have either $n=0$ (corresponding to the point $\mathcal{O}$ at ∞), or $u=0$ (corresponding to the point $(0,0)$ of order two). Hence $E(\mathbb{Q})\cong\mathbb{Z}/2\mathbb{Z}$.

12.4. The group of rational points of $y^2=x^3-x$. We will begin by applying the ideas from elementary number theory to rational points on the elliptic curve

$$E:\ y^2 = x^3-x.$$

We see that there are several rational points: the point at infinity, as well as the three points of order two, $(-1,0),(0,0),(1,0)$. Are there any others? By exercise 12.1.1 we can write $x = m/n^2,\ y = \ell/n^3$ with $(\ell m, n) = 1$ to obtain

$$(m-n^2)m(m+n^2) = \ell^2.$$

Here we have that the product of three integers equals a square, so we can write each of them as a squarefree integer times a square, where the product of the three squarefree integers must also equal a square. Note that the three squarefree integers cannot have a common factor, else that factor cubed is a square and so the integers were not squarefree. This means that we can write

$$m - n^2 = pqu^2, \quad m = prv^2, \quad m + n^2 = qrw^2,$$

for some squarefree integers p, q, r which are pairwise coprime. Moreover since the product of the integers is positive, and as $m-n^2 < m < m+n^2$ we see that $m+n^2 > 0$ and $m-n^2$ and m have the same sign. Hence we may assume that p has the same sign as m, and q and r are positive. Now note that $(m \pm n^2, m) = (n^2, m) = 1$ and so $|p| = r = 1$. Moreover $(m-n^2, m+n^2) = (m-n^2, 2n^2) = (m-n^2, 2)$ since $(m-n^2, n^2) = (m, n^2) = 1$. To summarize, we have proved that there are four possibilities for the value of (p,q,r), since $p = -1$ or 1 and $q = 1$ or 2. That leads to four sets of equations:

$$\begin{array}{rcl|rcl|rcl|rcl} m-n^2 &=& u^2 & m-n^2 &=& -u^2 & m-n^2 &=& 2u^2 & m-n^2 &=& -2u^2 \\ m &=& v^2 & m &=& -v^2 & m &=& v^2 & m &=& -v^2 \\ m+n^2 &=& w^2 & m+n^2 &=& w^2 & m+n^2 &=& 2w^2 & m+n^2 &=& 2w^2 \end{array}$$

Note that these four cases do each have a solution with $(m,n) = (1,0),(0,1),(1,1),(-1,1)$, respectively, which correspond to the four rational points $\mathcal{O},(0,0),(1,0),(-1,0)$, respectively, that we already know.

Let us suppose that we have another rational point on the curve $P = (m/n^2, \ell/n^3)$, so that we have four new rational points, $P = P+\mathcal{O}, P+(0,0), P+(1,0)$ and $P+(-1,0)$. We now show that if P gives rise to a solution to the second set of quadratic equations, then $P+(0,0)$ gives rise to a solution to the first set of quadratic equations (and, analogously, if P gives rise to a solution to the third or fourth set of quadratic equations, then $P+(1,0)$ or $P+(-1,0)$, respectively gives rise to a solution to the first set of quadratic equations):

The line between P and $(0,0)$ is $mny = \ell x$. If $P+(0,0) = (u,v)$ then $u + m/n^2 + 0 = (\ell/mn)^2$, so that $u = (\ell^2 - m^3)/m^2n^2 = -n^2/m$. since $\ell^2 = m^3 - mn^4$. Hence we see that m and n^2 in the first equations must be replaced by n^2 and $-m$ respectively, so that the three equations yield $n^2 + m = w^2,\ n^2,\ n^2 - m = u^2$, respectively, as claimed.

`Exercise 12.4.1.` Show that if P gives rise to a solution to the third or fourth set of quadratic equations, then $P+(1,0)$ or $P+(-1,0)$, respectively gives rise to a solution to the first set of quadratic equations.

Hence we have now proved that for any $P \in E(\mathbb{Q})$, at least one of $P, P+(0,0), P+(1,0)$ and $P+(-1,0)$ gives rise to a solution in coprime integers (m,n) to $m-n^2 = u^2,\ m = v^2,\ m+n^2 = w^2$. Substituting in $m = v^2$ yields that $v^2 - n^2 = u^2,\ v^2 + n^2 = w^2$, and multiplying these together yields a solution to

$$X^4 - Y^4 = z^2 \ \text{ with } (X,Y) = 1 \ \text{ and } X+Y \text{ odd}.$$

(Our solution is (v, n, uw); since $v^2 = m$ we have $(v, n) = 1$; since $v^2 + n^2 = w^2$ with $(v, n) = 1$ we know that they have opposite parity.) We will show that this has no non-trivial solution, by a Fermat-type descent argument. If there are solutions then we know from the Pythagorean equation $z^2 + (Y^2)^2 = (X^2)^2$ in coprime integers that there exist coprime integers r, s of different parity such that $X^2 = r^2 + s^2$ and $Y^2 = 2rs$ (as X, Y are of different parity). Now $(r, X)|(Y^2, X) = 1$, and so there exist coprime positive integers p, q of different parity such that $r = 2pq,\ s = p^2 - q^2$. Therefore $4pq(p^2 - q^2) = 2rs = Y^2$. Since $p, q, p^2 - q^2$ are positive and coprime we obtain that they are each squares, say x^2, y^2, z^2 and so $x^4 - y^4 = z^2$.

`Exercise 12.4.2.` Show that x, y, z is a "smaller" solution that X, Y, Z and therefore complete the proof. The challenge here is to define the size of a solution in such a way that the argument here works smoothly.

From all this we deduce that $E(\mathbb{Q}) = \{\mathcal{O}, (0, 0), (1, 0), (-1, 0)\}$.

Our goal in the next few sections is to generalize this proof to more elliptic curves. Let's rephrase the argument in a way that makes sense to generalize: We will suppose that $E:\ y^2 = x^3 + ax + b = (x - r_1)(x - r_2)(x - r_3)$ where, for now, r_1, r_2, r_3 are integers, though eventually we will have to consider the case where $x^3 + ax + b$ is irreducible.

`Step One`: Define a map $\phi : E(\mathbb{Q}) \to \mathbb{Q}^3$ where $\phi(P) = (x - r_1, x - r_2, x - r_3)$. We know that the three co-ordinates are rational numbers whose product is a square, and we classified the three numbers by writing them as a squarefree integer times a square. Now the three squarefree numbers have product that is a square, even in the case that one of $x - r_1, x - r_2, x - r_3$ equals 0. So we could instead define

$$\phi : E(\mathbb{Q}) \to \{(a, b, c) : a, b, c \in (\mathbb{Q}^*/(\mathbb{Q}^*)^2)^3 \text{ and } abc = 1\}.$$

`Step Two`: We observed that we can write $a = pq,\ b = rp,\ c = qr$ where p, q and r are squarefree divisors of $r_1 - r_2,\ r_1 - r_3$ and $r_3 - r_2$ respectively. Moreover if $r_1 > r_2 > r_3$ then $x - r_3 > 0$ and so we can assume that q and r are positive, and that p may be positive or negative. These conditions imply that

$$\phi(E(\mathbb{Q})) \subset \{(p, q, r) :\ p|r_1 - r_2,\ \ q|r_1 - r_3\ \ r|r_3 - r_2,\ \ \text{and } q, r \geq 1\}.$$

and so $\phi(E(\mathbb{Q}))$ is finite.

`Step Three`: For each element $v \in \phi(E(\mathbb{Q}))$ fix some $P_v \in E(\mathbb{Q})$ for which $\phi(P_v) = v$. Given any element $P \in E(\mathbb{Q})$, suppose that $v = \phi(P)$. The key new idea in our construction above, was that $\phi(P + P_v) = (1, 1, 1)$.

`Step Four`: Find a descent argument. That is, given a point $P \in \phi(E(\mathbb{Q}))$, for which $\phi(P) = (1, 1, 1)$, show that there is a "smaller" point $R \in \phi(E(\mathbb{Q}))$, for which $\phi(R) = (1, 1, 1)$.

This is a lot to ask for! Particularly since it cannot always end up in a contradiction as there are elliptic curves with infinitely many rational points. To pursue this argument we will need to better understand torsion points, the map ϕ, and to find a precise notion of the size of a solution.

12.5. The arithmetic of a torsion point. Elizabeth Lutz and Nagell showed that there are only finitely many torsion points in $E(\mathbb{Q})$. To be more precise they proved:

The Lutz-Nagell Theorem. *If $(x, y) \in E(\mathbb{Q})$ is a point of order $m > 1$ then*
(i) If $m = 2$ then $y = 0$ and x is an integer;
(ii) If $m > 2$ then x and y are integers, and y^2 divides $\Delta = 4a^3 + 27b^2$.
Therefore there are only finitely many torsion points in $E(\mathbb{Q})$.

The key idea in the proof is to show that if $x(P) = m/n^2$ with $(m, n) = 1$ and $n > 1$ then $x(kP) = M/N^2$ with $(M, N) = 1$ and N divisible by n. An immediate consequence of this is that if $P = (x, y)$ is a torsion point then x is an integer, and hence y is an integer.

Proof of the easy parts of the theorem. (i) We see that $2P = 0$ if and only if $-P = P$. So if $P = (x, y)$ then $y = -y$ so that $y = 0$ and hence x is an integer.

For the final part of the theorem, note that if $P = (x, y)$ is a torsion point, that is $mP = 0$ for some $m > 2$ then $m{\cdot}(2P) = 2{\cdot}nP = 0$ so $2P = (X, Y)$ is also a torsion point, and hence x, y, X, Y are all integers. Now $m^2 = 2x + X \in \mathbb{Z}$ so that $m = (3x^2 + a)/2y \in \mathbb{Z}$; that is $2y$ divides $3x^2 + a$. Hence y divides $9(3b - 2ax)(x^3 + ax + b) + (6ax^2 - 9bx + 4a^2)(3x^2 + a) = \Delta$. Since there are only finitely many divisors y of $4a^3 + 27b^2$, and each such y gives rise to at most three values of x, hence there are only finitely many torsion points in $E(\mathbb{Q})$. Actually we can better than this: Since $2P$ is also a torsion point, $x(2P) = ((x^2 - a)^2 - 8bx)/4y^2$ is an integer. In particular this implies that y^2 equals $x^3 + ax + b$ and divides $(x^2 - a)^2 - 8bx$ so that it must also divide

$$(3x^2 + 4a)((x^2 - a)^2 - 8bx) - (x^3 + ax + b)(3x^3 - 5ax - 27b) = 4a^3 + 27b^2. \tag{12.1}$$

`Exercise` 12.5.1. Prove that P has finite order if and only if there exists $i > j \geq 0$ such that $2^i P = \pm 2^j P$.

Mazur improved the Lutz-Nagell theorem showing that the torsion subgroup of $E(\mathbb{Q})$ contains at most 16 points. In fact this subgroup is either $\mathbb{Z}/N\mathbb{Z}$ for some $1 \leq N \leq 10$ or $N = 12$, or it is $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2N\mathbb{Z}$ for some $1 \leq N \leq 4$. After Mazur's Theorem it is not difficult to identify torsion points. First determine all integers y for which y^2 divides $4a^3 + 27b^2$, and then compute all such points $(x, y) \in E(\mathbb{Q})$.

`Exercise` 12.5.2. Show that $P \in E(\mathbb{Q})$ has finite order if and only if $x(P), x(2P), x(4P), x(8P), \infty$ are not distinct.

There can be torsion points in fields other than $\mathbb{Q}$. One can ask for them in $\mathbb{C}$; in fact their x-co-ordinates are roots of a polynomial with integer coefficients, so they are algebraic numbers. We will show later that the subgroup of torsion points of order N is isomorphic to $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$ for each $N \geq 1$, so that there are N^2 points of order dividing N.

`Exercise 12.5.3.` Prove that there are exactly $N^2 \prod_{p|N}(1 - \frac{1}{p^2})$ points of order N on $E(\mathbb{C})$.

We will need the following lifting lemma, which generalizes the result in exercise 8.3.3, explaining how the orders of points modulo different powers of p are related:

Lemma 12.0. *Suppose that $P = (x, y) \in E(\mathbb{Q})$ with $x = m/n^2,\ y = \ell/n^3$, where ℓ, m and $n = n(P)$ are integers for which $(\ell m, n) = 1$ (as in exercise 12.1.1). If p is a prime for which $p^k \| n(P)$ with $k \geq 1$ then $p^{k+\ell} \| n(rp^\ell P)$ for any $\ell \geq 0$ where $p \not| r$.*

Deduction of the Lutz-Nagell Theorem. Suppose that $P = (x, y)$ is a torsion point of E, but not $\mathcal{O}$, so that there exists an integer $s > 1$ for which $sP = 0$. If $x = m/n^2$ with $n \geq 1$ (as in exercise 12.1.1), and p is a prime for which $p^k \| n(P)$ with $k \geq 1$, then $p^{k+\ell} \| n(sP)$ where $s = rp^\ell$ and $p \not| r$, by Lemma 12.0. However this is incompatible with the fact that $n(sP) = 0$. Hence n cannot be divisible by any prime p, so that $n = 1$; that is x, and so y, are integers.

Proof of Lemma 12.0. To understand better the exact power of a prime p dividing n, it makes sense to study $x/y = mn/\ell$. Hence we make the change of variable $u = x/y,\ v = 1/y$ (which has inverse $x = u/v,\ y = 1/v$ so is 1-to-1), so we have the equation

$$v = y^2/y^3 = (x/y)^3 + a(x/y)(1/y) + b/y^3 = u^3 + auv^2 + bv^3.$$

Note that the map $x \to x,\ y \to -y$ can be realized as the map $u \to -u,\ v \to -v$, so that if $P = (u, v)$ then $-P = (-u, -v)$.

`Exercise` 12.5.4. Prove that if $n \equiv 0 \pmod q$ then $u \equiv 0 \pmod q$ and $v \equiv 0 \pmod{q^3}$.

Given two distinct points $P_i := (u_i, v_i),\ i = 1, 2$ with $n_i \equiv 0 \pmod q$ we determine $P_1 + P_2 = (u_3, v_3)$: To begin with we have

$$\begin{aligned} v_1 - v_2 &= u_1^3 - u_2^3 + au_1v_1^2 - au_2v_2^2 + bv_1^3 - bv_2^3 \\ &= (u_1^3 - u_2^3) + a(u_1 - u_2)v_1^2 + au_2(v_1^2 - v_2^2) + b(v_1^3 - v_2^3), \end{aligned}$$

so that the slope of the line between P_1 and P_2 is

$$\lambda := \frac{v_2 - v_1}{u_2 - u_1} = \frac{(u_1^2 + u_1u_2 + u_2^2) + av_1^2}{1 - au_2(v_1 + v_2) - b(v_1^2 + v_1v_2 + v_2^2)} \equiv 0 \pmod{q^2}.$$

`Exercise` 12.5.5. Argue (by continuity) that if $P_1 = P_2$ then $\lambda = (3u^2 + av^2)/(1 - 2auv - 2bv^2)$.

If the line between P_1 and P_2 has equation $v = \lambda u + \tau$ then $\tau = v_1 - \lambda u_1 \equiv 0 \pmod{q^3}$. If we substitute the equation of the line into the equation for the curve we obtain that

$$\lambda u + \tau = u^3 + au(\lambda u + \tau)^2 + b(\lambda u + \tau)^3$$

a cubic equation whose roots are u_1, u_2 and $-u_3$. Hence

$$u_1 + u_2 - u_3 = -\frac{2a\lambda\tau + 3b\lambda^2\tau}{1 + a\lambda^2 + b\lambda^3} \equiv 0 \pmod{q^5}.$$

With this tool in hand we proceed with an exercise:

`Exercise` 12.5.6. (i) Show that if $P \in E(\mathbb{Q})$ then $u(mP) \equiv mu(P) \pmod{n(P)^5}$ for all $m \geq 1$.
(ii) Deduce that if p is a prime and $p^k \| n(P)$ with $k \geq 1$ then $p^{k+\ell} \| n(mP)$ where $p^\ell \| m$. (Hint: Apply (i) for $P, pP, p^2P, \ldots$ in place of P.)

So now suppose that $P = (x, y)$ is a torsion point of E, so that there exists an integer m for which $mP = 0$. If $x = m/n^2$ and p is a prime for which $p^k \| n(P)$ with $k \geq 1$, then $p^{k+\ell} \| n(mP)$. However this is incompatible with the fact that $n(mP) = 0$. Hence p cannot divide n for every prime p, so that $n = 1$; that is x and y are integers.

Example: *Torsion on* $E_a: \ y^2 = x^3 + ax$ *where* $4 \not| a$: We see that E_a has the point $(0, 0)$ of order 2 and no other, unless $a = -b^2$ for some odd b, in which case there are also $(-b, 0)$ and $(b, 0)$. We shall now show that there are no other torsion points: Suppose that $P = (x, y) \in E(\mathbb{Q})$ has order > 2. By Nagell-Lutz, both x and y are integers and $y^2 | 4a^3$. Now $x(2P) = (x^2 - a)/2y$ and so $2|x^2 - a$. Therefore $y^2 = x^3 + ax = x(x^2 - a) + 2ax \equiv 0 \pmod 2$, and thus $2|y$ and hence $4|x^2 - a$. Therefore a is odd else $2|x$ which implies $4|a$, a contradiction. Therefore x is odd and $a \equiv x^2 \equiv 1 \pmod 4$, and so $0 \equiv y^2 = x^3 + ax \equiv x(x^2 + 1) \equiv 2 \pmod 4$, a contradiction.

`Exercise` 12.5.7. Using the Nagell-Lutz Theorem and Mazur's theorem, determine the torsion subgroup for each of the following curves:

i. $y^2 = x^3 - 2$, which has $\Delta = 2^2 3^3$
ii. $y^2 = x^3 + 8$, which has $\Delta = 2^6 3^3$
iii. $y^2 = x^3 + 4$ which has $\Delta = 2^4 3^3$
iv. $y^2 = x^3 + 4x$ which has $\Delta = 2^8$
v. $y^2 - y = x^3 - x^2$ (or $y^2 = x^3 - 432x + 8208$) which has $\Delta = 2^8 3^{12} 11$
vi. $y^2 = x^3 + 1$, which has $\Delta = 3^3$
vii. $y^2 = x^3 - 43x + 166$ which has $\Delta = 2^{15} 13$
viii. $y^2 + 7xy = x^3 + 16x$ (or $y^2 = (x - 147)(x^2 + 147x - 22482)$) which has $\Delta = -2^{16} 3^{16} 17$
ix. $y^2 + xy + y = x^3 - x^2 - 14x + 29$ (or $y^2 = x^3 - 219x + 1654$) which has $\Delta = 2^{17} 3^{15}$
x. $y^2 + xy = x^3 - 45x + 81$ (or $y^2 = (x - 75)(x^2 + 75x - 52722)$) which has $\Delta = -2^{18} 3^{17} 11$
xii. $y^2 + 43xy - 210y = x^3 - 210x^2$ (or $y^2 = (x - 3531)(x^2 + 3531x - 20871666)$) which has $\Delta = -2^{20} 3^{18} 5^3 7^4 13$
ii.ii. $y^2 = x(x - 2)(x + 2)$, which has $\Delta = -2^8$
ii.iv. $y^2 + xy - 5y = x^3 - 5x^2$ (or $y^2 = (x - 21)(x - 102)(x + 123)$) which has $\Delta = -2^8 3^{16} 5^4$
ii.vi. $y^2 + 5xy - 6y = x^3 - 3x^2$ (or $y^2 = (x - 66)(x - 111)(x + 177)$) which has $\Delta = -2^{10} 3^{18} 5^2$
ii.viii. $y^2 + 17xy - 120y = x^3 - 60x^2$ (or $y^2 = (x - 282)(x - 1011)(x + 1293)$) which has $\Delta = -2^{16} 3^{20} 5^4 7^2$

12.6. Embedding torsion in $\mathbb{F}_p$. We are interested in how the torsion subgroup $\mathrm{Tors}(E(\mathbb{Q}))$ "reduces" mod p. It is easiest to study those primes p for which $p \not| \Delta$ since then $x^3 + ax + b$ has no repeated roots mod p, and so the reduction map $\rho_p : E(\mathbb{Q}) \to E(\mathbb{Z}/p\mathbb{Z})$ is a homomorphism. Now if $P \neq Q \in E(\mathbb{Q})$ but $\rho_p(P) = \rho_p(Q)$, then if $R = P - Q \in E(\mathbb{Q}) \setminus \{\mathcal{O}\}$ we have $\rho_p(R) = 0$. But this implies that if $R = (m/n^2, \ell/n^3)$ with $(\ell m, n) = 1$ then $p|n$. However, the Lutz-Nagell theorem tells us that if $R \in \mathrm{Tors}(E(\mathbb{Q}))$ then $n = 1$ and so p cannot divide n. Hence we see that

$$\rho_p : \mathrm{Tors}(E(\mathbb{Q})) \to E(\mathbb{Z}/p\mathbb{Z}), \quad \text{is an injection.}$$

This implies, since $E(\mathbb{Z}/p\mathbb{Z})$ is $\cong E(\mathbb{F}_p)$:

Theorem 12.1. *If prime* $p \not| \Delta_E$ *then* $E(\mathbb{F}_p)$ *contains a subgroup which is isomorphic to* $\mathrm{Tors}(E(\mathbb{Q}))$.

This provides a useful way of bounding the possibilities for torsion, since we simply need to compute Δ, find the smallest prime (or primes) that does not divide Δ, determine the structure of $E(\mathbb{F}_p)$ and we know that this contains (as an abstract group), the $\mathrm{Tors}(E(\mathbb{Q}))$.

12.7. The growth of points. In exercise 12.1.6 we saw that if $P = (m/n^2, \ell/n^3)$ with $(\ell m, n) = 1$ then

$$x(2P) = \frac{(m^2 - an^4)^2 - 8bmn^6}{4n^2(m^3 + amn^4 + bn^6)} .$$

The numerator and denominator of $x(2P)$ are each polynomials of degree four in m and n^2, so we expect $2P$ to have co-ordinates about four times as long as those of P. We want to make this intuition precise. For a given rational number r/s with $(r, s) = 1$ we define its *height* to be $H(r/s) = \max\{|r|, |s|\}$. We extend this to points on E by taking $H(P) = H(x(P))$. Hence $H(P) = \max\{|m|, n^2\}$, and so $H(\mathcal{O}) = 1$. We believe that $H(2P)$ is around $H(P)^4$. The main difficulty is that to determine $H(2P)$, we must first divide the numerator and denominator of $x(2P) = R/S$ by (R, S). Now $(n, R) = (n, (m^2 - an^4)^2 - 8bmn^6) = (n, m^4) = 1$. Moreover $(m^3 + amn^4 + bn^6, R)$ divides

$$(3m^2 + 4an^4)R - (m^3 + amn^4 + bn^6)(3m^3 - 5amn^4 - 27bn^6) = \Delta n^{12} \tag{12.2}$$

by (12.1), and hence it divides Δ as $(R, n) = 1$. We deduce that (R, S) divides 4Δ,[23] and so $\max\{|R|, |S|\}/4\Delta \le H(2P) \le \max\{|R|, |S|\}$. Now as $|m|, n^2 \le H(P)$ we see that

$$|R| \le ((1 + |a|)^2 + 8|b|)H(P)^4 \quad \text{and} \quad |S| \le 4(1 + |a| + |b|)H(P)^4.$$

Therefore $H(2P) \ll H(E)^4 H(P)^4$, where $H(E) := \max\{|a|^3, b^2\}^{1/6}$.[24]

To get a good lower bound on $H(2P)$ we need to show that there cannot simultaneously be a lot of cancelation in both the numerator and denominator when we add the terms. If $|m| \ge 3H(E)n^2$ then $m^2 \ge 9|a|n^4$ and $|m|^3 \ge 27|b|n^6$ so that

$$|(m^2 - an^4)^2 - 8bmn^6| \ge (m^2 - m^2/9)^2 - 8m^4/27 = (40/81)m^4.$$

Otherwise $n^2 \ge H(P)/3H(E)$, and by (12.2), we have

$$\begin{aligned} 4\Delta n^{14} &\le 4|n^2(3m^2 + 4an^4)R| + |(3m^3 - 5amn^4 - 27bn^6)S| \\ &\le (15 + 21|a| + 27|b|)H(P)^3 \max\{|R|, |S|\} \ll 4\Delta H(E)^3 H(P)^3 H(2P). \end{aligned}$$

Therefore $H(2P) \gg H(P)^4/H(E)^{10}$. We have therefore proved what we set out to do:

Proposition 12.2. *If $P \in E(\mathbb{Q})$ then*

$$H(E)^{-10} H(P)^4 \ll H(2P) \ll H(E)^4 H(P)^4. \tag{12.3}$$

`Exercise 3.7`. Prove that if $P = (x, y)$ then $H(E)^{-3} H(x)^3 \ll H(y)^2 \ll H(E)^3 H(x)^3$. (Hint: Separate into the same two cases as above.)

---

[23] Also re-writing $R = (3m^2 + an^4)^2 - 8m\ell^2$ and $S = 4\ell^2 n^2$, we can deduce that $(R, S) = (R, 4\ell^2) = (3m^2 + an^4, 2\ell)^2$.

[24] Why do we choose this for the "height" of the curve, rather than say $\Delta = 4a^3 + 27b^2$? The definition of Δ reflects the fact that the sizes of a^3 and b^2 are comparable, but it may be that when we add $4a^3$ and $27b^2$ there is a lot of cancelation and Δ is in fact substantially smaller. It is obvious that $|\Delta| \le 31H(E)^6$, and the *abc*-conjecture (which is far from proved) implies that $|\Delta| \gg_\epsilon H(E)^{1-\epsilon}$. Quite a wide range of possibilities but it is evident that $H(E)$ provides our best hope of getting a fair measure for the size of the coefficients of E.

Corollary 12.3. *If $P \in E(\mathbb{Q})$ with $H(P) \gg H(E)^{10/3}$ then P has infinite order.*

Proof. If $H(P) \gg H(E)^{10/3}$ then $H(E)^{-10}H(P)^3 \gg 1$ and so $H(2P) > H(P)$ by Proposition 12.2. But then we can proceed by induction to show that $H(P) < H(2P) < H(4P) < \dots$. On the other hand if P has finite order then there exist integers $0 \le i < j$ such that $2^iP = 2^jP$ and so $H(2^iP) = H(2^jP)$, a contradiction.

In order to develop our plan above, we will also need to bound the height of $P + Q$ for a given, fixed rational point Q.

Lemma 12.4. *If $P, Q \in E(\mathbb{Q})$ we have $H(P + Q) \ll H(E)^3H(P)^3H(Q)^3$.*

Proof. Write $Q = (M/N^2, L/N^2)$ and $P = (m/n^2, l/n^3)$ so that

$$x(P+Q) = \left(\frac{y(P)-y(Q)}{x(P)-x(Q)}\right)^2 - x(P) - x(Q) = \left(\frac{Ln^3 - lN^3}{nN(Mn^2 - mN^2)}\right)^2 - \frac{M}{N^2} - \frac{m}{n^2}$$
$$= \frac{(Ln^3 - lN^3)^2 - (Mn^2 + mN^2)(Mn^2 - mN^2)^2}{n^2N^2(Mn^2 - mN^2)^2}\,.$$

Now since $l^2, |m|^3, n^6 \ll H(E)^3H(P)^3$ and $L^2, |M|^3, N^6 \ll H(E)^3H(Q)^3$ (as in exercise 3.7) the denominator is $\ll H(E)^3H(P)^3H(Q)^3$ and the numerator is $\ll H(E)^3H(P)^3H(Q)^3$

Let us suppose that we have another model E' for the elliptic curve E, obtained by mapping $x \to x' := \frac{\alpha x+\beta}{\gamma x+\delta}$ where $\alpha\delta - \beta\gamma = 1$. If $x = p/q$ with $(p, q) = 1$ then $|p'| \le |\alpha p| + |\beta q| \le (|\alpha| + |\beta|)H(x) \le \kappa H(x)$ where $\kappa := |\alpha| + |\beta| + |\gamma| + |\delta|$, and similarly $|q'| \le \kappa H(x)$, which imply that $H(x') \le \kappa H(x)$. Taking the inverse map we deduce that $H(x) \le \kappa H(x')$, and so

$$H(P') \asymp H(P). \tag{12.4}$$

We will return to the subject of heights later and obtain better bounds, and supply a more useful formulation.

13. Mordell's theorem – $E(\mathbb{Q})$ is finitely generated

13.1. The proof of Mordell's Theorem over the rationals. Given that $E(\mathbb{Q})$ is abelian we can write it as $T \times \mathbb{Z}^r$. Here T is the torsion subgroup, and r can be a integer ≥ 0, or even infinity. A remarkable theorem of Mordell shows that r is always finite.[25] His proof proceeds by descent: Given a point P on the elliptic curve with large co-ordinates he shows how to find a point R from a finite set S such that $P - R = 2Q$ for some other point $Q \in E(\mathbb{Q})$. This means that the co-ordinates of Q are about a quarter the length of those of P. One repeats this process with Q, and keeps on going until one arrives at a point of small height (call the set of such points H). This is rather like the Euclidean algorithm, and when we reverse the process we find that P can be expressed as linear combination of elements of S and H, and hence $r \leq |S| + |H|$.

There were several difficult calculations in Mordell's original proof in finding R. Weil made the astute observation that Mordell's process is tantamount to expressing all of $E(\mathbb{Q})$ as a finite set of cosets of $2E(\mathbb{Q})$, and hence it is enough to show that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. Weil came up with an elegant argument which generalizes to many other *algebraic groups* (that is generalizations of $E(\mathbb{Q})$). We now exhibit this argument in the special case that x^3+ax+b has three integer roots, and then will extend this argument to all number fields, using a little algebraic number theory.

Suppose that $x^3+ax+b = (x-r_1)(x-r_2)(x-r_3)$ with $r_1, r_2, r_3 \in \mathbb{Z}$. Given a rational point $P = (m/n^2, \ell/n^3)$ with $(\ell m, n) = 1$ we have $\ell^2 = (m - r_1n^2)(m - r_2n^2)(m - r_3n^2)$. Since $(m - r_in^2, m - r_jn^2) = (m - r_in^2, r_i - r_j)$ each $m - r_in^2 = \alpha_i\beta_i^2$ where α_i is a squarefree integer which divides $(r_i - r_j)(r_i - r_k)$. Hence we define a map

$$\phi : E(\mathbb{Q}) \to \{(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{Q}^*/(\mathbb{Q}^*)^2 : \ \alpha_1\alpha_2\alpha_3 \in (\mathbb{Q}^*)^2\}$$
$$\text{given by } \ \phi(P) = (x - r_1, x - r_2, x - r_3),$$

where α_i is a squarefree divisor of $(r_i - r_j)(r_i - r_k)$. If one of the $x - r_j$ equals 0 then we let α_j be the product of the other two α_i; for example if $x = r_1$ then $\phi((r_1, 0)) = ((r_1 - r_2)(r_1 - r_3), r_1 - r_2, r_1 - r_3)$. Note that $\phi(P) = \phi(-P)$.

We define multiplication in $(\mathbb{Q}^*/(\mathbb{Q}^*)^2)^3$ by $(\alpha_1, \alpha_2, \alpha_3)(\beta_1, \beta_2, \beta_3) = (\alpha_1\beta_1, \alpha_2\beta_2, \alpha_3\beta_3)$.

Proposition 13.1. *Suppose that $x^3+ax+b = (x-r_1)(x-r_2)(x-r_3)$ with $r_1, r_2, r_3 \in \mathbb{Z}$, and define $\phi : E(\mathbb{Q}) \to \{(a, b, c) \in (\mathbb{Q}^*/(\mathbb{Q}^*)^2)^3 : \ abc = 1\}$, by $\phi(P) = (x-r_1, x-r_2, x-r_3)$. Then*

$$\phi(E(\mathbb{Q})) \subset \{(pq, pr, qr) : \ p|r_1 - r_2, \quad q|r_1 - r_3 \quad r|r_3 - r_2, \quad \text{and } q, r \geq 1\}.$$

and so $\phi(E(\mathbb{Q}))$ is finite. Moreover $\phi(P_1 + P_2) = \phi(P_1)\phi(P_2)$. In fact $\ker\phi = 2E(\mathbb{Q})$ so that $E(\mathbb{Q})/2E(\mathbb{Q}) \cong \phi(E(\mathbb{Q}))$.

Corollary 13.2. *If $x^3 + ax + b$ splits over $\mathbb{Z}$ then the rank of $E(\mathbb{Q})$ is finite*

Proof. We can write any abelian group, such as $E(\mathbb{Q})$, in the form $T \oplus \mathbb{Z}^r$ where T is the set of points of finite order. We know that T here is finite by the Lutz-Nagell theorem.

[25] Mordell's argument works in any number field — see section 13.3.

Moreover $E(\mathbb{Q})/2E(\mathbb{Q}) = T/2T \oplus (\mathbb{Z}/2\mathbb{Z})^r$ and $E(\mathbb{Q})/2E(\mathbb{Q}) \cong \phi(E(\mathbb{Q}))$ which is finite, by the Proposition, so that r is finite.

Note that we did not use the notion of height to prove that the rank is finite.

Proof of Proposition 13.1. The first parts have already been proved. We now suppose that we have three points P_1, P_2, P_3 on the line $y = mx + b$. Then their x-co-ordinates are all roots of

$$(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2,$$

and so this monic polynomial equals $(x - x_1)(x - x_2)(x - x_3)$ where $x_j = x(P_j)$. Taking $x = r_i$ we deduce that $(x_1 - r_i)(x_2 - r_i)(x_3 - r_i) = (mr_i + b)^2 \in (\mathbb{Q}^*)^2$ so that

$$\phi(P_1)\phi(P_2)\phi(P_3) = (1, 1, 1),$$

In particular $\phi(P_1 + P_2) = \phi(-P_3) = \phi(P_3) = \phi(P_1)\phi(P_2)$, and so $\phi(2P) = (1, 1, 1)$.

On the other hand suppose that $\phi(Q) = (1, 1, 1)$ where $Q = (U, V)$, so that there exist $t_1, t_2, t_3 \in \mathbb{Q}$ such that $U - r_i = t_i^2$ for each i. Now

$$\begin{aligned}\det \begin{pmatrix} t_1 & r_1 & 1 \\ t_2 & r_2 & 1 \\ t_3 & r_3 & 1 \end{pmatrix} &= \det \begin{pmatrix} t_1 & U - t_1^2 & 1 \\ t_2 & U - t_2^2 & 1 \\ t_3 & U - t_3^2 & 1 \end{pmatrix} \\ &= \det \begin{pmatrix} t_1 & t_1^2 & 1 \\ t_2 & t_2^2 & 1 \\ t_3 & t_3^2 & 1 \end{pmatrix} = \pm(t_1 - t_2)(t_2 - t_3)(t_3 - t_1) \neq 0,\end{aligned}$$

so there exists rational numbers u, m, b such that

$$\begin{pmatrix} t_1 & r_1 & 1 \\ t_2 & r_2 & 1 \\ t_3 & r_3 & 1 \end{pmatrix} \begin{pmatrix} u \\ m \\ b \end{pmatrix} = \begin{pmatrix} t_1 r_1 \\ t_2 r_2 \\ t_3 r_3 \end{pmatrix}.$$

Therefore $mr_i + b = -t_i(u - r_i)$ for each i so that the monic polynomial

$$(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2$$

takes value $-(u - r_i)^2 t_i^2 = (r_i - u)^2(r_i - U)$ at $x = r_i$. Hence

$$(x - r_1)(x - r_2)(x - r_3) - (mx + b)^2 = (x - u)^2(x - U).$$

Taking $x = u$ yields the rational points $\pm P = (u, \pm(mu + b))$ on the curve, and one verifies that $Q = -2P$. Therefore $\ker \phi = 2E$. Since $\ker \phi = 2E$, know that the image of ϕ is isomorphic to $E/2E$.

`Exercise 13.1.1.` Prove that $|T/2T|$ equals the number of points of order 1 or 2. Deduce if there are 2^t points of order 1 or 2 (the possibilities being $t = 0, 1$ or 2), and the image of ϕ contains 2^s elements, then the rank of $E(\mathbb{Q})$ equals $r = s - t$.

In honor of their work the group of points $E(\mathbb{Q})$ is known as the *Mordell-Weil group*.

13.2. Another example: Four squares in an arithmetic progression. Fermat proved, by descent, that there are no four distinct squares in an arithmetic progression. Let's see how we can prove this using the Mordell-Weil group and our map ϕ. If $a-d,\ a,\ a+d$ and $a+2d$ are all squares, say $u_{-1}^2, u_0^2, u_1^2, u_2^2$ then $(-2d/a,\ 2u_{-1}u_0u_1u_2/a^2)$ is a point on the elliptic curve

$$E:\ y^2 = (x-1)(x-2)(x+2).$$

Here $t=2$ again but things are a bit more complicated, since $P=(0,2)$ is a point of order 4. Then $2P=(2,0)$ and the other points of order two are $R=(1,0)$ and $R+2P=(-2,0)$. We also have another point of order four namely $R-P=(4,6)$, as well as $-P$ and $P-R$.

Let R be the group $\{(\alpha_1,\alpha_2,\alpha_3)\in\mathbb{Q}^*/(\mathbb{Q}^*)^2 :\ \alpha_1\alpha_2\alpha_3\in(\mathbb{Q}^*)^2\}$. By the Proposition, we know that $\phi(E(\mathbb{Q}))\cong E(\mathbb{Q})/2E(\mathbb{Q})$ is a subgroup, and in fact a subgroup of G, which is generated by $(-1,-1,0),(1,2,2)$ and $(3,1,3)$. Moreover H is a subgroup of $\phi(E(\mathbb{Q}))$ where $H=\{\phi(\mathcal{O})=\phi(2P)=(1,1,1),\ \phi(P)=(-1,-2,2),\ \phi(P-R)=(3,2,6),\ \phi(R)=(-3,-1,3)\}$.

We now show that $\phi(E(\mathbb{Q}))\neq G$, else there exists $(x,y)\in E(\mathbb{Q})$ with $\phi(x,y)=(1,2,2)$. If so then we can write $x-1=(b/c)^2$ where $(b,c)=1$ by exercise 12.1.1, and hence we also have $b^2-c^2=2v^2,\ b^2+3c^2=2w^2$ for some integers v and w. Now b and c are both odd (since $b+c\equiv 0 \pmod 2$ and they are coprime), but then $2w^2\equiv 1+3=4 \pmod 8$ which is impossible.

Now since $(\mathbb{Z}/2\mathbb{Z})^2\cong H\subset\phi(E(\mathbb{Q}))\subsetneq G\cong(\mathbb{Z}/2\mathbb{Z})^3$, hence $E(\mathbb{Q})/2E(\mathbb{Q})\cong\phi(E(\mathbb{Q}))=H\cong(\mathbb{Z}/2\mathbb{Z})^2$. We have already seen that $E(\mathbb{Q})$ contains the torsion subgroup $(\mathbb{Z}/2\mathbb{Z})^2$, and hence we can deduce that the rank of $E(\mathbb{Q})$ is 0.

Now if S is a point that corresponds to an example where $a-d,\ a,\ a+d$ and $a+2d$ are all squares, then $\phi(S)=(-1,-2,2)$, and so $S=\pm P=(0,\pm 2)$ and hence $d=0$. Therefore there are no four distinct squares in an arithmetic progression.

In this case we have seen a new phenomenon: Let $S(E(\mathbb{Q}))$ be the set of triples of coprime squarefree integers (uv,uw,vw) with $u|r_1-r_2,\ \ v|r_1-r_3\ \ w|r_3-r_2$ and $v,w\geq 1$. In Proposition 13.1 we saw that $\phi(E(\mathbb{Q}))\subset S(E(\mathbb{Q}))$. Now let $S_2(E(\mathbb{Q}))$ be the set of triples $(pq,pr,qr)\in S(E(\mathbb{Q}))$ such that for all prime powers p^e there exist $m,n,y_1,y_2,y_3 \pmod{p^e}$, such that y_1^2,y_2^2,y_3^2 are not all $0 \pmod{p^e}$ such that

$$m-r_1n^2\equiv uvy_1^2,\ \ m-r_2n^2\equiv uwy_2^2,\ \ m-r_3n^2\equiv vwy_3^2 \pmod{p^e}.$$

We see that $\phi(E(\mathbb{Q}))\subset S_2(E(\mathbb{Q}))\subset S(E(\mathbb{Q}))$ in general. In this last case $E(\mathbb{Q})/2E(\mathbb{Q})\cong\phi(E(\mathbb{Q}))=S_2(E(\mathbb{Q}))\subsetneq S(E(\mathbb{Q}))$.

Here $S_2(E(\mathbb{Q}))$ is called the 2-*Selmer group*. More generally we can study the equivalence classes of $E(\mathbb{Q})/mE(\mathbb{Q})$ that are solvable modulo all prime powers to obtain $S_m(E(\mathbb{Q}))$, the *m-Selmer group*. In fact $S_2(E(\mathbb{Q}))\supset S_4(E(\mathbb{Q}))\supset S_8(E(\mathbb{Q}))\supset\ldots\supset E(\mathbb{Q})/2E(\mathbb{Q})$, and it believed that there exists some power of 2 such that $E(\mathbb{Q})/2E(\mathbb{Q})\cong S_{2^k}(E(\mathbb{Q}))$, though this is, for now, a conjecture.[26]

[26]To add: computing by day-and-night story from Tate's paper

`Exercise 13.2.1.` Let E be the elliptic curve $y^2 = x(x+2)(x+16)$.

(1) Prove that the point $P = (2, 12) \in E(\mathbb{Q})$ has infinite order. Deduce that $r \geq 1$
(2) Prove that there are no integer solutions to $u^2 + 2v^2 = 7w^2$. Deduce that $s < 4$.
(3) Show that $t = 2$ and hence deduce that r, the rank of $E(\mathbb{Q})$, is one.

In general one can write down generators of the Mordell-Weil group, say $P_1, P_2, \ldots, P_r$ and all points on $E(\mathbb{Q})$ can be written as $a_1P_1 + a_2P_2 + \ldots + a_rP_r$ for some integers $a_1, \ldots, a_r$. If P_j has infinite order then we take any $a_j \in \mathbb{Z}$; if P_j has order m_j then we take $a_j \pmod{m_j}$. We can add points, by adding the vectors of the a_j componentwise, and according to these rules.

Let's suppose that we are given $Q_1, \ldots, Q_s \in E(\mathbb{Q})$ for which $\{\phi(Q_1), \ldots, \phi(Q_s)\} = \phi(E(\mathbb{Q}))$. In the proof described above, given any $P \in E(\mathbb{Q})$, we first find Q_j such that $\phi(Q_j) = \phi(P)$ and then there exists $R \in E(\mathbb{Q})$ for which $2R = P + Q_j$. By Proposition 12.2 and Lemma 12.4 we have $H(R)^4 \ll H(E)^{10}H(2R) = H(E)^{10}H(P + Q_j) \ll H(E)^{13}H(P)^3H(Q_j)^3$. Therefore $H(R) < H(P)$ if $H(P) \gg H(E)^{13}\max_j H(Q_j)^3$. This shows us how to construct the generators of $E(\mathbb{Q})$ *given* a set of representatives of $E(\mathbb{Q})/2E(\mathbb{Q})$.

If there are infinitely many points in $E(\mathbb{Q})$, how are they spaced on the curve itself? Are they dense on the curve? This sort of question can be answered but requires methods from beyond this discussion.

How big the rank can get is an open question. Researchers have found elliptic curves for which $E(\mathbb{Q})$ has rank at least 28, and some people believe that ranks of $E(\mathbb{Q})$ can get arbitrarily large as we vary over elliptic curves E, though for now this is more a belief than a conjecture.

We have already seen something similar to the notion of Mordell-Weil groups when we were considering solutions to Pell's equation. There all solutions take the form $\pm\epsilon^a$ so that this group of units is generated by -1 and ϵ_d and has structure $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$, this ± 1 being torsion. There can be more torsion than just ± 1; for example, in $\mathbb{Z}[i]$ we also have the units $\pm i$ so the unit group structure is $\mathbb{Z}/4\mathbb{Z}$, generated by i.

`Exercise 13.2.2.` The size of ϵ_d^n grows exponentially in n. How fast does 2^kP grow (as a function of k)? Can you then deduce a result about the growth of nP?

13.3. Mordell's Theorem in number fields. Suppose that $E: \ y^2 = x^3 + ax + b = (x-r_1)(x-r_2)(x-r_3)$ and we wish to study $E(K)$ for some number field K that contains the roots r_1, r_2, r_3. This is again an abelian group so that we can write $E(K) \cong T(K) \oplus \mathbb{Z}^{r_E(K)}$. We will show here that $r_E(K)$ is bounded. The key difficulty in translating all the proofs above over to K is that the analogy to exercise 12.1.1 no longer holds. What does hold is that the ideals $(x) = M/N^2$, $(y) = L/N^3$ where L, M, N are integral ideals of K with $(LM, N) = 1$; however these ideals are not necessarily principal. We did take care to write the relevant proofs above for $E(\mathbb{Q})$ in such a way that the relevant parts of the proofs also hold for $E(K)$.

As before we define a map

$$\phi : E(K) \to \{(\alpha_1, \alpha_2, \alpha_3) \in K^*/(K^*)^2 : \ \alpha_1\alpha_2\alpha_3 \in (K^*)^2\}$$
$$\text{given by} \ \ \phi(P) = (x - r_1, x - r_2, x - r_3).$$

If one of the $x - r_j$ equals 0 then we let that entry be the product of the other two. We define multiplication in $(K^*/(K^*)^2)^3$ by $(\alpha_1, \alpha_2, \alpha_3)(\beta_1, \beta_2, \beta_3) = (\alpha_1\beta_1, \alpha_2\beta_2, \alpha_3\beta_3)$. The proof of Proposition 13.1 was written so that it carries over to this generality so that $\phi(P_1 + P_2) = \phi(P_1)\phi(P_2)$, and $\ker \phi = 2E(K)$ so that $E(K)/2E(K) \cong \phi(E(K))$. We will show that $\phi(E(K))$ is finite, which implies $r(K)$ is finite, since $|\phi(E(K))| = 2^{r(K)+2}$.

The key idea is that the class group $\mathcal{C}(K)$ of K is finite, and we will pick of a set of representative ideals, one for each equivalence class. Given $(x) = M/N^2$, $(y) = L/N^3$ where L, M, N are integral ideals of K with $(LM, N) = 1$ select C to be the representative of the inverse to the ideal class of N. Then $CN = (\nu)$ for some algebraic integer ν and $(x) = C^2M/(CN)^2$, so that $C^2M = (x\nu^2)$ is a principal ideal. That is we can write $x = \mu/\nu^2$ where μ and ν are algebraic integers as described above, and similarly $y = \lambda/\nu^3$ where $C^3L = (\lambda)$. Therefore

$$(x - r_1, x - r_2, x - r_3) = (\mu - r_1\nu^2, \mu - r_2\nu^2, \mu - r_3\nu^2) \text{ in } (K^*/(K^*)^2)^3.$$

Since $(\nu^2) = C^2N^2$ and $(\mu) = C^2M$ we know that $(\nu^2, \mu) = C^2$. Since the product of the three co-ordinates here is a square we know that there exist squarefree coprime integral ideals I_1, I_2, I_3 and ideals W_1, W_2, W_3 such that

$$(\mu - r_1\nu^2) = I_2I_3(CW_1)^2, \quad (\mu - r_2\nu^2) = I_1I_3(CW_2)^2, \quad (\mu - r_3\nu^2) = I_1I_2(CW_3)^2.$$

Now we can study the gcds of the pairs of elements:

$$C^2 \supseteq (\mu - r_1\nu^2, \mu - r_2\nu^2) \supseteq ((r_1 - r_2)\nu^2, (r_1 - r_2)\mu) = C^2(r_1 - r_2).$$

Therefore $(\mu - r_1\nu^2, \mu - r_2\nu^2) = I_3D^2$ for some squarefree ideal I_3 dividing $(r_1 - r_2)$.

Now select C_j to be the representative of the inverse of the ideal class of CW_j so that C_jCW_j is principal, say $= (w_j)$ for some algebraic integer w_j. Suppose that D_j is the representative of the ideal class of $(CW_j)^2$, so that $D_jI_iI_k$ is principal, say (g_j), as well as $D_jC_j^2 = (e_j)$. Hence we have

$$(e_j(\mu - r_j\nu^2)) = D_jC_j^2(\mu - r_j\nu^2) = D_jI_iI_k(C_jCW_j)^2 = (g_jw_j^2).$$

The algebraic integers on each side differ by a unit of the ring of units, $U(K)$, of K. We know that this is finitely generated, and so any unit can be written as $u_jv_j^2$ where u_j is a representative of an equivalence class of $U(K)/U(K)^2$, which is finite. Hence we have $e_j(\mu - r_j\nu^2) = u_jg_j(v_jw_j)^2$. Therefore

$$(x - r_1, x - r_2, x - r_3) = (u_1e_1g_1, u_2e_2g_2, u_3e_3g_3) \text{ in } (K^*/(K^*)^2)^3.$$

What are the total number of possibilities? We begin by selecting I_1, I_2, I_3, coprime squarefree ideals that divide $(r_3 - r_2), (r_1 - r_3), (r_1 - r_2)$, respectively. This yields the D_j, and so the g_j, for $j = 1, 2$. Then the C_j must come from an equivalence class for which C_j^2 is the inverse to D_j (in the class group), and so the number of possibilities is $|\mathcal{C}(K)|/|\mathcal{C}(K)^2|$. Finally we need to select the appropriate u_j for $j = 1, 2$ from $U(K)/U(K)^2$. Note that

once we have determined the first two entries of $\phi(P)$, we can simply take the third entry to be the product of the first two. Hence the number of possibilities is bounded by the number of triples of coprime squarefree ideals that divide $(r_3 - r_2), (r_1 - r_3), (r_1 - r_2)$, times $(|\mathcal{C}(K)/\mathcal{C}(K)^2|)^2 \cdot (|U(K)/U(K)^2|)^2$. This yields an upper bound

$$r_E(K) \le \omega_K(\Delta) + 2\omega_K(\Delta') + 2R_2(\mathcal{C}(K)) + 2R_2(U(K)) - 2$$

where $\Delta' = (r_3 - r_2,\ r_1 - r_3,\ r_1 - r_2)$, $\omega_K(I)$ is the number of distinct prime ideal factors of I in K, and $R_2(G)$ is the 2-rank of the group G, since $\Delta = ((r_3 - r_2)(r_1 - r_3)(r_1 - r_2))^2$ and $R_2(\text{Torsion}(E(K))) = 2$.

What can we say about the size of Torsion$(E(K))$*?* $E(K)$ does contain all three points of order two, namely $(r_j, 0), j = 1, 2, 3$. Otherwise we can follow the proof of Lutz-Nagell theorem above, beginning from the change of variable $(x, y) \to (u, v)$, and showing that if $P = (x, y)$ is a torsion point of $E(K)$ then x and y are algebraic integers (that is (x) and (y) are integral ideals). By (12.1) we then deduce that if P has order > 2 then y^2 divides Δ. There are indeed finitely many ideals (y) for which $(y)^2$ divides Δ; however, multiplying by units, there may be infinitely many such y.[27]

13.4. More precise bounds on naive height.

We again define $H(E) := \max\{|a|^3, |b|^2\}^{1/6}$ and revisit Lemma 12.4.

Proposition 13.3. *If* $P, Q \in E(\mathbb{Q})$ *then*

$$H(P + Q)H(P - Q) \ll H(E)^4 H(P)^2 H(Q)^2$$

Note that $\min\{H(P + Q), H(P - Q)\} \ll H(E)^2 H(P)H(Q)$. Since $\phi(-Q) = \phi(Q)$, we have $\phi(P - Q) = \phi(P + Q)$ in the proof of Mordell's Theorem, so we choose $\pm Q$ to minimize the height of $H(P \pm Q)$ in that proof, and so if $2R = P \pm Q_j$ then $H(R)^4 \ll H(E)^{12} H(P)H(Q_j)$. Therefore $H(R) < H(P)$ if $H(P) \gg H(E)^4 \max_j H(Q_j)^{1/3}$. In fact if Q_j is chosen to minimize $H(Q_j)$ over all rational points in the same equivalence class of $E(\mathbb{Q})/2E(\mathbb{Q})$ then $H(Q_j) \le H(P)$ and so $H(R) < H(P)$ if $H(P) \gg H(E)^6$.

Corollary 13.4. *If* $P, Q \in E(\mathbb{Q})$ *then*

$$H(E)^{-12} H(P)^2 H(Q)^2 \ll H(P + Q)H(P - Q) \ll H(E)^4 H(P)^2 H(Q)^2$$

Proof. We substitute $P \to P+Q,\ Q \to P-Q$ in Proposition 13.3, and then use Proposition 12.2, to obtain

$$H(E)^4 H(P + Q)^2 H(P - Q)^2 \gg H(2P)H(2Q) \gg H(E)^{-20} H(P)^4 H(Q)^4,$$

and the claimed lower bound follows.

Note that Corollary 13.4 contains with $P = Q$ yields a version of Proposition 13.3.

[27] We need to deal with this!

Proof of Proposition 13.3. Let $P_1 = P,\ P_2 = Q,\ P_3 = P+Q,\ P_4 = P-Q$ and write $x_j = x(P_j)$ and $P_j = (m_j/n_j^2, l_j/n_j^3)$ for each j. Now $P_1, P_2, -P_3$ are on a line, as are $-P_1, P_2, P_4$, and so, as $x(P_j) = x(-P_j)$,

$$x_1+x_2+x_3 = \left(\frac{y_2-y_1}{x_2-x_1}\right)^2 \quad \text{and} \quad x_1+x_2+x_4 = \left(\frac{y_2+y_1}{x_2-x_1}\right)^2.$$

Therefore

$$\begin{aligned} x_3, x_4 =& \frac{y_1^2+y_2^2 \pm 2y_1y_2}{(x_2-x_1)^2} - (x_1+x_2) = \frac{(x_1+x_2)(x_1x_2+a)+2b \pm 2y_1y_2}{(x_2-x_1)^2} \\ =& \frac{(m_1n_2^2+m_2n_1^2)(m_1m_2+an_1^2n_2^2)+2bn_1^4n_2^4 \pm 2l_1l_2n_1n_2}{(m_2n_1^2-m_1n_2^2)^2}, \end{aligned}$$

and then we deduce

$$\begin{aligned} x_3+x_4 =& \frac{2(m_1n_2^2+m_2n_1^2)(m_1m_2+an_1^2n_2^2)+4bn_1^4n_2^4}{(m_2n_1^2-m_1n_2^2)^2}, \\ x_3x_4 =& \frac{((m_1n_2^2+m_2n_1^2)(m_1m_2+an_1^2n_2^2)+2bn_1^4n_2^4)^2-(2l_1l_2n_1n_2)^2}{(m_2n_1^2-m_1n_2^2)^4} \\ =& \frac{(m_1m_2-an_1^2n_2^2)^2-4bn_1^2n_2^2(m_1n_2^2+m_2n_1^2)}{(m_2n_1^2-m_1n_2^2)^2}, \end{aligned}$$

the last equality following, somewhat surprisingly, after substituting $l_j^2 = m_j^3+am_jn_j^4+bn_j^6$. The result follows from:

`Exercise` 13.4.1. Write $x = a/b,\ y = A/B$ with $(a,b) = (A,B) = 1$ and $b, B \geq 1$.
(i) Show that the least common denominator of $x+y$ and xy is bB.
(ii) Show that $H(x)H(y) \leq 2\max\{|aB+Ab|, |aA|, bB\}$.
(iii) Deduce that if $x+y = R/D,\ xy = S/D$ then $H(x)H(y) \leq 2\max\{|R|, |S|, |D|\}$.

13.5. Néron-Tate height. We define $h(P) = \log H(P)$ so that Proposition 12.2 reads that

$$|h(2P) - 4h(P)| \leq 10h(E) + C; \tag{13.1}$$

and Corollary 13.4 reads that

$$|h(P+Q) + h(P-Q) - 2h(P) - 2h(Q)| \leq 12h(E) + C \tag{13.2}$$

for some absolute constant $C > 0$. If we replace P by $2^{k-1}P$ in (13.1) and divide through by 4^k then we obtain

$$\left|\frac{h(2^kP)}{4^k} - \frac{h(2^{k-1}P)}{4^{k-1}}\right| \leq \frac{10h(E)+C}{4^k} \quad \text{for each } n \geq 1.$$

Summing this up over $k = m+1, m+2, \dots, n$ we obtain

$$\left| \frac{h(2^n P)}{4^n} - \frac{h(2^m P)}{4^m} \right| \leq \frac{10h(E) + C}{3 \cdot 4^{m-1}},$$

and so the $h(2^n P)/4^n$ form a Cauchy sequence and therefore converge. We call this limit the *Néron-Tate height*,

$$\hat{h}(P) := \lim_{n \to \infty} \frac{\log H(2^n P)}{4^n} . \tag{13.3}$$

`Exercise` 13.5.2. This height has lots of useful properties. Prove the following
(i) $|\hat{h}(P) - \log H(P)| \leq 14 \log H(E) + 2C$.
(ii) For all B, there are finitely many $P \in E(\mathbb{Q})$ with $\hat{h}(P) \leq B$.
(iii) $\hat{h}(2P) = 4\hat{h}(P)$. (Hint: Use the definition and (13.1).)
(iv) P has finite order if and only if $\hat{h}(P) = 0$. (Hint: Use exercise 3.1, and use (ii) for $\Leftarrow$.)
(v) $\hat{h}(P) \geq 0$ for all P.

`Exercise` 13.5.3 (i) Prove that

$$\hat{h}(P+Q) + \hat{h}(P-Q) = 2\hat{h}(P) + 2\hat{h}(Q). \tag{13.4}$$

(Hint: Replace P by $2^n P$, and Q and $2^n Q$ in (13.2) and proceed as before.)
(ii) Prove that

$$\hat{h}(mP) = m^2 \hat{h}(P) \quad \text{for all} \quad m \geq 1.$$

(Hint: Prove this by induction on m, taking $Q = (m-1)P$ in (13.4).)

`Exercise` 13.5.4. Prove that if we have another model for the elliptic curve E, in which point P gets mapped to point P' then $\hat{h}(P) = \hat{h}(P')$. That is the Néron-Tate height does not change when we change the model of the curve, and is thus truly *canonical*. (Hint: Use (3.4).)

We analyze the algorithm suggested by Mordell's proof again, this time using $\hat{h}$: Again we pre-select the Q in each equivalence class of $E(\mathbb{Q})/2E(\mathbb{Q})$ to minimize $\hat{h}(Q)$. So, given P, select the representative Q in the same equivalence class of $E(\mathbb{Q})/2E(\mathbb{Q})$ as P, and we have $P \pm Q = 2R$, Selecting the sign to minimize $\hat{h}(P \pm Q)$ we have $4\hat{h}(R) = \hat{h}(2R) = \hat{h}(P \pm Q) \leq \hat{h}(P) + \hat{h}(Q) \leq 2\hat{h}(P)$ and so $\hat{h}(R) \leq \frac{1}{2}\, \hat{h}(P)$. Thus this algorithm always provides a "smaller" point unless $\hat{h}(P) = 0$, and so the algorithm ends in a torsion point (which may be $\mathcal{O}$) by exercise 13.5.2(iv).

13.6. An inner product. Define

$$\langle P, Q \rangle := \frac{1}{2} \left(\hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q) \right)$$

By exercise 13.5.2(iii) we see that $\langle P, P \rangle = \hat{h}(P)$. By definition, we have $\langle P, Q \rangle = \langle Q, P \rangle$.

Proposition 13.5. *Suppose that $P, Q \in E(\mathbb{Q})$. Then*
(i) $\langle P, -Q\rangle = -\langle P, Q\rangle$.
(ii) $\langle\ ,\ \rangle$ *is a bilinear form.*
(iii) $\hat{h}(mP + nQ) = m^2\hat{h}(P) + 2mn\langle P, Q\rangle + n^2\hat{h}(Q)$ *for any integers* m, n.

Proof. (i) follows immediately from (13.4). For (ii) we begin by noting, using (13.4) twice, that

$$\begin{aligned}\langle P+Q, R\rangle + \langle P-Q, R\rangle =& \frac{1}{2}\left(\hat{h}(P+R+Q) + \hat{h}(P+R-Q) - \hat{h}(P+Q) - \hat{h}(P-Q)\right) - \hat{h}(R)\\ &= \hat{h}(P+R) + \hat{h}(Q) - \hat{h}(P) - \hat{h}(Q) - \hat{h}(R) = 2\langle P, R\rangle.\end{aligned}$$

Now interchange P and Q, to obtain $\langle P+Q, R\rangle - \langle P-Q, R\rangle = 2\langle Q, R\rangle$, so that, adding the last two equations, we obtain

$$\langle P+Q, R\rangle = \langle P, R\rangle + \langle Q, R\rangle.$$

Applying this several times, this implies that for any integers m, n we have

$$\begin{aligned}\hat{h}(mP+nQ) = \langle mP+nQ, mP+nQ\rangle &= m^2\langle P, P\rangle + 2mn\langle P, Q\rangle + n^2\langle Q, Q\rangle\\ &= m^2\hat{h}(P) + 2mn\langle P, Q\rangle + n^2\hat{h}(Q).\end{aligned}$$

Corollary 13.6. $|\langle P, Q\rangle| \le \sqrt{\hat{h}(P)\hat{h}(Q)}$.

Proof. Suppose not. The quadratic form $\hat{h}(P)x^2 + 2\langle P, Q\rangle xy + \hat{h}(Q)y^2$ equals $\hat{h}(P)(\hat{h}(P)\hat{h}(Q) - \langle P, Q\rangle^2)$ at $x = -\langle P, Q\rangle,\ y = \hat{h}(P)$, and this is < 0 if $\hat{h}(P) \neq 0$.

`Exercise` 13.6.1. Deduce that if m/n is a good approximation to x/y then $m^2\hat{h}(P) + 2mn\langle P, Q\rangle + n^2\hat{h}(Q) < 0$.

In other words, $\hat{h}(mP+nQ) < 0$ by Proposition 13.5, which contradicts exercise 13.5.2(v). If $\hat{h}(P) = 0$ but $\hat{h}(Q) \neq 0$ then we construct the analogous argument by taking $y = -\langle P, Q\rangle,\ x = \hat{h}(Q)$. If $\hat{h}(P) = \hat{h}(Q) = 0$ then P and Q are torsion points by exercise 13.5.2(iv), so that $P + Q$ is also a torsion point and hence $\hat{h}(P+Q) = 0$ by exercise 13.5.2(iv), and therefore $\langle P, Q\rangle = 0$ by definition.

Corollary 13.7. *If Q has finite order then $\langle P, Q\rangle = 0$, and $\hat{h}(P+Q) = \hat{h}(P)$.*

Proof. By exercise 13.5.2(iv), we have $\hat{h}(Q) = 0$. Therefore $\langle P, Q\rangle = 0$ by Corollary 13.6, and then $\hat{h}(P+Q) = \hat{h}(P) + \hat{h}(Q) + 2\langle P, Q\rangle = \hat{h}(P)$.

The second part of Corollary 5b, suggests that $\hat{h}$ should be viewed as the height on $E(\mathbb{Q})/\mathrm{Torsion}(E(\mathbb{Q}))$.

13.7. The number of points in the Mordell-Weil lattice up to height x. Let $P_1, \dots P_r$ be a basis for $E(\mathbb{Q})/\text{Torsion}(E(\mathbb{Q}))$, so that any point in $E(\mathbb{Q})$ may be written as

$$m_1P_1 + m_2P_2 + \ldots + m_rP_r + t \text{ each } m_i \in \mathbb{Z},$$

and $t \in \text{Torsion}(E(\mathbb{Q}))$. By Proposition 13.5 we deduce that

$$\hat{h}(P) = F_{E(\mathbb{Q})}(m_1, \dots, m_r) \text{ where } F_{E(\mathbb{Q})}(x_1, \dots, x_r) := \sum_{1 \le i,j \le r} \langle P_i, P_j \rangle \, x_i x_j,$$

is a positive definite quadratic form. The determinant is denoted

$$\text{Regulator}(E(\mathbb{Q})) = R_{E(\mathbb{Q})} = \det(\langle P_i, P_j \rangle)_{1 \le i,j \le r}$$

and is called the *regulator* of $E(\mathbb{Q})$. This value is independent of the choice of basis of $E(\mathbb{Q})/\text{Torsion}(E(\mathbb{Q}))$ (this is a fact about the bases of any given lattice).

We will bound the number of points up to x, but first we need to prove a general result about positive definite quadratic forms. For a given quadratic form $F = \sum_{i,j} f_{i,j} x_i x_j$, where $f_{i,j} = f_{j,i}$ we define $\text{Disc}(F)$ to be the determinant of the matrix with (i,j)th entry $f_{i,j}$.

Lemma 13.8. *Suppose that $F(x_1, \dots, x_r) \in \mathbb{R}[x_1, \dots, x_r]$ is a positive definite quadratic form with Disc$(F) \neq 0$. Then*

$$\#\{m_1, \dots, m_r \in \mathbb{Z} : \ F(m_1, \dots, m_r) \le x\} \sim \frac{B_r}{|\text{Disc}(F)|^{1/2}} \ x^{r/2},$$

where B_r is the volume of the r-dimensional unit ball.

Proof. For any $m = (m_1, \dots, m_r) \in \mathbb{R}^r$ let U_m be the unit cube whose center is m, so that $U_m = \{m + \delta : \text{ each } |\delta_i| \le 1/2\}$. Let $\mathcal{U}(x)$ be the union of U_m with $m_1, \dots, m_r \in \mathbb{Z}$ and $F(m_1, \dots, m_r) \le x$, so that $\mathcal{U}(x)$ has volume equal to what we are trying to estimate.

Let $\mathcal{T}(x) = \{x_1, \dots, x_r \in \mathbb{R} : \ F(x_1, \dots, x_r) \le x\}$. As F is a quadratic form we have $\text{Vol}(\mathcal{T}(x)) = x^{r/2}\text{Vol}(\mathcal{T}(1))$.

Also, since F is a quadratic form, $|F(m+\delta) - F(m) - F(\delta)| \le 2\sqrt{|F(m)F(\delta)|}$ by the Cauchy-Schwarz inequality. Therefore if $F(m) \le x$ and $u \in U_m$ then $F(u) \le x + 2\sqrt{\theta x} + \theta$ where $\theta := \max F(\delta)$. Taking the union of the U_m in $\mathcal{U}(x)$, we deduce that if $u \in \mathcal{U}(x)$ then $F(u) \le x + 2\sqrt{\theta x} + \theta$. Hence $\mathcal{U}(x) \subset \mathcal{T}(x + 2\sqrt{\theta x} + \theta)$.

On the other hand suppose that $u \in \mathbb{R}^r$ with $F(u) \le x - 2\sqrt{\theta x} + \theta$. There exists an $m = (m_1, \dots, m_r) \in \mathbb{Z}^r$ such that $u \in U_m$; moreover, proceeding as above, with $\delta = u - m$, we have $F(m) \le F(u) + F(\delta) + 2\sqrt{|F(u)F(\delta)|} \le x$, and hence $\mathcal{T}(x - 2\sqrt{\theta x} + \theta) \subset \mathcal{U}(x)$.

Combining the last three paragraphs yields

$$(\sqrt{x} - \sqrt{\theta})^r \ \text{Vol}(\mathcal{T}(1)) \le \text{Vol}(\mathcal{U}(x)) \le (\sqrt{x} + \sqrt{\theta})^r \ \text{Vol}(\mathcal{T}(1)).$$

Taking $x \to \infty$ the result follows by proving that $\text{Vol}(\mathcal{T}(1)) = B_r/|\text{Disc}(F)|^{1/2}$, which follows from:

Exercise 13.7.1. (i) Show that for any quadratic form we can change variables via a matrix of determining one, to obtain an equivalent quadratic form that is diagonal, with all diagonal entries poitive.
(ii) Verify that this does not alter $\mathrm{Vol}(\mathcal{T}(1))$ or $|\mathrm{Disc}(F)|$.
(iii) Show that if $F = \sum_i a_i x_i^2$ with each $a_i > 0$ then $\mathrm{Vol}(\mathcal{T}(1)) = B_r/(a_1 a_2 \dots a_r)^{1/2}$, by an appropriate change of variables.

Proposition 13.9. *Let $r =$rank$(E(\mathbb{Q}))$. If $r = 0$ then $\#\{P \in E(\mathbb{Q})\} = |\mathrm{Torsion}(E(\mathbb{Q}))|$. If $r \geq 1$ then, as $x \to \infty$,*

$$\#\{P \in E(\mathbb{Q}) : \ \hat{h}(P) \leq x\} \sim \frac{|\mathrm{Torsion}(E(\mathbb{Q}))|}{|\mathrm{Regulator}(E(\mathbb{Q}))|^{1/2}} \ \ B_r \ x^{r/2},$$

where B_r is the volume of the r-dimensional unit ball.

Proof. The case $r = 0$ is trivial. For $r \geq 1$ we use our formula for $\hat{h}(P)$ and then Lemma 13.8 with $F = F_{E(\mathbb{Q})}$.

Exercise 13.7.2. Deduce that

$$\#\{P \in E(\mathbb{Q}) : \ H(P) \leq x\} \sim \frac{|\mathrm{Torsion}(E(\mathbb{Q}))|}{|\mathrm{Regulator}(E(\mathbb{Q}))|^{1/2}} \ \ B_r \ (\log x)^{r/2}.$$

Exercise 13.8. *Magic squares and elliptic curves*: A *magic square* is a 3-by-3 array of integers in which the row and column sums are all equal. For example,

4	3	8
9	5	1
2	7	6

A) Explain how all magic squares may be parametrized in terms of 5 variables. Show that, for integers x_1, x_2, x_3, a, b, we have such a parametrization from the square:

x_1	$x_2 + b$	$x_3 + a$
$x_3 + b$	$x_1 + a$	x_2
$x_2 + a$	x_3	$x_1 + b$

B) Prove that magic squares in which every entry is the square of an integer are in 1-1 correspondence with sets

$$(E_{a,b}, P_1, P_2, P_3)$$

where $E_{a,b}$ denotes the elliptic curve $E_{a,b} : \ y^2 = x(x + a)(x + b)$, each $P_i \in E(\mathbb{Z})$ and $2P_i = (x_i, y_i)$.

C) Find such a magic square (i.e. of squares) with all entries distinct and non-zero (use a computer and (b)).

D) A *super magic square* is one where row, column and diagonal sums are all equal (i.e. $4+5+6 = 2+5+8$ above). Show that super magic squares are parametrized with three integers n, x, d as in

$x - n$	$x - d + n$	$x + d$
$x + d + n$	x	$x - d - n$
$x - d$	$x + d - n$	$x + n$

`E)` Prove that super magic squares, where every entry is a square, are in 1-1 correspondence with sets (E_n, P_1, P_2, P_3) where the elliptic curve $E_n : \ y^2 = x^3 - n^2 x$, each $P_i \in E(\mathbb{Z})$ with $2x_{2P_1} = x_{2P_2} + x_{2P_3}$.

`F)` *Either* show that such squares cannot exist *or* find an example. (This is an open problem.)

14. The Weierstrass $\wp$-function, and endomorphisms of elliptic curves

In this chapter we will head in a seemingly different direction: We are going to classify doubly periodic functions $f : \mathbb{C} \to \mathbb{C}$, but then we find that they parametrize elliptic curves!

14.1. Double periodicity and the Weierstrass $\wp$-function. One can classify the periodic functions $f : \mathbb{R} \to \mathbb{C}$ which are meromorphic inside their period: If $f(x)$ has period t, that is $f(x + t) = f(x)$ for all x, then $F(x) := f(tx)$ has period one (so we can assume that the period is 1, wlog). Examples include $\cos 2\pi x$ and $\sin 2\pi x$ which are both rational functions of $e^{2i\pi x}$, and one can prove that *all* such f are rational functions of $e^{2i\pi x}$. We now will try to establish a similar classification result for *doubly periodic functions* $f : \mathbb{C} \to \mathbb{C}$; by "doubly periodic" we mean that there are two periods τ_1, τ_2 such that $f(z + \tau_1) = f(z)$ and $f(z + \tau_2) = f(z)$ (and so $\tau_1, \tau_2 \neq 0$). But then $F(z) := f(z\tau_1)$ has the two periods $1, \tau$ where $\tau = \tau_2/\tau_1$. So, wlog, the periods are $1, \tau$.

`Exercise 14.1.1.` Prove that (i) if $\tau \in \mathbb{R} \setminus \mathbb{Q}$ then $f(z)$ is a constant; and (ii) if $\tau = r/s \in \mathbb{Q}$, with $(r, s) = 1$ then $f(z) = f(z + 1/s)$ generates both the periods and so f only has one period.

Hence we may assume that $\tau \notin \mathbb{R}$. In general we define

$$\text{The lattice } \Lambda = \Lambda_\tau := \{m + n\tau : \ m, n \in \mathbb{Z}\} \text{ denoted } \langle 1, \tau \rangle;$$

$$\text{and the parallellogram } \ P_\tau := \{x + y\tau : \ 0 \leq x < 1, \ 0 \leq y < 1\}.$$

`Exercise 14.1.2.` (i) Prove that $\Lambda = -\Lambda$. (ii) Prove that there exists a constant c_τ such that $\#\{w \in \Lambda : |w| \leq x\} \sim c_\tau x^2$. (iii) Prove that for any $z \in \mathbb{C}$ there exists a unique $y \in P_\tau$ such $z - y \in \Lambda$ and hence $f(z) = f(y)$.

One can think of this as f acting on $\mathbb{C}$ modulo the lattice Λ ("$\mathbb{C}$ mod Λ"), and the parallellogram P_τ provides a set of representatives for $\mathbb{C}/\Lambda$. Indeed any such parallelogram is called a *fundamental parallellogram* for $\mathbb{C}/\Lambda$. We also define C_τ to be the boundary of P_τ, that is the contour going along the lines from 0 to 1, then from 1 to $1 + \tau$, then from $1 + \tau$ to τ, and finally from τ back to 0.

We define the *Eisenstein series* of weight k to be

$$G_k(\Lambda) := \sum_{\substack{w \in \Lambda \\ w \neq 0}} \frac{1}{w^k} \quad \text{for each integer} \quad k \geq 3.$$

`Exercise` 14.1.3. Prove that (i) $G_k(\Lambda) = 0$ when k is odd; and (ii) the sum defining $G_k(\Lambda)$ converges for all $k \geq 3$. (Hint: Use exercise 14.1.2.)

The key to understanding periodic functions on Λ is the *Weierstrass* $\wp$-function:

$$\wp(z) = \wp_\Lambda(z) := \frac{1}{z^2} + \sum_{\substack{w \in \Lambda \\ w \neq 0}} \left\{ \frac{1}{(w - z)^2} - \frac{1}{w^2} \right\},$$

Exercise 14.1.4. Prove that (i) $\wp(z) = \wp(-z)$; and (ii) the series defining $\wp(z)$ converges at every $z \notin \Lambda$ (keeping together the two terms in each parentheses).

Now $w^{-2}((1-z/w)^{-2}-1) = \sum_{j\geq 1}(j+1)z^j/w^{j+2}$ and therefore

$$\wp(z) = \frac{1}{z^2} + \sum_{j\geq 1}(j+1)G_{j+2}(\Lambda)z^j = \frac{1}{z^2} + \sum_{k\geq 2}(2k-1)G_{2k}(\Lambda)z^{2k-2}.$$

The derivative is a lot easier to work with, in that proof of convergence is straightforward:

$$\wp'(z) = \sum_{w\in\Lambda} \frac{2}{(w-z)^3} = -\frac{2}{z^3} + \sum_{k\geq 2}(2k-1)(2k-2)G_{2k}(\Lambda)z^{2k-3}.$$

It is evident that $\wp'(z+w) = \wp'(z)$, which one proves simply by shuffling the relevant terms; note that this proof does not work (easily) for $\wp(z)$ because of convergence issues. However, since $\wp'(z+w) = \wp'(z)$, so $\wp(z+w) = \wp(z) + c_w$ for some constant c_w, by the fundamental theorem of calculus. Now $c_w = \wp(w/2) - \wp(-w/2) = 0$ since $w \in \Omega$ and as $\wp$ is an even function, and therefore $\wp(z+w) = \wp(z)$.

Exercise 14.1.5. Show that (i) $\wp^{(k)}(z)$ is Λ-periodic and (ii) $\wp^{(k)}(-z) = (-1)^k\wp^{(k)}(z)$ for all $k \geq 1$. (iii) Deduce that if $2v \in \Omega$ but $v \notin \Omega$ then $\wp^{(k)}(v) = 0$ for all odd k.

We now show that any non-constant Λ-periodic function, $f(z)$, must have at least one pole inside Λ. For, if not, then $f(z)$ is analytic in Λ, and since Λ is compact we know that $f(z)$ is bounded in Λ. By periodicity we then know that $f(z)$ is bounded and analytic on all of $\mathbb{C}$ and so must be a constant function.

We need to do a bit of integration using the residue theorem. Suppose that $f(z)$ is a non-constant, meromorphic, Λ-periodic function and select $v \in P_\tau$ such that $v + C_\tau$ does not go through a pole of f, which is possible as f only has finitely many poles in a bounded region like $P_\tau + C_\tau$. Now $\int_{v+C_\tau} f(z)dz = 0$, because as we integrate along opposite sides of $v + C_\tau$ we obtain cancelation due to the periodicity of f; for example $\int_{x=0}^{1} f(v+x)dx + \int_{x=1}^{0} f(v+\tau+x)dx = 0$. By the residue theorem we deduce that the sum of the residues at the poles of f, inside the contour $v+C_\tau$, equals 0. Note that f must then have at least two poles (including multiplicity), else the sum of the residues of the poles is non-zero. Replacing f by f'/f we deduce that the number of poles of f inside $v + C_\tau$, equals the number of zeros (all including multiplicity). Taking the example $f = \wp(z) - c$ and selecting v close to $-(1+\tau)/2$, we only have the one pole at $z = 0$, which occurs with multiplicity 2, and so there are exactly two zeros of $\wp(z) - c = 0$ in P_τ, for every $c \in \mathbb{C}$. Since $\wp(z)$ is even we know that if $\wp(z) = c$ then $\wp(-z) = c$, which gives us our two zeros, except in the case that $z = -z + w$ for some $w \in \Lambda$, in other words $2z \in \Lambda$. Looking at the parallellogram P_τ, we see that the set of such points is $\{0, 1/2, \tau/2, (1+\tau)/2\}$.

In analogy to the theorem, above, about all real periodic functions being rational functions of $e^{2i\pi x}$ we now prove that every meromorphic, Λ-periodic function $f(z)$ is a rational function of $\wp(z)$ and $\wp'(z)$: The functions $F(z) = f(z) + f(-z)$ and $G(z) = (f(z) - f(-z))/\wp'(z)$ are both even, meromorphic, Λ-periodic functions. We will show that any such function is a rational function in $\wp(z)$ and hence the result follows for

$f(z) = (F(z) + \wp'(z)G(z))/2$. Now let us suppose that $F(v) = 0$. Then $F(-v) = 0$ as F is even. Indeed suppose that the zeros of F inside P_τ are $v_1, -v_1, \dots, v_m, -v_m$ (including multiplicities) and the poles at $y_1, -y_1, \dots, y_n, -y_n$. Then the function

$$H(z) := F(z) \cdot \frac{\prod_{j=1}^{n}(\wp(z) - \wp(y_j))}{\prod_{i=1}^{m}(\wp(z) - \wp(v_i))}$$

can only have poles or zeros at 0, inside P_τ. From what saw before the number of zeros of H must equal the number of poles, but since 0 is the only candidate, it can be neither. Therefore $H(z)$ is an analytic function on $\mathbb{C}$, which is therefore bounded (since it is periodic and bounded in P_τ) and thus a constant (since the only bounded analytic functions on $\mathbb{C}$ are the constants). The result follows.

We now show that if $f(z)$ is a non-zero, meromorphic, Λ-periodic function, with zeros $z_1, \dots, z_k$ and poles $y_1, \dots, y_\ell$ (all counted with multiplicity), then $\sum_i z_i - \sum_j y_j \in \Lambda$. Note that we already know that this is true for any $\wp(z) - c$, and for any $r\wp'(z) + s\wp(z) + t$. We shall prove it by induction on the number of zeros and poles of f. If f has none then the result is trivial. As we saw above f must have at least two poles (including multiplicities). We may assume that f only has poles at 0, else we multiply $f(z)$ through by $\prod_j(\wp(z) - \wp(y_j))$ since, as we have seen, the sum of the zeros of this is $\sum_j y_j + (-y_j) = 0$ mod Λ. Since f has at least a double pole at 0, it must have two (or more) zeros z_1, z_2, so we compute the equation of line $ry + sx + t$ that goes between P_{z_1} and P_{z_2}. We saw above that the Λ-periodic function $g(z) := r\wp'(z) + s\wp(z) + t$ has a triple pole at 0, and then its three zeros are at z_1, z_2 and $-z_1 - z_2$ mod Λ. Replacing $f(z)$ by $f(z)(\wp(z) - \wp(z_1 + z_2))/g(z)$ we have a new such function with one less zero, and no new poles, and so the result follows by induction. This result is more surprising than it might seem at first sight, for if we select $v \in P_\tau$ such that $v + C_\tau$ does not go through a pole of f, then

$$\int_{v+C_\tau} \frac{zf'(z)}{f(z)}\, dz = \sum_i z_i - \sum_j y_j \in \Lambda.$$

Note that zf'/f *is not* periodic. However if we integrate along opposite sides and compare, we obtain, after some cancelation

$$\int_v^{v+\tau} \frac{f'(z)}{f(z)}\, dz - \tau \int_v^{v+1} \frac{f'(z)}{f(z)}\, dz \in \Lambda.$$

14.2. Parametrizing elliptic curves. Now $\wp'(z)^2$ is an even Λ-periodic function and so must be a rational function of $\wp(z)$. The rational function must in fact be a polynomial, for if $\wp(z) - c$ is a factor of the denominator then its zeros would be poles of $\wp'(z)$; however the only poles of $\wp'(z)$ are the elements of Λ which are poles, not zeros, of $\wp(z) - c$. But then $\wp'(z)^2$ must equal a polynomial of degree 3 in $\wp(z)$, since 0 is a pole of order 6 of $\wp'(z)^2$, and a pole of order $2d$ of any polynomial of degree d in $\wp(z)$. Comparing the coefficients of $1/z^6$ in the Fourier expansions of $\wp'(z)^2$ and $\wp(z)^3$ we deduce that

$\wp'(z)^2 = 4\wp(z)^3 + a\wp(z)^2 + b\wp(z) + c$ for some constants a, b, c (which may depend on Λ). Now the Laurent expansion of $\wp'(z)^2 - 4\wp(z)^3$ at $z = 0$ is

$$\begin{aligned}
&= (-2z^{-3} + 6G_4 z + 20G_6 z^3 + O(z^5))^2 - 4(z^{-2} + 3G_4 z^2 + 5G_6 z^4 + O(z^6))^3 \\
&= (4z^{-6} - 24G_4 z^{-2} - 80G_6) - 4(z^{-6} + 9G_4 z^{-2} + 15G_6) + O(z^2) \\
&= -60G_4 z^{-2} - 140G_6 + O(z^2) = -60G_4(\Lambda)\wp(z) - 140G_6(\Lambda) + O(z^2).
\end{aligned}$$

Hence $\wp'(z)^2 - 4\wp(z)^3 + 60G_4(\Lambda)\wp(z) + 140G_6(\Lambda)$ is a Λ-periodic function which equals 0 at $z = 0$, and so has no poles. Therefore it equals 0, and we have proved that

$$\wp'(z)^2 = 4\wp(z)^3 - 60G_4(\Lambda)\wp(z) - 140G_6(\Lambda). \tag{14.1}$$

Consider the elliptic curve

$$E: \ y^2 = 4x^3 + ax + b \text{ where } a = -60G_4(\Lambda) \text{ and } b = -140G_6(\Lambda).$$

From the last paragraph

$$P_z := (\wp(z), \wp'(z)) \in E(\mathbb{C}) \text{ for every } z \in \mathbb{C}.$$

Now, for every $x \in \mathbb{C}$ we have seen that there are exactly two values $u \in \mathbb{C}/\Lambda$, that is u and $-u$, for which $\wp(u) = \wp(-u) = x$. By the previous paragraph we know that the points $P_u = (\wp(u), \wp'(u)), P_{-u} = (\wp(-u), \wp'(-u)) = (\wp(u), -\wp'(u)) = -P_u$ lie on the curve and give rise to the two points (x, y) and $(x, -y)$. This almost yields a 1-1 correspondence between $\mathbb{C}/\Lambda$ and $E(\mathbb{C})$; the only question arises when $2u \in \Lambda$. If $u = 0$ then $(\wp(u), \wp'(u))$ yields the point at infinity. If $u \neq 0$ then $\wp'(u) = 0$ and these three u-values correspond to the three points $(x, 0)$ of order two. Hence we have established a 1-1 correspondence, and we ask whether the bijection $\wp : \mathbb{C}/\Lambda \to E(\mathbb{C})$ extends to give an isomorphism of groups? In other words, is it true that if $u, v, w \in \mathbb{C}$ with $u + v + w = 0$ then $P_u + P_v + P_w = 0$ in $E(\mathbb{C})$. Certainly this works if $w = 0$ since we have seen that $P_u + P_{-u} + P_0 = 0$.

We will show that for any $v \in \mathbb{C} \setminus \Lambda$, the Λ-periodic function

$$F_v(u) := \wp(u + v) + \wp(u) + \wp(v) - \frac{1}{4}\left(\frac{\wp'(u) - \wp'(v)}{\wp(u) - \wp(v)}\right)^2$$

equals 0. The poles of this function can only lie at the poles of $\wp(u + v)$ or $\wp(u)$ or $\wp'(u) - \wp'(v)$, or the zeros of $\wp(u) - \wp(v)$; that is when $u \in v + \Lambda$ or $u \in \Lambda$ or $u \in -v + \Lambda$. Now if $u \in v + \Lambda$ where $2v \notin \Lambda$ (in this case $u \in -v + \Lambda$ which we consider later), then $\wp(u) - \wp(v)$ has a simple zero, and clearly $\wp'(u) - \wp'(v)$ also has a zero and so the zeros cancel. The next case is when $u \in \Lambda$, so we need only study the case $u = 0$. Using the Taylor series for $\wp(u)$ and $\wp'(u)$ we obtain, for small u,

$$F_v(u) = \wp(v) + O(u) + \frac{1}{u^2} + O(u) + \wp(v) - \frac{1}{4}\left(\frac{-2/u^3 + O(1)}{1/u^2 - \wp(v) + O(u)}\right)^2 = O(u),$$

and so $F_v(u)$ has no pole for $u \in \Lambda$; in fact $F_v(0) = 0$. This leaves only the case $u \in -v+\Lambda$. Now, the Taylor series around $u = -v$ yields

$$\begin{aligned}\wp(u) &= \wp(-v) + (u+v)\wp'(-v) + \frac{(u+v)^2}{2}\wp''(-v) + O((u+v)^3)\\ &= \wp(v) - (u+v)\wp'(v) + \frac{(u+v)^2}{2}\wp''(v) + O((u+v)^3)\end{aligned}$$

and $\wp'(u) = \wp'(-v) + (u+v)\wp''(-v) + O((u+v)^2) = -\wp'(v) + (u+v)\wp''(v) + O((u+v)^2)$, so that

$$\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)} = \frac{-2\wp'(v) + (u+v)\wp''(v) + O((u+v)^2)}{-(u+v)\wp'(v) + \frac{(u+v)^2}{2}\wp''(v) + O((u+v)^3)} = \frac{2}{u+v} + O(u+v),$$

if $\wp'(v) \neq 0$.[28] If $\wp'(v) = 0$ then $2v \in \Omega$ but $v \notin \Omega$, and so

$$\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)} = \frac{(u+v)\wp''(v) + O((u+v)^3)}{\frac{(u+v)^2}{2}\wp''(v) + O((u+v)^4)} = \frac{2}{u+v} + O(u+v),$$

by exercise 14.1.5. This implies that $F_v(u) = 1/(u+v)^2 - 1/(u+v)^2 + O(1) = O(1)$ when u is close to $-v$. We have therefore shown that $F_v(u)$ is an anlytic, Λ-periodic function with $F_v(0) = 0$ and therefore $F_v(u) = 0$, as claimed. Therefore we have proved that

$$\wp(u+v) + \wp(u) + \wp(v) = \frac{1}{4}\left(\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)}\right)^2$$

and so the third point of $E(\mathbb{C})$ on the line joining P_u and P_v is $P_{\pm(u+v)}$. To determine the sign, we use the equation of the line so that

$$\wp'(\pm(u+v)) = \left(\frac{\wp'(u)-\wp'(v)}{\wp(u)-\wp(v)}\right)(\wp(u+v) - \wp(v)) + \wp'(v).$$

Expanding this around $u = 0$ reveals that

$$\pm\wp'(v) + O(u) = \left(\frac{-2/u^3 - \wp'(v)}{1/u^2 - \wp(v)}\right)(1 + O(u^4))(u\wp'(v) + O(u^2)) + \wp'(v) = -\wp'(v) + O(u).$$

We have thus established that P_u, P_v and $P_{-(u+v)}$ lie on a line, and therefore

$$P_{u+v} = P_u + P_v.$$

In other words

[28] This is surprising, in that we might expect an error term of $O(1)$ here. This works since the first two coefficients of the numerator and denominator are in the same ratio.

Theorem 14.1. *The map $\wp : (\mathbb{C}/\Lambda, +) \to (E(\mathbb{C}), +)$ is an isomorphism of groups.*

`Exercise` 14.2.1. Explain how this yields a proof of the associative law for addition of points on $E(\mathbb{C})$.

It is difficult to get a good intuitive idea of addition of points in $E(\mathbb{C})$ because the addition law is hard to think through. However we are all used to addition in $\mathbb{C}$, so various questions about the structure of $E(\mathbb{C})$ are often most easily studied in this isomorphic context. For example, torsion points, as we see in the next subsection. It should be stressed that although $\wp$ provides an isomorphism, it provides few hints as to the structure of $\wp^{-1}(E(\mathbb{Q}))$, since this map does not preserve rationality. However it can help us understand the distribution of points on $E(\mathbb{R})$. (*Exercise to be added.*)

Strange identities. We established (14.1) by studying the Taylor expansion up to the z^0 term. Since (14.1) holds, this means that all subsequent coefficients of the difference of the two sides must equal 0. A quick calculation reveals[29] that the difference is

$$(108G_4^2 - 252G_8)z^2 + (180G_4G_6 - 396G_{10})z^4 + (100G_6^2 + 420G_4G_8 - 108G_4^3 - 572G_{12})z^6 + \dots$$

where $G_j = G_j(\Lambda)$. We therefore deduce that

$$G_8 = \frac{3}{7}\, G_4^2, \quad G_{10} = \frac{5}{11}\, G_4G_6, \text{ and } \quad G_{12} = (18G_4^3 + 25G_6^2)/143.$$

In fact we can proceed like this to show that $G_{2k}(\Lambda)$ can be expressed as a polynomial in G_4 and G_6 for all $k \geq 2$, and that the terms are of the form $c_{i,j}G_4^iG_6^j$ where $4i + 6j = 2k$. These surprising formulas will be developed by a different, easier method in section 15.4.

14.3. Torsion points in $\mathbb{C}$. Can we describe all the points $P \in E(\mathbb{C})$ for which $NP = 0$? Now if $P = P_u$ then $\wp(Nu) = N\wp(u) = NP_u = 0 = P_0 = \wp(0)$ if and only if $Nu \in \Lambda$, that is $u = \frac{1}{N}\,\Lambda$. The quotient $\frac{1}{N}\,\Lambda/\Lambda$ has a set of representatives $\{\frac{1}{N}\,(i+j\tau) : 0 \leq i, j \leq N-1\}$, so the points can be thought of as $\frac{1}{N}\,(u + v\tau)$ where we take u and v mod N. Therefore, the subgroup $E[N](\mathbb{C})$ of points of $E(\mathbb{C})$ of order N has the structure

$$E[N](\mathbb{C}) \cong \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \tag{14.2}$$

for every $N \geq 1$.

We have seen how, given a point $P = (x, y)$, we can explicitly determine the coordinates of $2P$. In general:

$$[m]P = \left(\frac{\phi_m(P)}{\psi_m(P)^2}, \frac{\omega_m(P)}{\psi_m(P)^3}\right)$$

where the *division polynomials* are given by $\psi_1 = 1$, $\psi_2 = 2y$, $\psi_3 = 3x^4 + 6ax^2 + 12bx - a^2$ and $\psi_4 = 4y(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3$, and then one determines them

[29] Preferably using Maple than by hand!

by

$$\psi_{2m+1} = \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3 \text{ for each } m \geq 2,$$
$$\omega_m = (\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2)/4y, \text{and}$$
$$\psi_{2m} = 2\omega_m\psi_m \text{ for each } m \geq 3, \text{ and then}$$
$$\phi_m(x) = x\psi_m(x)^2 - \psi_{m-1}(x)\psi_{m+1}(x) = x^{m^2} + \dots$$

One can prove this by induction, and various properties:

- We have $\psi_n(x), \phi_{2m+1}(x) \in \mathbb{Z}[a, b, x]$ and $w_n(x), \phi_{2m}(x) \in y\mathbb{Z}[a, b, x]$.
- $\psi_n(x)^2 = n^2x^{n^2-1} + cx^{n^2-2} + \dots$ and $\phi_n(x) = x^{n^2} + \dots$ More precisely
$$\psi_n(x) = nx^{(n^2-1)/2} + \dots, \ w_n(x) = yx^{(3n^2-3)/2} + \dots \text{ for } n \text{ odd};$$
$$\psi_n(x) = nyx^{(n^2-4)/2} + \dots, \ w_n(x) = x^{3n^2/2} + \dots \text{ for } n \text{ even}.$$

`Exercise` 14.3.1. Prove that $(\psi_m(x), \psi_{m-1}(x)) = 1$ for all $m \geq 1$, by induction on m.

Another proof uses the $\wp$-function. By Theorem 14.1 we know that if $P = P_z$ then $NP = P_{Nz} = (\wp(Nz), \wp'(Nz))$. Now $\wp(Nz)$ is an even, meromorphic, Λ-periodic function and so is a rational function of $\wp(z)$. Now $\wp(Nz) = \infty$ if and only if $z \in \{\frac{1}{N}(i + j\tau) : 0 \leq i, j \leq N - 1\}$; in other words if $\wp(Nz) = f_N(\wp(z))/g_N(\wp(z))$ then

$$g_N(x) = \prod_{\substack{0 \leq i,j \leq N-1 \\ (i,j) \neq (0,0)}} \left(x - \wp\left(\frac{1}{N}(i + j\tau) \right) \right).$$

Now, since $\wp$ is a homomorphism, the number of solutions z of $\wp(Nz) = c$ counting multiplicities, is the same as the number of solutions of $\wp(Nz) = \infty$, that is N^2, and therefore $f_N = \phi_N$ had degree N^2. For every root $\wp(z)$ of f_N we also have the root $\wp(-z) = \wp(z)$. Now z and $-z$ are distinct unless $2z \in \Lambda$. Hence we can write $g_N(z) = \kappa_N\psi_N(z)^2$ for some polynomial $\psi_N(z)$ if N is odd, and $g_N(z) = \kappa_N g_2(z)\psi_N(z)^2$ if N is even, where $g_2(z) = \left(x - \wp\left(\frac{1}{2}\right)\right)\left(x - \wp\left(\frac{\tau}{2}\right)\right)\left(x - \wp\left(\frac{1+\tau}{2}\right)\right) = x^3 + ax + b = y^2$, and κ_N is some constant.

`Exercise` 14.3.2. (i) Prove that $\psi_m(x)$ divides $\psi_{md}(x)$ for all $d \geq 1$. (ii) Prove that $(\psi_m(x), \psi_n(x)) = \psi_{(m,n)}(x)$ for all $m, n \geq 1$. (Hint: Use the last approach.)

We saw above that $E[n](\mathbb{C}) \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ and so has n^2 elements. The proof used the $\wp$-function so gave us little sense of what these points look like, or how to find them. Using the division polynomials $\phi_n(x)$ one can get a better sense of what they look like: Exercise 14.3.1 implies that $(\phi_m(x), \psi_m(x)^2) = (\psi_{m-1}(x)\psi_{m+1}(x), \psi_m(x)^2) = 1$. Therefore

$$[m]P = \mathcal{O} \text{ if and only if } \psi_m(x)^2 = 0 \text{ or } P = \mathcal{O}.$$

Now for any given $x \in \mathbb{C}$ there are at most two points $(x, y) \in E(\mathbb{C})$. If n is odd then $\psi_n(x)$ has at most $(n^2 - 1)/2$ distinct roots, and so these correspond to at most $n^2 - 1$ roots. However by the above display we know that $\psi_n(x)$ must correspond to exactly $n^2 - 1$ roots, and hence the roots of $\psi_n(x)$ are distinct. Similarly, if n is even, $\psi_n(x)/y$ has at most $(n^2 - 4)/2$ distinct roots, and the polynomial $(\psi_n(x)/y)^2$ corresponds to all points

for which $[n]P = 0$ except the four of order 1 or 2. Hence, the roots of $\psi_n(x)$ are distinct. Thus we have proved that the roots of $\psi_n(x)$ each give rise to exactly 2 n-torsion points, and indeed all of them of order > 2.

Let ℓ be any prime. If $\Lambda = \langle 1, \tau \rangle$ then the ℓ^k-torsion points are exactly the points of the form $\frac{i+j\tau}{\ell^k}$ for $0 \le i, j < \ell^k$. Notice that this forms a finer and finer mesh, as k gets larger, covering the fundamental parallellogram, so that if f is any continuous function of $\mathbb{C}/\Lambda$ then its values can be understood in terms of the values in this set as $k \to \infty$. Thus the *Tate module* is given by

$$T_\ell[E] = \lim_{k\to\infty} E[\ell^k], \text{ which is written } \lim_{k\leftarrow} E[\ell^k],$$

since evidently we have $E[\ell] \subset E[\ell^2] \subset \dots$ Now we have seen that, over $\mathbb{C}$ we have $E[\ell^k] \cong \mathbb{Z}/\ell^k\mathbb{Z} \times \mathbb{Z}/\ell^k\mathbb{Z}$, and so, taking limits,

$$T_\ell[E] \cong \mathbb{Z}_\ell \times \mathbb{Z}_\ell,$$

where $\mathbb{Z}_\ell$ is all expressions $\frac{a_1}{\ell} + \frac{a_2}{\ell^2} + \frac{a_3}{\ell^3} + \dots$, where each $0 \le a_i \le \ell - 1$, and we can perform addition and multiplication mod 1.

14.4. The converse theorem. We have shown that every lattice Λ may be associated to an elliptic curve E via the Weierstrass parametrization. It turns out that the converse is true: For every elliptic curve E there is a lattice Λ which may be associated to the elliptic curve E — we will only prove this later when we discuss the j-function.

The bijection works as follows: Two lattices Λ, Λ' are homothetic (as in section 4.4) if there exists a complex number λ such that $\Lambda' = \lambda\Lambda$. Two elliptic curves $E:\ y^2 = x^3+ax+b$ and $E':\ y^2 = x^3 + a'x + b'$ are equivalent if the ratios $a^3 : b^2$ and $(a')^3 : (b')^2$ are equal. There is a 1-1 correspondence between equivalence classes of lattices and equivalence classes of elliptic curves given (in one direction) by the Weierstrass $\wp$-function. Going in the other direction, that is finding a lattice given an elliptic curve, involves Riemann surface theory, which lies a little deeper than we wish to go. However we will sketch that part of the theory that is accessible: If we are given $x = \wp(z)$ then we have

$$\left(\frac{dx}{dz}\right)^2 = 4x^3 - ax - b$$

so that

$$dz = \frac{dx}{\sqrt{4x^3 - ax - b}}$$

and therefore, we get the *elliptic integral*

$$z(x) - z_0 = \int_{x_0}^{x} \frac{dx}{\sqrt{4x^3 - ax - b}}$$

which needs some interpreting. In the case that $4x^3-ax-b$ has three real roots $e_1 < e_2 < e_3$ then we can select the periods to be

$$w_1 := 2i \int_{-\infty}^{e_1} \frac{dx}{\sqrt{b + ax - 4x^3}} \quad \text{and} \quad w_2 := 2 \int_{e_3}^{\infty} \frac{dx}{\sqrt{4x^3 - ax - b}}$$

where we take the square root to be positive and simply integrate along the x-axis. That is $\Lambda' = \langle w_1, w_2\rangle$ which is homothetic to $\Lambda = \langle 1, \tau\rangle$ where $\tau = w_1/w_2$.

One can evaluate such integrals by re-writing the elliptic curve as $y^2 = 4x(x-1)(x-\lambda)$. Indeed one can show that

$$\frac{1}{\pi}\int_0^\lambda \frac{dx}{\sqrt{x(x-1)(x-\lambda)}} = \sum_{j\geq 0}\binom{-\frac{1}{2}}{j}^2\lambda^j, \tag{14.3}$$

which should be compared to the formula for $\#\{(x,y):\ y^2 \equiv x(x-1)(x-\lambda) \pmod p\}$ given at the end of section 9.5.

`Exercise 14.4.1.` We denote this integral by $\sigma(\lambda)$ and prove (14.3).

i. Show that it satisfies the *Picard-Fuchs* equation

$$\lambda(\lambda-1)\frac{d^2\sigma}{d\lambda^2} + (2\lambda-1)\frac{d\sigma}{d\lambda} + \frac{\sigma}{4} = 0,$$

with $\sigma = 0$, and then derive (14.3).

ii. Alternatively, establish that

$$\frac{1}{\pi}\int_0^1 \frac{y^j}{\sqrt{y(1-y)}}\,dy = \frac{2}{\pi}\int_0^{\pi/2}\sin^{2j}\theta\,d\theta = \frac{1}{4^j}\binom{2j}{j} = (-1)^j\binom{-\frac{1}{2}}{j}.$$

Then derive (14.2) by substituting $x = \lambda y$ into the definition of $\sigma(\lambda)$, and using the binomial theorem to expand $(1-\lambda y)^{-1/2}$.

14.5. Maps between elliptic curves over $\mathbb{C}$. Complex multiplication. We are interested in meromorphic homomorphisms $\phi : E \to E'$ between elliptic curves (and thus if ϕ is a homomorphism it must preserve the group structure). Suppose that E and E' are parametrized by Λ and Λ' respectively. Now

$$(\phi\circ\wp_\Lambda)(z) + (\phi\circ\wp_\Lambda)(-z) = \phi(P_z + P_{-z}) = \phi(\mathcal{O}_E) = \mathcal{O}_{E'},$$

and so $\rho(z) = \rho(-z)$ where $\rho := x\circ\phi\circ\wp_\Lambda :\ \mathbb{C}/\Lambda \to E'$ (and $x(.)$ is the map that takes the x-coordinate). Hence the map ρ is an even, meromorphic Λ-periodic function, and thus a rational function of $\wp(z)$ (as we have seen above).

We define $\psi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ by $\psi = \wp_{\Lambda'}^{-1}\circ\phi\circ\wp_\Lambda$, which is therefore an analytic homomorphism between lattices. Note that $\phi\circ\wp_\Lambda = \wp_{\Lambda'}\circ\psi$ by definition, and we have $\ker\psi = \ker\rho$, and $\wp_\Lambda(\ker\psi) = \ker(\phi)$. We already know two examples of such maps:

• For any integer N, we have the map $z \to Nz$ on $\mathbb{C}/\Lambda$ for given $N \in \mathbb{Z}$, which defines the map $P \to NP$ on E that we have already studied.

• For fixed $u \in \mathbb{C}$ the map $z \to z+u$ on $\mathbb{C}/\Lambda$ defines the map $P \to P + P_u$ on E.

We wish to determine all such maps. If we fix any $w \in \Lambda$ then we want $\psi(z+w) = \psi(z) \pmod{\Lambda'}$ for all $z \in \mathbb{C}$. However $\psi(z+w) - \psi(z)$ is a continuous function of z so it cannot jump from one point of the lattice Λ' to another, and therefore it must be constant. But then $\psi'(z+w) - \psi'(z) = 0$ for all $w \in \Lambda$, and therefore $\psi'(z)$ is a Λ-periodic analytic function. The only such function is a constant, and hence $\psi(z) = mz + s$ for some $m, s \in \mathbb{C}$. But then $\psi(0) = s \in \Lambda'$, so we can reduce to studying the maps

$$\psi(z) = mz.$$

Now if $w \in \Lambda$ then we must have $mw = \psi(w) \in \Lambda'$; that is

$$m\Lambda \subset \Lambda', \tag{14.5.1}$$

and evidently this is an "if and only if" condition. We deduce that

$$\ker \psi = \{z \in \mathbb{C}/\Lambda : \ mz \in \Lambda'\},$$

which is evidently finite, and therefore a finite subgroup of $\mathbb{C}/\Lambda$. Now $\rho(z) = \wp_{\Lambda'}(mz)$ as $\phi \circ \wp_\Lambda = \wp_{\Lambda'} \circ \psi$. Since $0 \in \ker \psi$, we deduce that for any z_0,

$$\rho(z) = \kappa(\wp_\Lambda(z) - \wp_\Lambda(z_0)) \cdot \frac{h_m(\wp_\Lambda(z) - \wp_\Lambda(z_0))}{h_m(\wp_\Lambda(z))} + \wp_{\Lambda'}(mz_0)$$
$$\text{where } \ h_m(x) := \prod_{\substack{z \in \ker \psi \\ z \neq 0}} (x - \wp_\Lambda(z)),$$

for some constant κ.[30] This is much like what we saw earlier when m was an integer, but this is more general as it works for any complex number m. Henceforth we will assume that m is not real (and so we call this *complex multiplication*), and we define $\overline{\psi}(z) = \overline{m}z$. Note that $(\overline{\psi} \circ \psi)(z) = \overline{m}mz = Nz$ where $N =$Norm$(m) := \overline{m}m$, and so we have $\overline{\psi} : C/\Lambda' \to C/\Lambda$. Hence $\overline{\psi}$ is the *dual* map to ψ, and returns us to the original lattice, so we see that (14.5.1) implies that $\overline{m}\Lambda' \subset \Lambda$ (and therefore (14.5.1) defines an equivalence relation). We also deduce that $\ker \psi \subset \ker[N] = \{\frac{i+j\tau}{N} : \ 0 \le i, j \le N-1\}$ where $\Lambda := \langle 1, \tau \rangle$.

One can also determine our map by its action on the kernel, and then on one other point, as in our formula for ρ above, or by using the Tate module (see below). If the kernel has composite order then one can decompose the map into two parts, killing one part of the kernel, and then the remainder. However this does not always provide a complex multiplication type homomorphism for if it did then we would be able to always factor such m into two other integers of smaller norm in the appropriate field, and so the class number would always be one, which is not true.

If m is an algebraic integer and rational prime $p|m$ then we can write $m = pM$ where M is an algebraic integer, and our map can be decomposed as $z \to pz$ (which is a map $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$) followed by $z \to Mz$ (which is $\mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$). We understand the first map so we restrict our attention to the case that m is an algebraic integer that is not divisible by any rational prime. Writing $\Lambda = \langle 1, \tau \rangle$ and $\Lambda' = \langle 1, \tau' \rangle$ we have $m\Lambda \subset \Lambda'$ if and only if there exist integers a, b, c, d such that $m = a + b\tau'$ and $m\tau = c + d\tau'$. Now $(a, b) = 1$ since m is not divisible by any rational prime, and $ad - bc \neq 0$ else $\tau = \frac{c+d\tau'}{a+b\tau'} \in \mathbb{R}$, a contradiction. It is easiest to write this down explicitly as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ \tau' \end{pmatrix} = m \begin{pmatrix} 1 \\ \tau \end{pmatrix} \Longleftrightarrow (ad - bc) \begin{pmatrix} 1 \\ \tau' \end{pmatrix} = m \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} 1 \\ \tau \end{pmatrix}.$$

[30] One might object to this description of $h_m(x)$ in that one needs to determine the value of $\wp$ at various points. Actually only knowing that it is a degree N polynomial divided by a degree $N-1$ polynomial is enough, because we can use the Taylor expansion of $\wp(mz)$ to try to write it as a rational function of this degree in (the Taylor expansion) of $\wp(z)$, using the continued fraction expansion. Details of this may be found in [Sta].

Therefore $m\Lambda \subset \Lambda'$ if and only if $m^*\Lambda' \subset \Lambda$ where $m^*m = ad - bc$. Note that m^* is therefore an integer multiple of $\overline{m}$. Above we saw that $\ker\psi \subset \{\frac{i+j\tau}{N}\colon\ 0 \le i, j \le N-1\}$, and so

$$\psi\left(\frac{i+j\tau}{N}\right) = \frac{im+jm\tau}{N} = \frac{i(a+b\tau')+j(c+d\tau')}{N} = \frac{(ia+jc)+(ib+jd)\tau'}{N} \in \Lambda',$$

if and only if $\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} i \\ j \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod N$. We deduce that $|\ker\psi| = N$ by the following exercise:

`Exercise` 14.5.1.(i) Prove that if a, b, c, d are integers with $N := |ad-bc| \neq 0$ and $(a, b, c, d) = 1$ then there are exactly N pairs of residue classes $i, j \pmod N$ such that $\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} i \\ j \end{pmatrix} \equiv \begin{pmatrix} 0 \\ 0 \end{pmatrix} \pmod N$. (Hint: Use the Chinese Remainder Theorem.)

(ii) Show that these vectors $\begin{pmatrix} i \\ j \end{pmatrix} \pmod N$ form a subspace of the vectors in $(\mathbb{Z}/N\mathbb{Z})^2$, and can be decomposed, via the Chinese Remainder Theorem, into subspaces of $(\mathbb{Z}/p^k\mathbb{Z})^2$ for each $p^k \| N$.

(iii) Deduce that if m above is not divisible by a rational prime then there exists an integer k such that $\ker\psi = \{\frac{i(\tau+k)}{N}\colon\ 0 \le i \le N-1\}$ where $N = \text{Norm}(m)$.

(iv) Deduce that $\deg\rho = \text{Norm}(m)$.

We can replace τ by $\tau + k$ as a basis element for Λ. Hence, by a suitable change of basis, we may assume that $\ker\psi = \{\frac{i\tau}{N}\colon\ 0 \le i \le N-1\}$ and therefore $m, m\tau/N \in \Lambda'$. Now $m\tau/N = \tau/\overline{m}$, and so if Λ' belongs to the ring of integers of the field then $\overline{m}|\tau$. Let us suppose that $\tau = \sigma\overline{m}$, and so $\Lambda = \langle 1, \sigma\overline{m}\rangle$ whereas $\Lambda' \supseteq \langle m, \sigma\rangle$. Now $\tau = \frac{d\tau'+c}{b\tau'+a}$, hence $\mathbb{Q}(\tau') = \mathbb{Q}(\tau)$, and so we can take $\Lambda = \Lambda'$ (so as to create *endomorphisms*) if and only if $m, \tau/\overline{m} \in \mathbb{Z}[\tau]$.

If there are two maps $\psi_j : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ for $j = 1, 2$ where $m_1/m_2 \notin \mathbb{Q}$ then $\psi_2^* \circ \psi_1 : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$ is an endomorphism which is not multiplication by an integer. On the other hand if $\psi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$ is any endomorphism then $\psi_1 \circ \psi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$ yields a homomorphism. Thus the two sets of maps are in 1-to-1 correspondence, and so we can restrict our study to the set of endomorphisms.

So if $\psi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$ with $\psi(z) = mz$ then $\tau = \frac{d\tau+c}{b\tau+a}$, and so $b\tau^2 + (a-d)\tau + c = 0$, and so τ as an imaginary quadratic algebraic number.

`Exercise` 14.5.2. Prove that the set of m for which we have an endomorphism $\psi(z) = mz$ form a ring R. (Hint: Use the fact that $m\Lambda \subset \Lambda$.)

Now m is an eigenvalue of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $m^2 - (a+d)m + (ad-bc) = 0$. Hence m is an algebraic integer and $m \in \mathbb{Z}[\tau]$. As R is a ring, and $1 \in R$, we know that $a + b\tau \in R$ if and only if $b\tau \in R$. But then our matrix equation holds if and only if $a = 0$ and $b\tau^2 - d\tau - c = 0$. This last equation must therefore be a multiple of the minimum polynomial for τ. We have therefore proved the following:

Proposition 14.5.3. *Given τ, the set of analytic maps $\psi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda$ are all of the form $\tau \to m\tau + s$ where $m \in \mathbb{Z}$, unless $\tau = \frac{-b+\sqrt{d}}{2a}$ for some integers a, b, c with $(a, b, c) = 1$ and $d = b^2 - 4ac < 0$. In this case the endomorphism ring is given by*

$\{z \to mz + s : \ m \in \mathbb{Z}[a\tau]\}$. *We call this* complex multiplication by the order with conductor a^2d.

So let us consider the map $\psi(z) = (b + a\tau)z$. For a point P on our elliptic curve we have a value u such that $P_u = P$ and so we can extend ψ to a map ϕ_E on E as follows:

$$\phi_E(P_u) = \phi_E(\wp(u)) = \wp(\psi(u)) = \wp((b + a\tau)u) = P_{(b+a\tau)u},$$

we can see an underlying commutative diagram: $\phi_E \circ \wp = \wp \circ \psi$. The tricky bit is to understand what this means for the co-ordinates of E.

In a couple of cases it is easy enough:

• Consider the lattice $\Lambda := \langle 1, i\rangle$ where $i^2 = -1$, and the multiplication $z \to iz$ since $i\Lambda = \Lambda$. Therefore $\wp(iz) = -\wp(z)$ and $\wp'(iz) = i\wp'(z)$. Now if $y^2 = x^3 + ax + b$ then we also have $(iy)^2 = (-x)^3 + a(-x) + b$ and hence $b = 0$. That is $E : \ y^2 = x^3 + ax$ and the isogeny is the map $(x, y) \to (-x, iy)$.

• Consider the lattice $\Lambda := \langle 1, w\rangle$ where $w^2+w+1 = 0$, and the multiplication $z \to wz$ since $w\Lambda := \langle w, w^2\rangle = \langle w, 1\rangle$. Therefore $\wp(wz) = w\wp(z)$ and $\wp'(iz) = \wp'(z)$. This implies that $a = 0$, so that $E : \ y^2 = x^3 + b$ and the isogeny is the map $(x, y) \to (wx, y)$

Notice that neither of these maps are defined over $\mathbb{Z}$ but rather some field extension, which is typical.

Endomorphisms acting on the Tate module. We now show how an endomorphism ψ acts on $T_\ell[E]$: For each k, we have a basis $1/\ell^k, \tau/\ell^k$ for $E[\ell^k]$ and suppose that these are mapped to $(\alpha_k + \beta_k\tau)/\ell^k, (\gamma_k + \delta_k\tau)/\ell^k$, respectively. Note that

$$(\alpha_{k-1} + \beta_{k-1}\tau)/\ell^{k-1} = \psi(\ell/\ell^k) = \ell\psi(1/\ell^k) = (\alpha_k + \beta_k\tau)/\ell^{k-1}$$

and so $\alpha_k \equiv \alpha_{k-1} \pmod{\ell^{k-1}}$, and therefore

$$\alpha = \lim_{k\to\infty} \alpha_k \quad \text{exists in} \quad \mathbb{Z}_\ell.$$

Similarly remarks may be made about β, γ, δ. Hence we may write

$$\psi(u) = (\alpha + \beta\tau)u \ \text{ and } \ \psi(\tau v) = (\gamma + \delta\tau)v$$

so that

$$\psi\left(\begin{pmatrix} u & v\end{pmatrix}\begin{pmatrix}1\\ \tau\end{pmatrix}\right) = \begin{pmatrix} u & v\end{pmatrix}\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}\begin{pmatrix}1\\ \tau\end{pmatrix};$$

in other words we can represent the action $\psi : T_\ell[E] \to T_\ell[E]$ by a 2-by-2 matrix,

$$M_\psi := \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} \in \ \mathrm{SL}_2(\mathbb{Z}_\ell).$$

How big is the kernel of ψ? Now consider the dual isogeny $\hat\psi$. We know that $\psi \circ \hat\psi = N$ for some integer N. This means that $M_\psi M_{\hat\psi} = NI$ and so

$$\text{if } M_\psi = \begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix} \text{ then } M_{\hat\psi} = \begin{pmatrix}\delta & -\beta\\ -\gamma & \alpha\end{pmatrix}, \text{ and } N = \alpha\delta - \beta\gamma = \det M_\psi = \det M_{\hat\psi}.$$

Moreover $|\ker\psi| \cdot |\ker\hat\psi| = |\ker[N]| = N^2$, and so we can deduce that $|\ker\psi| = N = |\det M_\psi|$, as above.

In a later version of these notes we might discuss Galois representations a bit?

14.6. Homomorphisms between elliptic curves over arbitrary fields. Let K be a field of characteristic $\neq 2$. We just developed the theory of endomorphisms of elliptic curves using the Weierstrass parametrization. The problem with this is that it only works over fields of characteristic 0. So we now re-develop this theory without reference to the $\wp$-function. Again *isogenies* are group homomorphisms $\phi : E(K) \to E'(K)$ for some elliptic curves E and E' defined over K, that are defined by rational functions.

`Exercise` 14.6.1. Show (directly) that if $E:\ y^2 = x^3 + ax^2 + bx + c$ then any rational function of x, y on E can be written in the form $a(x) + b(x)y$ where $a(x), b(x) \in K(x)$.

By the previous exercise we may write $\phi(x, y) = (a(x) + b(x)y, c(x) + d(x)y)$ with $a(x), b(x), c(x), d(x) \in K(x)$. Now $\phi(P) + \phi(-P) = \phi(P + (-P)) = \phi(\mathcal{O}) = \mathcal{O}'$, and so

$$\begin{aligned}(a(x) + b(x)y, c(x) + d(x)y) = \phi(x, y) = -\phi(x, -y) &= -(a(x) - b(x)y, c(x) - d(x)y)\\ &= (a(x) - b(x)y, -c(x) + d(x)y)\end{aligned}$$

so that $b(x) = d(x) = 0$. Hence any such homomorphism can be written in the form

$$\phi((x, y)) = (a(x), d(x)y), \quad \text{where, } a(x), d(x) \in K(x).$$

Example: Let $E : y^2 = x^3 + ax^2 + bx$ and $E' : y^2 = x^3 - 2ax^2 + (a^2 - 4b)x$, and define $\phi : E \to E'$ by $\phi(x, y) = (y^2/x^2, y(x^2 - b)/x^2)$. Note that ϕ has degree 2, since if $\phi(x, y) = (X, Y)$ with $X \neq 0$ then $x^2 + (a - X)x + b = 0$ and $y = Yx^2/(x^2 - b)$.

If $E:\ y^2 = f(x)$ and $E':\ y^2 = g(x)$ then the existence of the map ϕ is equivalent to the identity $f(x)d(x)^2 = g(a(x))$; in other words $(a(x), d(x))$ in a point on $E''(K(x))$ where $E'':\ f(x)Y^2 = g(X)$, an elliptic curve defined over the function field $K(x)$.

An *endomorphism* is a group homomorphism $\phi : E(K) \to E(K)$ (that is $E' = E$); and an *algebraic endomorphism* is an isogeny with $E' = E$; confusingly we drop the word "algebraic" as is traditional, and so call this either an endomorphism or an isogeny (though it is both!).

`Exercise` 14.6.2. Prove that the endomorphisms $\phi : E(K) \to E(K)$ (i) are closed under addition; (ii) are closed under composition; (iii) form a ring.

Example: Let $E : y^2 = x^3 + bx$ and consider the map $\phi : E \to E$ given by $\phi(x, y) = (b/x, by/x^2)$ as well as the map $\psi(x, y) = (-x, iy)$.

`Exercise` 14.6.3. In this last example, what are the endomorphisms $\psi \circ \psi$, $\psi \circ \phi$ and $\psi + \phi$?

`Exercise` 14.6.4. Let $\phi : E(K) \to E(K)$ and define $\phi^{-1}(Q) := \{P \in E(\overline{K}) :\ \phi(P) = Q\}$ where $\overline{K}$ is the algebraic closure of K.

(i) Use the group structure of E to show that if $\phi(P) = Q$ then $\phi^{-1}(Q) = P + \ker\phi$.
(ii) Deduce that $|\phi^{-1}(Q)| = |\ker\phi|$.
(iii) Prove that if ϕ is defined by a separable rational function then $\deg\phi = |\ker\phi|$.

We define the *height* of ϕ to be $h(\phi) := \deg\phi$ which, as we have seen

`Exercise` 14.6.5. Show that if ϕ, ϕ_1, ϕ_2 are endomorphisms on $E(K)$ then (i) $h(\phi_1 \circ \phi_2) = h(\phi_1)h(\phi_2)$; (ii) $h(\phi + \phi) = 4h(\phi)$; (iii) $h(\phi) = 0$ if and only if $\phi(P) = \mathcal{O}$ for all $P \in E(K)$ (we denote this map by 0); (iv) Deduce that if $\psi \circ \phi = 0$ then either $\psi = 0$ or $\phi = 0$.

We can define an inner product

$$\langle \phi, \psi \rangle := \frac{1}{2}\,(h(\phi+\psi) - h(\phi) - h(\psi)).$$

Exercise 14.6.6. Prove that $\langle ., . \rangle$ is bilinear. (Hint: Remember how we approached the analogous question for $\hat{h}$.)

Hence $h(.)$ is a quadratic form; that is

$$h(x\phi^2 + y\phi + zI) = x^2 h(\phi^2) + y^2 h(\phi) + z^2 + 2xy\langle \phi^2, \phi\rangle + 2xz\langle \phi^2, I\rangle + 2yz\langle \phi, I\rangle. \tag{14.4}$$

where $\phi^2 := \phi \circ \phi$. Now by exercise 14.6.5(i), we see that

$$\langle \phi^2, \phi\rangle = \frac{1}{2}\,(h(\phi^2+\phi) - h(\phi^2) - h(\phi)) = h(\phi)\cdot\frac{1}{2}\,(h(\phi+I) - h(\phi) - h(I)) = h(\phi)\langle \phi, I\rangle.$$

Now, by (14.4), we have $h(\phi \pm I) = h(\phi) + 1 \pm 2\langle \phi, I\rangle$ so that

$$\begin{aligned}(h(\phi)+1)^2 + 4\langle \phi, I\rangle^2 &= \frac{1}{2}\,(h(\phi+I)^2 + h(\phi-I)^2) = \frac{1}{2}\,(h((\phi+I)^2) + h((\phi-I)^2))\\ &= \frac{1}{2}\,(h(\phi^2+2\phi+I) + h(\phi^2-2\phi+I))\\ &= h(\phi^2) + 4h(\phi) + 1 + 2\langle \phi^2, I\rangle = h(\phi^2+I) + 4h(\phi),\end{aligned}$$

using (14.4) several times.

Now let's factor $X^2 - 2\langle \phi, I\rangle X + h(\phi) = (X-\alpha)(X-\beta)$, say. Then $2\langle \phi, I\rangle = \alpha+\beta$ and $2\langle \phi^2, I\rangle = 4\langle \phi, I\rangle^2 - 2h(\phi) = (\alpha+\beta)^2 - 2\alpha\beta = \alpha^2+\beta^2$. Moreover $2\langle \phi^2, \phi\rangle = 2h(\phi)\langle \phi, I\rangle = \alpha\beta(\alpha+\beta)$. Substituting all this into (14.4) yields

$$\begin{aligned}h(x\phi^2 + y\phi + zI) &= x^2\alpha^2\beta^2 + y^2\alpha\beta + z^2 + xy\alpha\beta(\alpha+\beta) + xz(\alpha^2+\beta^2) + yz(\alpha+\beta)\\ &= (x\alpha^2 + y\alpha + z)(x\beta^2 + y\beta + z).\end{aligned}$$

Therefore $h(\phi^2 - 2\langle \phi, I\rangle\phi + h(\phi)I) = 0$,[31] and so $\phi^2(P) - 2\langle \phi, I\rangle\phi(P) + h(\phi)P = \mathcal{O}$, by exercise 14.6.5(iii). So we have proved that all endomorphisms satisfy a polynomial of degree two, with integer coefficients (since, by definition, $2\langle \phi, I\rangle \in \mathbb{Z}$). Also since h is positive-definite we know, taking $x = 0$ in (14.4), that

$$|\langle \phi, I\rangle| \le \sqrt{h(\phi)}.$$

Now this implies that the discriminant of $X^2 - 2\langle \phi, I\rangle X + h(\phi) = (X-\alpha)(X-\beta)$ is ≤ 0, so $\beta = \overline{\alpha}$, and therefore $|\alpha| = |\beta| = \sqrt{h(\phi)}$.

Exercise 14.6.7. Deduce that if $a\phi + bI = 0$ where a, b are integers then either $a = b = 0$, or a divides b and $h(\phi) = (b/a)^2$.

We write $t := \langle \phi, I\rangle \in \mathbb{Z}$ for convenience, so that $\phi^2 - 2t\phi + h(\phi)I = 0$. We can re-write this as

$$\overline{\phi} \circ \phi = \phi \circ \overline{\phi} = h(\phi)I \quad \text{where } \overline{\phi} := 2tI - \phi$$

is the dual isogeny.

Exercise 14.6.8. Verify that $\overline{\phi}^2 - 2t\overline{\phi} + h(\phi)I = 0$.

[31] We could have proved this directly by using (14.4) here and the last display.

14.7. The Frobenius map and the number of points on $E(\mathbb{F}_{p^k})$. In a field K of characteristic p we have the *Frobenius endomorphism* $\phi_p : E(K) \to E(K)$ defined by $\phi((x,y)) = (x^p, y^p)$, where E is defined over $\mathbb{F}_p$. That this maps E to E follows from the fact that

$$y^p = (x^3 + ax + b)^p = x^{3p} + a^p x^p + b^p = (x^p)^3 + ax^p + b$$

since $a, b \in \mathbb{F}_p$. To verify that this is indeed an endomorphism we need look at $\phi_p(P_1 + P_2)$. If $\lambda = \frac{y_2 - y_1}{x_2 - x_1}$ is the slope of the line between P_1 and P_2 then $x_1 + x_2 + x_3 = \lambda^2$. This implies that $x_1^p + x_2^p + x_3^p = (\lambda^p)^2$. Now $x_j^p = x(\phi_p(P_j))$ and $\lambda^p = \frac{y_2^p - y_1^p}{x_2^p - x_1^p}$ is the slope of the line between $\phi_p(P_1)$ and $\phi_p(P_2)$.

`Exercise` 14.7.1. Prove that the Frobenius map commutes with every other endomorphism. (Hint: Think what happens when you write the maps explicitly.)

Now evidently $h(\phi_p) = p$, and so we have $(\phi_p^2 - 2\langle \phi_p, I\rangle \phi_p + p)P = \mathcal{O}$. Write $x^2 - 2tx + p = (x - \alpha)(x - \overline{\alpha})$, where $2t = 2\langle \phi_p, I\rangle$. By the previous subsection we know that $|t| \le \sqrt{p}$.

We wish to determine the size of $E(\mathbb{F}_p)$. We shall consider $K = \overline{\mathbb{F}}_p$ above, the algebraic closure of $\mathbb{F}_p$. Now if $(x,y) \in E(K)$ then $\phi_p(x,y) = (x,y)$ if and only if $x^p = x$ and $y^p = y$, which implies that $x, y \in \mathbb{F}_p$. In other words

$$\#E(\mathbb{F}_p) = \#\{P \in E(K) : \ \phi_p(P) = P\}.$$

Now the condition $\phi_p(P) = P$ is the same as saying $(\phi_p - I)(P) = 0$, and so

$$\#E(\mathbb{F}_p) = |\ker(\phi_p - I)| = |\deg(\phi_p - I)| = h(\phi_p - I) = h(\phi_p) - 2t + 1 = p + 1 - 2t,$$

by (14.4). That is, $p + 1 - \#E(\mathbb{F}_p) = 2t = \alpha_p + \overline{\alpha_p}$, and so

$$\boxed{|p + 1 - \#E(\mathbb{F}_p)| \le 2\sqrt{p}.} \tag{14.5}$$

Now $|\alpha_p| = \sqrt{p}$, and α_p and $\overline{\alpha_p}$ are called the *Frobenius eigenvalues*. This is the result we promised way back in section 9.4, when we worked with the two simplest examples, $y^2 = x^3 + ax$ and $y^2 = x^3 + b$.

Now $x^2 - 2tx + p = (x - \alpha)(x - \overline{\alpha})$ divides $(x^k - \alpha^k)(x^k - \overline{\alpha}^k) = x^{2k} - 2tx^k + p^k$. In particular is we take $x = \phi_p$ then $((\phi_p^k)^2 - 2t_k\phi_p^k + p^k)(P) = 0$, where $2t_k = 2\langle \phi_p^k, I\rangle = \alpha^k + \overline{\alpha}^k$.

`Exercise` 14.7.2. (i) Prove that $2t_k \equiv (2t)^k \pmod p$. (Hint: Use the expansion in terms of α and $\overline{\alpha}$.) (ii) Deduce that if $p > 2$ then $p | t_k$ if and only if $t = 0$; in which case $\phi_p^2 + pI = 0$, and $t_k = 0$ if k is odd, with $t_k = (-p)^{k/2}$ if k is even. (iii) Prove that if $t = 0$ then $[p](x,y) = (x^{p^2}, -y^{p^2})$ in $E(\overline{\mathbb{F}_p})$.

Let $q = p^k$. The Frobenius map in $\mathbb{F}_q$ is given by $\phi_q(x,y) = (x^q, y^q) = \phi_p^k(x,y)$ so that $\phi_q = \phi_p^k$, and so $(\phi_q^2 - 2t_k\phi_q + q)(P) = 0$ for all $P \in E(\mathbb{F}_q)$; moreover $\phi_q(u) = u$ if and only if $u \in \mathbb{F}_q$, and so we can deduce, as above that

$$\#E(\mathbb{F}_q) = q + 1 - 2t_k = p^k + 1 - \alpha^k - \overline{\alpha}^k. \tag{14.6}$$

Therefore

$$\sum_{k\geq 1} \#E(\mathbb{F}_{p^k})\ \frac{x^k}{k} = \sum_{k\geq 1}(p^k+1-\alpha^k-\overline{\alpha}^k)\ \frac{x^k}{k}$$
$$= \log\,(1-\alpha x) + \log\,(1-\overline{\alpha}x) - \log\,(1-x) - \log\,(1-px),$$

and so

$$\exp\left(\sum_{k\geq 1} \#E(\mathbb{F}_{p^k})\ \frac{x^k}{k}\right) = \frac{(1-\alpha x)(1-\overline{\alpha}x)}{(1-x)(1-px)} = \frac{1-2tx+px^2}{(1-x)(1-px)} = 1+\ \frac{\#E(\mathbb{F}_p)x}{(1-x)(1-px)}\ .$$

It is usual to take $x = p^{-s}$ here and take the product over all primes p; this then yields

$$\prod_p \frac{1-2tp^{-s}+p^{1-2s}}{(1-p^{-s})(1-p^{1-s})} \asymp \frac{\zeta(s)\zeta(s-1)}{L(E,s)}\ . \tag{14.7}$$

By "$\asymp$" we mean that the factors this is not exactly the definition of $L(E,s)$; the definition is correct when $p\not|\Delta$, but one has to do things a little differently when $p|\Delta$.

An alternate proof for determining $\#E(\mathbb{F}_p)$. Let $P=(x,y)$ and $Q := \phi_p(P) = (x^p, y^p)$, so that $h(Q) = p$. Noe the slope of the line between $-P$ and Q is $\frac{y^p+y}{x^p-x}$, and the x-co-ordinate of $Q-P$ is given by

$$\left(\frac{y^p+y}{x^p-x}\right)^2 - x^p - x\ = \frac{y^{2p}+y^{p+1}+y^2-(x^p+x)(x^p-x)^2}{(x^p-x)^2}$$
$$= \frac{2(x^3+ax+b)^{\frac{p+1}{2}} + x^{2p+1} + x^{p+2} + ax^p + ax + 2b)}{(x^p-x)^2}$$

since $y^{2p} = x^{3p} + ax^p + b$ in $\mathbb{F}_p$. Hence $h(Q-P)$, which is the degree of this rational function, is given by $2p+1-\deg((y^p+y)^2,(x^p-x)^2)$. Now $x^p - x = \prod_{t\in\mathbb{F}_p}(x-t)$, and

$$(y^p+y)^2 = y^2(1+(y^2)^{\frac{p-1}{2}}) = (x^3+ax+b)\left(1+(x^3+ax+b)^{\frac{p-1}{2}})\right).$$

Therefore

$$\deg((y^p+y)^2,(x^p-x)^2) = \sum_{t\in\mathbb{F}_p}\left\{1-\left(\frac{t^3+at+b}{p}\right)\right\},$$

and so

$$h(Q-P)-h(Q)-1 = \sum_{t\in\mathbb{F}_p}\left(\frac{t^3+at+b}{p}\right) = \#\{(t,u)\in\mathbb{F}_p^2:\ u^2=t^3+at+b\}-p.$$

Now p was a generic point so this is really $h(\phi_p - I) - h(\phi_p) - 1$ as before.

Isogenous curves. Suppose that $\psi : E(\mathbb{F}_p) \to E'(\mathbb{F}_p)$ is a homomorphism defined by a rational function. Evidently the Frobenius map commutes with ψ, but we need to be a little careful and denote ϕ_E to be Frobenius acting on E, and $\phi_{E'}$ to be Frobenius acting on E'; and therefore we can write $\psi \circ \phi_E = \phi_{E'} \circ \psi$. This implies that $\psi \circ (I - \phi_E) = (I - \phi_{E'}) \circ \psi$. Taking heights we deduce that $h(\psi)h(I - \phi_E) = h(\psi \circ (I - \phi_E)) = h((I - \phi_{E'}) \circ \psi) = h(I - \phi_{E'})h(\psi)$ and so

$$\#E(\mathbb{F}_p) = h(I - \phi_E) = h(I - \phi_{E'}) = \#E'(\mathbb{F}_p).$$

Now suppose that $\psi : E(\mathbb{Q}) \to E'(\mathbb{Q})$ is a homomorphism defined by a rational function. Evidently ψ may be reduced to a homomorphism mod p for all but finitely many primes p, and therefore $\#E(\mathbb{F}_p) = \#E'(\mathbb{F}_p)$ for all but finitely many primes p. Notice that this implies the $L(E, s)$ and $L(E', s)$ are the same other than for the Euler product factor at finitely many primes.

What about the converse? Is it true that if $\#E(\mathbb{F}_p) = \#E'(\mathbb{F}_p)$ for all but finitely many primes p then there is a homomorphism $\psi : E(\mathbb{Q}) \to E'(\mathbb{Q})$? This is a deep conjecture of Tate, that was resolved by Faltings (1983?).

14.8. Classifying endomorphisms when $t \neq 0$. Suppose that ψ is an endomorphism of $E(\mathbb{F}_q)$ where $q = p^k$. By exercise 14.7.1 we know that ϕ_q commutes with ψ. Now, by the work of section 14.6, we know that ψ, ϕ_q and $\psi\phi_q$ all satisfy polynomials of degree two over the integers, say

$$\psi^2 - 2T\psi + hI \;=\; \phi_q^2 - 2t_k\phi_q + qI \;=\; (\psi\phi_q)^2 - 2\tau(\psi\phi_q) + qhI \;=\; 0,$$

the last coefficient since $h(\psi\phi_q) = h(\psi)h(\phi_q) = qh$. Now, since ψ and ϕ_q commute we have

$$2\tau(\psi\phi_q) - qhI = (\psi\phi_q)^2 = \psi^2\phi_q^2 = (2T\psi - hI)(2t_k\phi_q - qI),$$

and therefore

$$(a\phi_q + bI)\psi = 2h(qI - t_k\phi_q).$$

where $a = 2(\tau - 2t_kT),\ b = 2qT$. Applying $a\overline{\phi_q} + bI$ to both sides, we obtain

$$N\psi = c\phi_q + dI$$

for some integers c, d, N. Now $N = b^2 + 2t_kab + qa^2$ which is $\neq 0$ unless $a = b = 0$, or $t_k^2 = q$ and $at_k + b = 0$. In the latter case, $t = 0$ by exercise 14.7.2, contradicting the hypothesis, so we must have $a = b = 0$, which implies that $2h(qI - t_k\phi_q) = 0$. But then by exercise 14.6.5(iii), we have $t_k\phi_q = qI$ and so $t_k = \sqrt{q}$, which again implies that $t = 0$ by exercise 14.7.2, contradicting the hypothesis.

Now suppose that $\sigma \in \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$. Since ϕ_p is defined over $\mathbb{F}_p$, so is $\phi_q = \phi_p^k$ and so $\sigma(\phi_q) = \phi_q$. Therefore applying σ to the equation above we obtain $N\sigma(\psi) = c\phi_q + dI = N\psi$. Hence $N(\sigma(\psi) - \psi) = 0$ and so $\sigma(\psi) = \psi$ by exercise 14.6.5(iv), for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{F}_p}/\mathbb{F}_p)$. That means that ψ is defined over $\mathbb{F}_p$.

`Exercise` 14.8.1. Use that $N\psi = c\phi_q + dI$ to prove that $T^2 - h(\psi) = (t_k^2 - q)(c/N)^2 \in (t_k^2 - q)\mathbb{Q}^2 \in (t^2 - p)\mathbb{Q}^2$.

Combining all of this information, and denoting the ring of integers of $\mathbb{Q}(\sqrt{d})$ by $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$, we have proved:

Proposition 14.8. *Let ϕ_p be the Frobenius endomorphism on E and suppose that it satisfies $\phi_p^2 - 2t\phi_p + pI = 0$, with $t \neq 0$. Let $\alpha = t + \sqrt{d}$ be a root of $x^2 - 2tx + p = 0$, where $d = t^2 - p$. Under the embedding $\phi_p \to \alpha$, we find that $End(E(\overline{F_p}))$ injects into a subring of the ring of integers of $\mathbb{Q}(\sqrt{d})$, and moreover its elements are all defined over $\mathbb{F}_p$. The first remark implies that $End(E(\overline{F_p}))$ is a commutative ring which is isomorphic to an order of $\mathcal{O}_{\mathbb{Q}(\sqrt{d})}$. Note also that if E is defined over $\mathbb{F}_p$ then* $\mathrm{End}_{\mathbb{F}_p}(E) = \mathbb{Z}[\phi_p]$.

Example: We want an example in which $\mathcal{O}_{\mathbb{Q}(\sqrt{d})} \subsetneq \mathrm{End}(E(\overline{F_p})) \subsetneq \mathbb{Z}[t + \sqrt{d}]$.

Suppose that E has complex multiplication by the order with conductor d (< 0). When we reduce mod $\mathcal{P}$ (for an appropriate prime ideal $\mathcal{P}$ dividing the rational prime p) to $\overline{\mathbb{F}}_p$ we might guess that the distinct elements of $\mathrm{End}_{\overline{\mathbb{Q}}}(E)$ reduce to distinct elements of $\mathrm{End}_{\overline{\mathbb{F}}_p}(E)$. Indeed this is true. More can be said if p does not divide d. As one might guess this by comparing Proposition 14.5.3 with Proposition 14.8,

$$\mathrm{End}_{\overline{\mathbb{Q}}}(E) \ = \mathrm{End}_{\overline{\mathbb{F}}_p}(E).$$

Even more is true: If E and E' have complex multiplication by the same order with conductor d (< 0) but are not isomorphic over $\mathbb{C}$, then their reductions are not isomorphic in $\mathbb{F}_p$. Hence the set of elliptic curves over $\mathbb{C}$ with complex multiplication by the same order with conductor d injects (via reduction) into the set of elliptic curves in $\overline{\mathbb{F}}_p$ with the same endomorphism ring. And this goes also in the other direction meaning that if E' is an elliptic curve over $\overline{\mathbb{F}}_p$ whose endomorphism ring is the order of conductor d, then there exists an elliptic curve E over $\mathbb{C}$ with complex multiplication by the same order with conductor d, whose reduction in $\overline{\mathbb{F}}_p$ is isomorphic to E'. This is called the *Deuring lifting lemma*. From this we can deduce that the isomorphism classes of elliptic curves over $\mathbb{F}_p$ for which $\#E(\mathbb{F}_p) = p + 1 - u$ (where $u \neq 0$) are in 1-to-1 correspondence with the ideal classes in the order of conductor $u^2 - 4p$ (using Proposition 14.8 with $u = 2t$); and hence the total number is $H(u^2 - 4p)$, the class number including imprimitive classes (so that $H(d) = \sum_{f^2|d} h(d/f^2)$).

`Exercise` 14.8.2. Show that the total number of isomorphism classes of elliptic curves over $\mathbb{F}_p$ is $2(p-1)$.

Kronecker showed that $2(p-1) = \sum_{u:\ |u|<2\sqrt{p}} H(u^2 - 4p)$. Hence, subtracting what we have in the previous paragraph, we deduce that there are $H(-4p)$ isomorphism classes of elliptic curves over $\mathbb{F}_p$ for which $\#E(\mathbb{F}_p) = p + 1$. It turns out that the Deuring lifting lemma extends to this case (when $u = 0$). Now when $t = u = 0$, we have that

$$\phi_p^2 + pI = 0.$$

This implies that

$$\#E(\mathbb{F}_{p^k}) = h(\phi_{p^k} - I) = h(\phi_p^k - I) = h(\phi_p^k) + h(I) = p^k + 1.$$

The endomorphism structure of E is more interesting. We know that $\mathrm{End}_{\overline{\mathbb{F}}_p}(E)$ contains the order of conductor $-4p$, and in fact it contains much more. For one thing it is not longer commutative being a quaternion algebra (as in section 7.5) of rank four. We call these *supersingular* elliptic curves.

14.8b Attempt at supersingular primes. We are given $E/\mathbb{F}_p$ for which $a_p(E) = 0$. Suppose that ℓ is a prime that is not the norm of an element in $\mathbb{Z}[\sqrt{-p}]$. Define a sequence of ℓ-isogenies $\lambda_j : E_j \to E_{j+1}$, with $E_0 := E$ where we have $E_j[\ell] = \langle P_j, Q_j \rangle$ with $\lambda_j(P_j) = \mathcal{O}$ and $\lambda(Q_j) = Q_{j+1}$ for each j. Now all of these E_j are isogenous, and as there only finitely many different elliptic curves in $\mathbb{F}_p$ we know that, eventually we obtain a cycle, that is $E_m = E_{m+r}$ for some $m \geq 0,\ r \geq 1$. Acting on this with the dual isogenies, that is $\overline{\lambda}_0 \circ \overline{\lambda}_1 \circ \ldots \circ \overline{\lambda}_{m-1}$, we see that $E = E_r$. Hence we have an endomorphism

$$\alpha : E \to E, \text{ defined by } \alpha = \lambda_{r-1} \circ \lambda_{r-2} \circ \ldots \circ \lambda_0, \text{ of degree } \ \ell^r,$$

which is not a norm in $\mathbb{Z}[\sqrt{-p}]$. By considering the action of α on the Tate module for the ℓ-torsion we see that α satisfies a quadratic equation of the form $x^2 - tx + \ell^r I$ but this does not factor in $\mathbb{Z}[\sqrt{-p}]$. Moreover the dual isogeny $\beta = tI - \alpha$. Now $\alpha(P) = 0$ and $\alpha(Q) = tQ - \beta(Q) = tQ$.

We know that α is defined over $\mathbb{F}_{p^2}$ because $\alpha \circ \overline{\alpha} = [\ell^r]$.

`Exercise` Verify that we can construct the quaternions from i, j for which $i^2 = j^2$, $i^4 = 1$, and $iji = j$.

We wish to show that $\alpha\phi_p = \overline{\phi_p}\alpha$.

Since we are working in $\mathbb{F}_{p^2}$ we know that $\phi_p^2 = I$, and so $\phi_p(x, y) = (\overline{x}, \overline{y})$.

If $\alpha(x, y) = (f(x), yg(x))$ then $\phi\alpha(x, y) = (\overline{f(x)}, \overline{yg(x)})$. On the other hand $\overline{\alpha}(x, y) = (\overline{f}(x), y\overline{g}(x))$, and so $\overline{\alpha}\phi(x, y) = (\overline{f}(\overline{x}), \overline{yg}(\overline{x})) = (\overline{f(x)}, \overline{yg(x)}) = \phi\alpha(x, y)$, as desired.

This needs cleaning up! I do not really understand

14.9. Torsion points in $\mathbb{F}_p$. What about the n-torsion points in $\overline{\mathbb{F}}_p$? We start with an elliptic curve $E: \ y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Z}$ and consider this reduced mod p, and hence embedded in $\overline{\mathbb{F}}_p$. For any point $P \in E(\overline{\mathbb{F}}_p)$ one can use the same formulae to describe nP. Therefore the n-torsion points of order > 1 correspond to the roots of $\psi_n(x)^2 = 0$ in $\mathbb{F}_p$. The addition law survives the reduction to $\mathbb{F}_p$ and so we know that the subgroup of n-torsion points on $E(\overline{\mathbb{F}}_p)$, that is $E[n](\overline{\mathbb{F}}_p)$, is isomorphic to a subgroup of $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. But which subgroup?

The first thing to note is that we can focus on prime power torsion, since

$$E[mn](\overline{\mathbb{F}}_p) \cong E[m](\overline{\mathbb{F}}_p) \times E[n](\overline{\mathbb{F}}_p) \ \text{ if } \ (m,n) = 1$$

by the Chinese Remainder Theorem. More subtle is that we only need study prime torsion because the map $[n] : E(\overline{\mathbb{F}}_p) \to E(\overline{\mathbb{F}}_p)$ is a homomorphism, and hence every element of the image has the same number of pre-images, and that number is the size of the kernel. Hence as $E[q](\overline{\mathbb{F}}_p)$ is a subgroup of $\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$ so, for q prime,

$$E[q](\overline{\mathbb{F}}_p) \ \cong \ \text{ to one of } \{\mathcal{O}\}, \ \mathbb{Z}/q\mathbb{Z} \ \text{ or } \ \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z};$$

and therefore

$$E[q^r](\overline{\mathbb{F}}_p) \ \cong \ \text{ to one of } \{\mathcal{O}\}, \ \mathbb{Z}/q^r\mathbb{Z} \ \text{ or } \ \mathbb{Z}/q^r\mathbb{Z} \times \mathbb{Z}/q^r\mathbb{Z}$$

respectively, for each $r \geq 1$.

We now show that if $q \neq p$ then $E[q](\overline{\mathbb{F}}_p) \ \cong \ \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$ and hence $E[q^r](\overline{\mathbb{F}}_p) \ \cong \ \mathbb{Z}/q^r\mathbb{Z} \times \mathbb{Z}/q^r\mathbb{Z}$. To prove this we study the reduction map $\rho_p : E(\overline{\mathbb{Q}}) \to E(\overline{\mathbb{F}}_p)$. This is obviously a homomorphism (as may be seen by manipulating the formulae for adding points). Hence if $P, Q \in E[q](\mathbb{C})$ and $\rho(P) = \rho(Q)$ then $P - Q \neq \mathcal{O}$ yet $\rho(P - Q) = \mathcal{O}$ (in $E(\mathbb{F}_p)$). In other words more than one element of $E[q](\mathbb{C})$ reduces to $\mathcal{O}$ in $E(\mathbb{F}_p)$ and so $\psi_q(x) \pmod p$ must have smaller degree than $\psi_q(x)$ over $\mathbb{C}$. However this is not true, unless $q = p$ since $\psi_q(x) = qx^{\frac{q^2-1}{2}} + \dots$

We now show that $E[q](\overline{\mathbb{F}}_q)$ cannot be $\mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$, else there would be $q^2 - 1$ points (x, y) of order q, and therefore $\psi_q(x)^2$ would have $q^2 - 1$ roots (including multiplicity). However we know that $\psi_q(x)^2 = q^2 x^{q^2-1} + \dots$ and so has $\leq q^2 - 2$ roots.

This leaves only two possibilities so we next determine whether $\mathcal{O}$ is the only point P with $qP = \mathcal{O}$, or whether there are others. To study this we go back to our discussion of the Frobenius map $\phi_q((x, y)) = (x^q, y^q)$ and recall that we showed that there exists an integer T such that $|T| \leq 2\sqrt{q}$ and ϕ_q satisfies the equation $\phi_q^2 - T\phi_q + qI = 0$. This can be re-written as

$$qI = (T\phi_q - \phi_q^2) = \phi_q \circ (TI - \phi_q) = (TI - \phi_q) \circ \phi_q.$$

Now $\ker \phi_q = \{\mathcal{O}\}$ and so $\ker[q] = \ker(TI - \phi_q)$. That is, if $qP = 0$ then $[T]P = \phi_q(P)$, and therefore $[|T|]P = \pm\phi_q(P)$. Hence for $P = (x, y)$ we need that x is a root of $x^q \psi_{|T|}(x)^2 - \phi_{|T|}(x) = T^2 x^{q+T^2-1} + \dots - (x^{T^2} + \dots)$, which is non-zero mod q, unless $q|T$. Therefore

$$\text{If } q \not| T \text{ then } E[q](\overline{\mathbb{F}}_q) \;\cong\; \mathbb{Z}/q\mathbb{Z}.$$

If $T = 0$ then $qP = -\phi_q^2 P$ which has kernel $\{\mathcal{O}\}$, since ϕ_q does too. In fact if $q|T$ and $T \neq 0$ then $T = \pm q$ with $q = 2$ or 3 since $|T| \leq 2\sqrt{q} < 2q$. Therefore $\phi_q^2 = T\phi_q - qI = (\pm\phi_q - I) \circ q$ and hence $\ker[q] \subset \ker \phi_q^2 = \{\mathcal{O}\}$ and so $\ker[q] = \{\mathcal{O}\}$ Therefore,

$$\text{If } q|T \text{ then } E[q](\overline{\mathbb{F}}_q) \;\cong\; \{\mathcal{O}\}.$$

15. Modular Forms

15.1. The magic of Eisenstein series. When working with the Weierstrass $\wp$-function for the lattice $\Lambda = \langle 1, \tau\rangle$ we defined the *Eisenstein series* of weight k as defined by

$$g_k(\tau) := G_k(\Lambda) = \sum_{\substack{w\in\Lambda\\ w\neq 0}} \frac{1}{w^k} = \sum_{\substack{m,n\in\mathbb{Z}\\ (m,n)\neq(0,0)}} \frac{1}{(m+n\tau)^k},$$

for each integer $k \geq 3$. We saw that this sum is absolutely convergent when $\mathrm{Im}(\tau) \neq 0$, and so g_k is analytic in $\mathcal{H}$ (there is no loss of generality assuming that $\tau \in \mathcal{H}$, since if τ is not then $\tau \notin \mathbb{R}$ so we can replace it by $-\tau$). From the definition it is evident that

$$g_k(\tau + 1) = g_k(\tau)$$

taking $(m, n) \to (m - n, n)$, which is an invertible map. More interestingly

$$\tau^{-k} g_k(-1/\tau) = \tau^{-k} \sum_{\substack{m,n\in\mathbb{Z}\\ (m,n)\neq(0,0)}} \frac{1}{(m-n/\tau)^k} = \sum_{\substack{m,n\in\mathbb{Z}\\ (m,N)\neq(0,0)}} \frac{1}{(m\tau+N)^k} = g_k(\tau),$$

taking $n = -N$. Hence it is not difficult to keep track of the value of $g_k(z)$, under the maps $z \to z + 1$ and $z \to -1/z$, and hence any combination of these maps. We know that these maps generate any map of the form $z \to \frac{az+b}{cz+d}$ for $\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \mathrm{PSL}(2, \mathbb{Z})$.

`Exercise` 15.1. Show that if $\begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \mathrm{PSL}(2, \mathbb{Z})$ then

$$g_k\left(\frac{az+b}{cz+d}\right) = (cz+d)^k g_k(z).$$

We saw that the $\wp$-function was periodic with respect to the lattice generated by the transformations $z \to z+1$ and $z \to z+\tau$, which tiles the upper half plane. Now we have seen that $g_k(z)$ is also more-or-less periodic (other than the simple factor $(cz + d)^k$), but with respect to $\mathrm{PSL}(2, \mathbb{Z})$, that is the group generated by the more complicated transformations $z \to z + 1$ and $z \to -1/z$. This implies that every value of the periodic function is "equivalent" to a unique value in the fundamental domain: For $\wp(z)$ that is the fundamental parallellopiped, and for $g_k(z)$, the fundamental domain is the domain D shown in chapter *.* during our discussion of this group. There is a problem however; that domain D is not closed since there is no point in D at the very top, that is at $i\infty$. It will make our analysis easier if we include $i\infty$ in D (so that D is now compact), but then we need to understand the behaviour of $g_k(\tau)$ there:

We wish to determine $g_k(it)$ for $t \in \mathbb{R}$ and getting increasingly large. If we fix $n \neq 0$ then

$$\sum_{\substack{m\in\mathbb{Z}\\ m\neq 0}} \frac{1}{|m+int|^k} \leq \sum_{m:\ 1\leq|m|\leq|n|t} \frac{1}{|nt|^k} + \sum_{m:\ |m|>|n|t} \frac{1}{m^k} \leq \frac{2|n|t}{|nt|^3} + 2 \sum_{m:\ m>|n|t} \frac{1}{m^3}$$

$$\leq \frac{2|n|t}{|nt|^3} + 2 \sum_{m:\ m>|n|t} \frac{1}{m^3} \leq \frac{3}{|nt|^2}\ .$$

Adding in the terms with $m = 0$ and summing over all $n \neq 0$, gives $6t^{-2}\zeta(2) + 2t^{-3}\zeta(3)$, which $\to 0$ as $t \to \infty$. Hence we find that

$$\lim_{t\to\infty} g_k(it) = \sum_{\substack{m\in\mathbb{Z}\\ m\neq 0}} \frac{1}{|m|^k} = 2\zeta(k).$$

Hence we can define $g_k(i\infty) = 2\zeta(k)$ and hence $g_k(\tau)$ is continuous inside and on D.

15.2. The Fourier expansion of an Eisenstein series. Since $g_k(\tau + 1) = g_k(\tau)$ we know that $g_k(.)$ must have a Fourier expansion; that is, it can be written in the form $g_k(\tau) = \sum_{n\in\mathbb{Z}} a_n e^{2i\pi n\tau}$. It is convenient to write $q := e^{2i\pi\tau}$, so that if $\tau = u + iv$ with $v > 0$ then $|q| = e^{-2\pi v} < 1$. We begin from the well-known identity for the sine function in terms of its zeros:[32]

$$\sin \pi z = \pi z \prod_{n\geq 1} \left(1 - \frac{z^2}{n^2}\right).$$

If we take the logarithmic derivative of both sides we obtain

$$\frac{\pi \cos \pi z}{\sin \pi z} = \frac{1}{z} + \sum_{n\geq 1} \left(\frac{1}{z+n} + \frac{1}{z-n}\right).$$

The left hand side can be written in terms of exponentials, that is

$$\frac{\pi \cos \pi z}{\sin \pi z} = \frac{i\pi(e^{i\pi z} + e^{-i\pi z})}{(e^{i\pi z} - e^{-i\pi z})} = \frac{i\pi(q+1)}{(q-1)} = -i\pi(1 + 2q + 2q^2 + \dots),$$

multiplying top and bottom by $e^{i\pi z}$, and then writing $q = e^{2i\pi z}$. Now equating the last two lines, and differentiating with respect to z, we obtain (as $dq/dz = 2i\pi q$)

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^2} = (2i\pi)^2 \sum_{n\geq 1} n q^n.$$

Differentiating $k - 2$ times we obtain

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^k} = \frac{(-2i\pi)^k}{(k-1)!} \sum_{n\geq 1} n^{k-1} q^n.$$

We can pair up the terms $(-m, -n)$ with (m, n) in the expansion of G_k, to obtain that $g_k(\tau) = 0$ for k odd. For k even this means we need to substitute in $z = n\tau$ into the last

[32] Each of the terms $1 - z^2/n^2$ in the product can be factored as $(1 - z/n)(1 + z/n)$ so it is tempting to rewrite the product as the product of $1 - z/n$, over all non-zero integers n. The problem is that this is not absolutely convergent so one needs to be careful.

equation for $n = 1, 2, \dots$; this causes $q \to q^m$. Hence we have for k even (where the $\zeta(k)$ corresponds to the term where $n = 0$),

$$g_k(\tau) = 2\zeta(k) + 2\,\frac{(-2i\pi)^k}{(k-1)!}\sum_{m,n\geq 1} n^{k-1}q^{mn}$$
$$= 2\zeta(k)E_k(\tau) \ \text{ where } E_k(\tau) := 1 + \frac{2(-1)^{k/2}k}{B_k}\sum_{N\geq 1}\sigma_{k-1}(N)q^N$$

is called the *weight k Eisenstein series*, and where $\sigma_j(N) := \sum_{d|N} d^j = \sum_{dm=N} d^j$, since $\zeta(k) = B_k(2\pi)^k/2\cdot k!$, and B_k is the kth Bernoulli number.[33]

15.3. Modular forms. After such benefits from developing the theory of Λ-periodic functions, it seems like an interesting idea to further develop functions which are more-or-less periodic with respect to $\mathrm{PSL}(2,\mathbb{Z})$ like $g_k(\tau)$. So we define f to be *weakly modular of weight k*, if f satisfies

$$f(z) = (cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right) \quad \text{for all} \quad \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in \mathrm{PSL}(2,\mathbb{Z}).$$

and f is meromorphic on $D \cong \mathcal{H}/\mathrm{PSL}(2,\mathbb{Z})$. Since $\mathrm{PSL}(2,\mathbb{Z})$ is generated by $S := \begin{pmatrix} 1 & 1 \\ 0 & 1\end{pmatrix}$ and $T := \begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix}$, we deduce that this holds if and only if

$$f(z+1) = f(z) \ \text{ and } \ f(-1/z) = z^k f(z).$$

`Exercise` 15.3.1. Let $M = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in \mathrm{PSL}(2,\mathbb{Z})$. (i) Verify that $\mathrm{Im}(Mz) = \mathrm{Im}(z)/|cz+d|^2$. (ii) Verify also that $\frac{d}{dz}(Mz) = \frac{1}{(cz+d)^2}$. (iii) With an appropriate abuse of notation, deduce that the "differential" $f(z)(dz)^{k/2}$ is invariant under the action of $\mathrm{PSL}(2,\mathbb{Z})$. (iv) Show that if $z = x+iy$ then $f(z)y^{k/2-1}dy$ is invariant under the action of $\mathrm{PSL}(2,\mathbb{Z})$. (v) Show that if $z = x+iy$ then $|f(z)y^{k/2}|$ is invariant under the action of $\mathrm{PSL}(2,\mathbb{Z})$.

We know f has a Fourier series which we write as

$$f(z) = F(q) := \sum_{n\in\mathbb{Z}} a_n q^n \ \text{ where } \ q = e^{2i\pi z}.$$

This is meromorphic inside the disk $0 < |q| < 1$. We call f a *modular function* if F is also meromorphic at 0 and we say that f *is meromorphic at* ∞. Evidently this means that $F(q)$ has a Laurent expansion at 0 and so there exists n_0 such that $a_n = 0$ for all $n < n_0$.

[33] This is a famous result of Euler. We define $\sum_{n\geq 0} B_n \frac{x^n}{n!} = \frac{x}{e^x-1}$. Here $B_1 = -1/2$ and $B_n = 0$ for all odd $n \geq 3$. Also $B_2 = 1/6,\ B_4 = B_8 = 1/30,\ B_6 = 1/42,\ B_{10} = 5/66, \dots$

If $f(z)$ is analytic everywhere, including at ∞, then $f(z)$ is a *modular form*. If $f(\infty) = F(0) = a_0 = 0$ then f is a *cusp form*. Therefore a modular form of weight k is given by a series $f(z) = \sum_{n\geq 0} a_n q^n$ (where $q = e^{2i\pi z}$) which converges for $|q| < 1$, and which satisfies an identity $f(-1/z) = z^k f(z)$. It is a cusp form when $a_0 = 0$.

Exercise 15.3.2.(i) Prove that if f and g are modular (or cusp) forms of weight k, then so is $f+g$, as well as cf for any constant $c \neq 0$. (ii) Deduce that the modular forms of weight k form a vector space over $\mathbb{C}$, which will call M_k. (iii) Show that if f_j is a modular form of weight k_j for $j = 1, 2$ then $f_1 f_2$ is a modular form of weight $k_1 + k_2$.

We therefore proved above that $g_k(\tau)$ is a modular form (but not a cusp form) of weight k.

Exercise 15.3.3.(i) Show that if $A(\tau)$ is a modular form of weight k then there exists a constant c such that $A(\tau) - cg_k(\tau)$ is a cusp form of weight k. (ii) Deduce that if C_k is the vector space of cusp forms of weight k, then $M_k = \langle C_k, g_k \rangle$.

In chapter 6 we studied the elliptic curve $E_\Lambda : \ y^2 = 4x^3 - ax - b$ where $a = 60G_4(\Lambda)$ and $b = 140G_6(\Lambda)$. The discriminant of the polynomial $4x^3 - ax - b$ is $16(a^3 - 27b^2)$. Hence we define

$$\Delta(\tau) := (60g_4(\tau))^3 - 27(140g_6(\tau))^2 = \frac{64}{27}\ \pi^{12}(E_4(\tau)^3 - E_6(\tau)^2),$$

which is a modular form of weight 12 (by exercises 15.3(iii) and then (ii)). Note that at $\tau = it$ with $t \to \infty$ we have $g_4(\infty) = 2\zeta(4), g_6(\infty) = 2\zeta(6)$, and so $\Delta(\infty) = (120\zeta(4))^3 - 27(280\zeta(6))^2 = 0$. Hence $\Delta(\tau)$ is a cusp form of weight 12.

Exercise 15.3.4. Show that $\Delta(\tau) = (2\pi)^{12}(q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 + \dots)$

15.4. Determining spaces of modular forms. Let $v_w(f)$ denote the order of f at w; that is $f(z) = (z-w)^{v_w(f)}(c_0 + c_1(z-w) + c_2(z-w)^2 + \dots)$, with $c_0 \neq 0$. For a modular function f of weight k we can restrict attention to the fundamental domain,

$$D := \{z \in \mathbb{C} : \ \mathrm{Im}(z) > 0, \ -\frac{1}{2} \leq \mathrm{Re}(z) < \frac{1}{2}, \ |z| \geq 1, \ \text{if } |z| = 1 \text{ then } \mathrm{Re}(z) \leq 0.\}$$

(Remember that $D \cong \mathcal{H}/\mathrm{PSL}(2,\mathbb{Z})$.) We also let $v_\infty(f)$ be the order of $F(q)$ at $q = 0$. Note that since f is modular, each $v_\alpha(f)$ is a non-negative integer. The key result is that, for $i = e^{i\pi/2}, \ w = e^{2i\pi/3}$,

$$v_\infty(f) + \frac{1}{2}\, v_i(f) + \frac{1}{3}\, v_w(f) + \sum_{\substack{\alpha \in D \\ \alpha \neq i, w\infty}} v_\alpha(f) = \frac{k}{12}\ . \tag{15.1}$$

Proof. If we integrate along the boundary of D, that is along the straight line from $\frac{1+i\sqrt{3}}{2}$ vertically to ∞, then back down the vertical line from ∞ to $\frac{-1+i\sqrt{3}}{2}$, and then along that part of the unit circle from $\frac{-1+i\sqrt{3}}{2}$ to $\frac{1+i\sqrt{3}}{2}$. For this last part we write the contour as $z = e^{i\theta}$ for θ from $2\pi/3$ to $\pi/3$, so that $dz = iz d\theta$. Hence we have

$$\int_D \frac{f'}{f}\,(z)dz = \int_{t=\sqrt{3}/2}^{\infty} \frac{f'}{f}\,(1/2+it)dt - \int_{t=\sqrt{3}/2}^{\infty} \frac{f'}{f}\,(-1/2+it)dt + \int_{\theta=2\pi/3}^{\pi/3} \frac{f'}{f}\,(e^{i\theta})ie^{i\theta}d\theta.$$

As $(f'/f)(z+1) = (f'/f)(z)$, the first two terms cancel. Now $f(-1/z) = z^k f(z)$ and so $(f'/f)(-1/z) = kz + z^2(f'/f)(z)$. Hence if $z = e^{i\theta}$ then

$$(f'/f)(e^{i(\pi-\theta)})e^{i(\pi-\theta)} + (f'/f)(e^{i\theta})e^{i\theta} = -k,$$

and so integrating from θ from $2\pi/3$ to $\pi/2$ and adding in the $\pi - \theta$ term, we obtain

$$\frac{1}{2i\pi}\int_D \frac{f'}{f}(z)dz = -\frac{k}{2\pi}\int_{\theta=2\pi/3}^{\pi/2} 1d\theta = \frac{k}{12}.$$

On the other hand the sum of the residues is $\sum_{\alpha\in D} v_\alpha(f)$, and the result follows as long as f has no zeros or poles on the boundary of D.

Now we use (15.1) to determine all $f \in M_k$:

If $k = 0$ or 2 then any $f \in M_k$ has no zeros, and is thus a constant; that is $M_0 = \langle 1\rangle$.. For $k = 2$ this is impossible and so $M_2 = \emptyset$.

If $k = 4$ then f has a simple zero at w and no other zero.

If $k = 6$ then f has a simple zero at i and no other zero.

If $k = 8$ then f has a double zero at w and no other zero.

If $k = 10$ then f has simple zeros at i and w and no other zero.

For $k = 4, 6, 8, 10$ we have $g_k \in M_k$ and if $f \in M_k$ then the above criteria show us that f/g_k has no poles or zeros and is in M_0, and so is a constant. Hence $M_k = \langle g_k\rangle$.

We notice however that $g_4^2 \in M_8$, $g_4g_6 \in M_{10}$ and so $g_8 = c_8g_4^2$ and $g_{10} = c_{10}g_4g_6$ for some constant c_8 and c_{10}.

`Exercise` 15.4.1. Determine the values of c_8 and c_{10}.

Now $g_4(i) \neq 0$ and $g_6(i) = 0$, so that $\Delta(i) = (60g_4(i))^3 - 27(140g_6(i))^2 \neq 0$, and so Δ is not identically zero. We know that $\Delta(\tau)$ has a zero at ∞ and has weight 12, and hence ∞ is the only zero of $\Delta(\tau)$ by (15.1), and is a simple zero.

`Exercise` 15.4.2. Suppose that $f \in C_k$ for some $k \geq 12$. (i) Prove that f/Δ is a modular form of weight $k - 12$. (ii) Deduce that $C_k = \Delta M_{k-12}$. (iii) Deduce that $M_k = \langle g_k, \Delta M_{k-12}\rangle$. (iv) Deduce that $\dim M_k = \dim M_{k-12} + 1$ for all $k \geq 12$. (v) Conclude that $\dim M_k = [\frac{k}{12}]$ plus 1 unless $k \equiv 2 \pmod{12}$.

We have a better description of M_k:

$$M_k = \langle g_4^i g_6^j : \ 4i + 6j = k, \ i, j \geq 0\rangle,$$

which we prove by induction on k. Above we saw that this is true for each $k \leq 10$. For any given $k \geq 12$, we first select integers $u, v \geq 0$ such that $4u + 6v = k$, and therefore for every f there exists a constant c such that $f - cg_4^u g_6^v$ is a cusp form. But we have seen that every cusp form can be written as Δ times an element of M_{k-12}. Now since $\Delta = (60g_4)^3 - 27(140g_6)^2$, the result follows from the induction hypothesis.

Of particular interest is that g_k is a polynomial in g_4 and g_6.

It is worth noting that the modular forms $g_4^i g_6^j : \ 4i + 6j = k$ are linearly independent over $\mathbb{C}$, and so form a basis for M_k. For if they were not independent then there would exists complex numbers c_j, not all zero, such that $\sum_j c_j g_4^i g_6^j = 0$. Now k is even. If $k \equiv 0$

(mod 4) then $2j \equiv 4i + 6j = k \equiv 0 \pmod 4$ and so j is even, say $j = 2J$, in which case $i = k/4 - 3J$. Therefore $0 = \sum_J c_{2J} g_4^{k/4-3J} g_6^{2J} = g_4^{k/4} \sum_J c_{2J} (g_6^2/g_4^3)^{2J}$, and so g_6^2/g_4^3 is the root of a polynomial in $\mathbb{C}[x]$, and therefore the ratio $g_6(\tau)^2 : g_4(\tau)^3$ is fixed. But the ratio is $0 : 1$ at $\tau = i$ and $1 : 0$ at $\tau = w$, which is impossible. An analogous argument works when $k \equiv 2 \pmod 4$.

Another basis for M_k is given by $\langle E_4^i E_6^j : \ 4i + 6j = k, \ i, j \geq 0 \rangle$, which is useful because the coefficients of E_4 and E_6 are rational numbers.

15.5. The j-function. The j-function is defined by

$$j(\tau) = 1728 \ \frac{(60 g_4(\tau))^3}{\Delta(\tau)} = 1728 \ \frac{E_4(\tau)^3}{E_4(\tau)^3 - E_6(\tau)^2} \ .$$

Since Δ is a cusp form and g_4 is a modular form (but not a cusp form) of weight 12, we see that $j(\tau)$ is a modular function of weight 0 which is holomorphic on $\mathcal{H}$, and has a simple pole at ∞.

Now let $f_c(\tau) := 1728(60 g_4(\tau))^3 - c\Delta(\tau)$ which belongs to M_{12}. By formula (15.1) we see that it either has a triple zero at w, or a double zero at i, or a simple zero at some other point of D, and *no other zero*. Hence there exists a unique value of τ in D for which $j(\tau) = c$. In other words

$$j : D \to \mathbb{C} \text{ is a bijection.}$$

Note that $j(\tau) = \infty$ if and only if $\Delta(\tau) = 0$, that is if for $E_\tau : y^2 = x^3 + ax + b$, the polynomial $x^3 + ax + b$ has a repeated root and so the elliptic curve E_τ degenerates into something linear.

For the elliptic curves $E(a, b) : \ y^2 = x^3 + ax + b$ over $\mathbb{C}$, we defined the j-invariant:

$$j(E(a, b)) = 1728 \ \frac{a^3}{4a^3 + 27b^2} \ .$$

If we multiply through by λ^6, then the map $y \to y/\lambda^3, \ x \to x/\lambda^2$ gives an isomorphism $E(a, b) \cong_{\mathbb{C}} E(\lambda^4 a, \lambda^6 b)$, for any non-zero $\lambda \in \mathbb{C}$. Note that $j(E(a, b)) = j(E(\lambda^4 a, \lambda^6 b))$.

`Exercise` 15.5. (i) Prove that $E(a, b) \cong_{\mathbb{C}} E(A, B)$ if and only if $A^3 : B^2 = a^3 : b^2$. Show that, up to sign, there is a unique $\lambda \in \mathbb{C}$ such that $A = \lambda^4 a, \ B = \lambda^6 b$. (ii) Prove that if $E(a, b) \cong_{\mathbb{C}} E(A, B)$ then $j(E(a, b)) = j(E(A, B))$. (iii) Prove that $j(E_\tau) = j(\tau)/4$ where $E_\tau = E_\Lambda$ is as given above.

Given any elliptic curve $E : \ y^2 = x^3 + ax + b$ we know that there exists a unique $\tau \in D$ such that $j(\tau) = 4j(E)$. But then $j(E_\tau) = j(\tau)/4 = j(E)$, and so $E_\tau \cong_{\mathbb{C}} E$ by exercise 15.8(ii). Now suppose that E_τ, after a suitable change of variables, is the elliptic curve $y^2 = x^3 + Ax + B$. We select λ as in exercise 15.8(i). Now if we work with the lattice $\Lambda' = \lambda\Lambda = \langle \lambda, \lambda\tau \rangle$ then evidently $g_k(\Lambda') = \lambda^{-k} g_k(\Lambda)$ for all k, by definition, and so $E_{\Lambda'} : \ y^2 = x^3 + \lambda^{-4} Ax + \lambda^{-6} B = x^3 + ax + b$, and therefore $E_{\Lambda'} = E$. Observing that $\lambda\Lambda = -\lambda\Lambda$ we deduce that

The lattices $\{m\omega_1 + n\omega_2 : \ m, n \in \mathbb{Z}\}$ with $\omega_1, \omega_2 \in \mathbb{C}^*$ but $\omega_1/\omega_2 \notin \mathbb{R}$
are in 1-to-1 correspondence with
The elliptic curves $E : \ y^2 = x^3 + ax + b$ with $a, b \in \mathbb{C}$, and not both 0,

via the map $\Lambda \to E_\Lambda$.

Another property of $j(\tau)$ is that f is a modular function of weight 0 if and only if it is rational function of j. That f is a modular function of weight 0 if it is rational function of j is immediate. In the other direction assume that f is a modular function of weight 0. For each pole $\tau_0 \in \mathbb{C}$ of f (including multiplicity) we can multiply f through by $j(\tau) - j(\tau_0)$ to obtain a modular function g of weight 0 which is analytic except perhaps at ∞. If the pole at ∞ has order m then $\Delta^m g$ is analytic and thus a modular form of weight $12m$. As we saw in the previous section, this means that $\Delta^m g$ can be written as a linear combination of terms $g_4^i g_6^j$ with $4i + 6j = 12m$, and thus of $g_4^{3k} g_6^{2\ell}$ with $k + \ell = m$. Hence g can be written as a polynomial in (g_4^3/Δ) and (g_6^2/Δ), and thus as a polynomial in j. Hence f is a rational function of j.

Now by a simple calculation with the Eisenstein series we have

$$j(\tau) = \frac{1}{q} + 744 + 196884q + 21493760q^2 + 864299970q^3 + \dots$$

These coefficients have all sorts of interesting properties. For example, in 1982 Griess constructed that the "monster group" (the largest sporadic simple group) as the automorphism group of a 196884-dimensional commutative nonassociative algebra.

15.6. The j-invariant and complex multiplication. Let R be an order of the ring of integers of $\mathbb{Q}(\sqrt{-d})$, an imaginary quadratic field. Now each ideal Λ of R yields a lattice Λ of $\mathbb{C}$ and we know that the endomorphism ring of $\mathbb{C}/\Lambda$ is R. In fact $\mathbb{C}/\Lambda$ only depends on the ideal class of Λ in R.

On the other hand if the endomorphism ring of E is R then $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ for some ideal Λ of R, coming from a unique ideal class. Hence there is a 1-to-1 correspondence between ideal classes of R and isomorphism classes of elliptic curves with endomorphism ring R. Since there are only finitely many classes of ideals, there are only finitely many such elliptic curves.

Let E be an elliptic curve with endomorphism ring R, so that $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ for some ideal Λ of R. If σ is any automorphism then $E^\sigma(\mathbb{C}) \cong \mathbb{C}/\Lambda^\sigma$, which will also have endomorphism ring R. Hence $j(E)^\sigma = j(E^\sigma)$ belongs to a finite set (the size of which is the size of the class group of R). Hence $j(E)$ is an algebraic number, with degree at most the size of the class group of R. In particular if $\mathbb{Q}(\sqrt{-d})$ has class number one, then the values of $j(\tau)$ with $\tau = \sqrt{-d}$ or $\frac{1+\sqrt{-d}}{2}$ must be rational numbers. Hence we have for the nine fields of class number one

$$j(2i) = 12^3,\ j(\sqrt{-8}) = (-20)^3,\ j\left(\frac{1+\sqrt{-3}}{2}\right) = 0,\ j\left(\frac{1+\sqrt{-7}}{2}\right) = (-15)^3,$$
$$j\left(\frac{1+\sqrt{-11}}{2}\right) = (-32)^3,\ j\left(\frac{1+\sqrt{-19}}{2}\right) = (-96)^3,\ j\left(\frac{1+\sqrt{-43}}{2}\right) = (-960)^3,$$
$$j\left(\frac{1+\sqrt{-67}}{2}\right) = (-5280)^3,\ j\left(\frac{1+\sqrt{-163}}{2}\right) = (-640320)^3.$$

So there is more to be understood – the j-values are integers, and even cubes! In fact their prime factors are all very small also. One amusing observation comes from substituting in

$\tau = \frac{1+\sqrt{-163}}{2}$ into the q-expansion for $j(\tau)$. Then we see that $q = -e^{-\pi\sqrt{163}} \approx -e^{-40}$, and so

$$(-640320)^3 = j\left(\frac{1+\sqrt{-163}}{2}\right) = -e^{\pi\sqrt{163}} + 744 - 196884e^{-\pi\sqrt{163}} + \dots$$

which implies that $e^{\pi\sqrt{163}}$ is very close to the integer $(640320)^3+744$. Indeed the difference is about 15.5×10^{-13}, so it is easy to fool someone with a handheld calculator into believing that $e^{\pi\sqrt{163}}$ is an integer.

15.7. Almost an Eisenstein series. Although the sum for g_2 is not absolutely convergent, we saw that if we define

$$g_2(\tau) := 2\zeta(2) + \sum_{n\neq0}\sum_{m\in\mathbb{Z}} \frac{1}{(m+n\tau)^2} = \frac{\pi^2}{3}\left(1 - 24\sum_{N\geq1}\sigma_1(N)q^N\right),$$

we obtain a Fourier expansion. The logarithmic derivative of the *Dedekind η-function*,

$$\eta(\tau) := q^{\frac{1}{24}}\prod_{n\geq1}(1-q^n),$$

(which we already saw in chapter 11 while discussing partitions) is

$$\begin{aligned}\frac{\eta'(\tau)}{\eta(\tau)} &= (2i\pi)q\left(\frac{1}{24q} + \sum_{n\geq1}\frac{(-nq^{n-1})}{1-q^n}\right) = \frac{2i\pi}{24}\left(1 - 24\sum_{n\geq1}\frac{nq^n}{1-q^n}\right)\\ &= \frac{i\pi}{12}\left(1 - 24\sum_{m,n\geq1} nq^{mn}\right) = \frac{i\pi}{12}\left(1 - 24\sum_{N\geq1}\sigma_1(N)q^N\right) = \frac{i}{4\pi}\,g_2(\tau).\end{aligned}$$

Evidently $g_2(\tau+1) = g_2(\tau)$. However $g_2(-1/\tau) = 2\zeta(2) + \tau^2\sum_{n\neq0}\sum_{m\in\mathbb{Z}}\frac{1}{(m\tau-n)^2}$, and we cannot swap the order of summation of m and n, as we did for weight > 2, because we do not have absolute convergence. So we will have to proceed more carefully: We will show that $g_2(-1/\tau) = \tau^2 g_2(\tau) - 2i\pi\tau$. Now

$$g_2(\tau) = \sum_{n\in\mathbb{Z}}{\sum_{m\in\mathbb{Z}}}' \frac{1}{(m+n\tau)^2} \quad\text{and}\quad \tau^{-2}g_2(-1/\tau) = \sum_{m\in\mathbb{Z}}{\sum_{n\in\mathbb{Z}}}' \frac{1}{(m+n\tau)^2}$$

where $'$ indicates missing out the term $(m,n) = (0,0)$, and we define

$$F(\tau) = \sum_{n\in\mathbb{Z}}{\sum_{m\in\mathbb{Z}}}'' \frac{1}{(m+n\tau)(m-1+n\tau)} \quad\text{and}\quad H(\tau) = \sum_{m\in\mathbb{Z}}{\sum_{n\in\mathbb{Z}}}'' \frac{1}{(m+n\tau)(m-1+n\tau)}$$

where $''$ indicates missing out the terms $(m,n) = (0,0)$ and $(1,0)$. In each case the inner sum is absolutely convergent, but all of the terms together are not, so there may be some

difference when we swap the order of summation going from $g_2(\tau)$ to $\tau^{-2}g_2(-1/\tau)$, or from F to H. Now comparing terms we see that $1(m+n\tau)(m-1+n\tau) - \frac{1}{(m+n\tau)^2} = \frac{1}{(m+n\tau)^2(m-1+n\tau)}$, and the sum of this over all m and n is absolutely convergent, so we can deduce that $g_2(\tau) - F(\tau) = \tau^{-2}g_2(-1/\tau) - H(\tau)$. Next we wish to evaluate F and H. Now $F(\tau)$ is easy because one can telescope the identity $\frac{1}{(m+n\tau)(m-1+n\tau)} = \frac{1}{m-1+n\tau} - \frac{1}{m+n\tau}$ so that every inner sum equals 0 except the term for $n = 0$ which equals 2; hence $F(\tau) = 2$. Evaluating $H(\tau)$ is trickier. We will proceed by:

$$\begin{aligned} H(\tau) &= \lim_{M\to\infty} \sum_{-M<m\leq M} {\sum_{n\in\mathbb{Z}}}'' \frac{1}{(m+n\tau)(m-1+n\tau)} \\ &= \lim_{M\to\infty} \sum_{n\in\mathbb{Z}} {\sum_{-M<m\leq M}}'' \frac{1}{m-1+n\tau} - \frac{1}{m+n\tau} \\ &= 2 - \lim_{M\to\infty} \sum_{n\in\mathbb{Z}} \frac{1}{M-n\tau} + \frac{1}{M+n\tau} = 2 - \lim_{M\to\infty} \sum_{n\in\mathbb{Z}} \frac{2M}{M^2-n^2\tau^2}\ . \end{aligned}$$

Exercise 15.7.1. Show that this equals

$$2 - 2\int_{-\infty}^{\infty} \frac{du}{1-\tau^2u^2} = 2 - 2i\pi/\tau.$$

Combining all this information we have $g_2(\tau) - \tau^{-2}g_2(-1/\tau) = F(\tau) - H(\tau) = 2i\pi/\tau$, which implies that $g_2(-1/\tau) = \tau^2 g_2(\tau) - 2i\pi\tau$.

Hence we deduce that

$$\frac{\tau^2}{\eta(-1/\tau)}\frac{d\eta(-1/\tau)}{d\tau} = \frac{\eta'(-1/\tau)}{\eta(-1/\tau)} = \frac{i}{4\pi}\ g_2(-1/\tau) = \frac{i}{4\pi}\ \tau^2 g_2(\tau) + \frac{\tau}{2} = \tau^2\ \frac{\eta'(\tau)}{\eta(\tau)} + \frac{\tau}{2},$$

which can be conveniently rewritten as

$$\frac{d\eta(-1/\tau)}{\eta(-1/\tau)} = \frac{d\eta(\tau)}{\eta(\tau)} + \frac{1}{2}\frac{d\tau}{\tau}\ .$$

Integrating, and then substituting in $\tau = i$ (as $-1/i = i$), we deduce that

$$\eta(-1/\tau) = (\tau/i)^{1/2}\eta(\tau).$$

Exercise 15.7.2. (i) Show that $\eta(\tau+1) = e^{i\pi/12}\eta(\tau)$. (ii) Deduce from the last two equations that $h(\tau) := \eta(\tau)^{24}$ is weakly modular of weight 12. (iii) Use exercise 10.5 to exhibit the Fourier expansion of $h(.)$, so as to justify that $h(.)$ is not only a modular form, but also a cusp form of weight 12. (iv) Deduce that $\eta(\tau)^{24}$ is a constant multiple of $\Delta(\tau)$, and compare leading coefficients (see exercise 15.3.4) to show that

$$\Delta(\tau) = (2\pi)^{12}\eta(\tau)^{24} = (2\pi)^{12}q\prod_{n\geq 1}(1-q^n)^{24}.$$

15.8. Sublattice and subgroups. Given a lattice $\Lambda = \langle 1, \tau\rangle$ we wish to find all of the sublattices Λ' of index N. The set $S = \Lambda/\Lambda'$ is a group of order N, and a quotient of $\Lambda \cong \mathbb{Z} \times \mathbb{Z}$, so must be isomorphic to a subgroup of $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$.

`Exercise` 15.8.1. Show that all subgroups S of $\mathbb{Z}/N\mathbb{Z}\times\mathbb{Z}/N\mathbb{Z}$ of index N have bases of the form $(a, b), (0, d)$ with $0 \le b < d,\ ad = n$; and that these are distinct. (Hint: Use Gaussian elimination mod N to show that any subgroup has a basis of the form $(a, b), (0, d)$ with $0 \le b < d$ and $a, d|n$. Then prove that we can assume $ad = n$ if the index is n.)

Hence the sublattices Λ' of $\Lambda = \langle 1, \tau\rangle$ of index N can be described as $\Lambda' = \langle a\tau + b, d\rangle$ for integers a, b, d satisfying $0 \le b < d,\ ad = n$. We will need to better understand their structure. To this end note that the key idea in the proof of exercise 15.8.1 is that $\langle a\tau + b, c\tau + d\rangle$ equals $\langle A\tau + B, C\tau + D\rangle$ if and only if there exists $M \in \mathrm{SL}(2, \mathbb{Z})$ such that $\begin{pmatrix} a & b \\ c & d\end{pmatrix} = M\begin{pmatrix} A & B \\ C & D\end{pmatrix}$; that is the distinct lattices of index N are isomorphic to the distinct cosets of the integer matrices of determinant N, under the action of multiplication on the left by elements of $\mathrm{SL}(2, \mathbb{Z})$. These cosets have representatives $M(N) := \left\{ \begin{pmatrix} a & b \\ 0 & d\end{pmatrix} \pmod N \text{ with } ad = n,\ 0 \le b < d\right\}$.

`Exercise` 15.8.2. Prove that the distinct cosets of the integer matrices of determinant N, under the action of multiplication on the left by elements of $\mathrm{SL}(2, \mathbb{Z})$, have representatives:
(i) $\left\{ M\begin{pmatrix} 1 & 1 \\ 0 & 1\end{pmatrix} \pmod N :\ M \in M(N)\right\}$; and (ii) $\left\{ M\begin{pmatrix} 0 & 1 \\ -1 & 0\end{pmatrix} \pmod N :\ M \in M(N)\right\}$.

If we have two such matrices, perhaps from two different such sets then

$$\begin{pmatrix} a & b \\ 0 & d\end{pmatrix}\begin{pmatrix} A & B \\ 0 & D\end{pmatrix} = \begin{pmatrix} aA & aB + bD \\ 0 & dD\end{pmatrix} \tag{15.2}$$

If $(m, n) = 1$ then $M(m)M(n) = M(mn)$ since, in (15.2), $aA \cdot dD$ runs though all products of two factors of mn, each exactly once, while $aB + bD$ runs through all residue classes mod dD, each exactly once.

We see that $M(p) = \left\{ \begin{pmatrix} 1 & b \\ 0 & p\end{pmatrix} :\ 0 \le b \le p - 1\right\} \bigcup \left\{ \begin{pmatrix} p & 0 \\ 0 & 1\end{pmatrix}\right\}$.

Suppose that $AD = p^\ell$ and $ad = p$ in (15.2). When $a = 1,\ d = p$ and $0 \le b \le p - 1$, we get the matrices $\begin{pmatrix} A & B + bD \\ 0 & pD\end{pmatrix}$. Now $B + bD$ runs through all residue classes mod pD, each exactly once; so this gives all of $M(p^{\ell+1})$ except those matrices of the form $\begin{pmatrix} p^{\ell+1} & 0 \\ 0 & 1\end{pmatrix}$. If $a = p, d = 1, b = 0$ and so we get the matrices $\begin{pmatrix} pA & pB \\ 0 & D\end{pmatrix}$. This includes $\begin{pmatrix} p^{\ell+1} & 0 \\ 0 & 1\end{pmatrix}$ once, and all the other such matrices are of the form $\begin{pmatrix} pa & pb \\ 0 & pd\end{pmatrix}$. But here $0 \le B < D = pd$ and $b \equiv B \pmod d$ with $0 \le b \le d$, so each such matrix is count p times. Hence *the set of matrices in $M(p)M(p^\ell)$ counting multiplicity, is given by the matrices in $M(p^{\ell+1})$ together with each of the matrices in $pM(p^{\ell-1})$, each p (more) times.*

`Exercise` 15.8.3. Prove that the set of matrices in $M(p^a)M(p^b)$ counting multiplicity, is given by the

union, over c in the range $0 \le c \le \min\{a,b\}$, of p^c copies of the matrices in $p^c M(p^{a+b-2c})$. Deduce that $M(p^a)M(p^b) = M(p^b)M(p^a)$.

15.9. Hecke operators. The *Hecke operator* $T(N)$ on a modular function f of weight k is defined by

$$T(N)f(z) := N^{k-1} \sum_{\substack{ad=N \\ 0\le b<d}} d^{-k} f\left(\frac{az+b}{d}\right) = N^{k-1} \sum_{M\in M(N)} d(M)^{-k} f(Mz),$$

where $Mz := \frac{az+b}{d}$ and $d(M) = d$. Note that $T(1)f = f$. We just saw that if $(m,n) = 1$ then $M(m)M(n) = M(mn)$, and so

$$T(mn)f = T(m)T(n)f = T(n)T(m)f \quad \text{whenever } (m,n) = 1. \tag{15.3}$$

We also saw that the set of matrices in $M(p)M(p^\ell)$ counting multiplicity, is given by the matrices in $M(p^{\ell+1})$ together with each of the matrices in $pM(p^{\ell-1})$, each p (more) times, for all $\ell \ge 1$ which yields

$$T(p)T(p^\ell)f = T(p^{\ell+1})f + p^{k-1}T(p^{\ell-1})f \tag{15.4}$$

since $(pM)z = Mz$ but $d(pM)^{-k} = (pd(M))^{-k}$.

`Exercise` 15.9.1. Verify (15.4) by the method suggested in the text.

`Exercise` 15.9.2. Prove that the $T(n)$ commute. (Hint: Use exercise 15.8.3 and (15.3).)

By definition $T(n)f$ is analytic on $\mathcal{H}$.

`Exercise` 15.9.3. Prove that (i) $T(n)f(z+1) = T(n)f(z)$, and (ii) $T(n)f(-1/z) = z^k T(n)f(z)$. (Hint: Use exercise 15.8.2.)

Hence $T(n)f$ is a modular form of weight k.

Now $e^{2i\pi Mz} = e^{2i\pi b/d} q^{a/d}$ and so if $f(z) = \sum_{n\ge 0} c(n)q^n$ then for given a, d,

$$\sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right) = \sum_{n\ge 0} c(n) \sum_{b=0}^{d-1} e^{2i\pi bn/d} q^{an/d} = d \sum_{m\ge 0} c(dm) q^{am}$$

since the only non-zero terms are where $d|n$ so we write $n = dm$. Hence

$$T(N)f(z) = \sum_{ad=N} a^{k-1} \sum_{m\ge 1} c(dm) q^{am} = \sum_{r\ge 0} \left(\sum_{a|(N,r)} a^{k-1} c(rN/a^2) \right) q^r, \tag{15.5}$$

and so $T(N)f$ is a cusp form if and only if f is. Thus $T(N)f(z) = \sigma_{k-1}(N)c(0) + c(N)q + \dots$

We will study a very special class of $f(z)$; those that are eigenforms for every $T(n)$; that is, there exists a constant $\lambda(n)$ such that

$$T(n)f = \lambda(n)f \text{ for all } n \ge 1.$$

In this case we have $\sigma_{k-1}(n)c(0)+c(n)q+\ldots = T(n)f = \lambda(n)f = \lambda(n)c(0)+\lambda(n)c(1)q+\ldots$ In particular $c(n) = \lambda(n)c(1)$ for all $n \geq 1$, and so we may assume $c(1) \neq 0$ else f is a constant. Now we can multiply an eigenform through by a scalar multiple and so we may assume that $c(1) = 1$. Hence $c(n) = \lambda(n)$ for all n, including $c(1) = \lambda(1) = 1$, and we can write

$$f(z) = c(0) + \sum_{n\geq 1} \lambda(n)q^n.$$

`Exercise` 15.9.2. (i) Prove that $\lambda(n)$ is a multiplicative function. (Hint: Use (15.3).) (ii) Also show that $\lambda(p)\lambda(p^\ell) = \lambda(p^{\ell+1}) + p^{k-1}\lambda(p^{\ell-1})$ for all $\ell \geq 1$.

There is a standard transformation in analysis (the Mellin transform) that sends series of the form $\sum_{n\geq 1} a(n)q^n$ to Dirichlet series $\sum_{n\geq 1} a(n)/n^s$. Since $\lambda(.)$ is multiplicative we see that our eigenform, less its constant coefficient, gets mapped to

$$\sum_{n\geq 1} \frac{\lambda(n)}{n^s} = \prod_{p \text{ prime}} \left(1 + \frac{\lambda(p)}{p^s} + \frac{\lambda(p^2)}{p^{2s}} + \ldots\right) = \prod_{p \text{ prime}} \left(1 - \frac{\lambda(p)}{p^s} + \frac{p^{k-1}}{p^{2s}}\right)^{-1},$$

since $\left(1 + \frac{\lambda(p)}{p^s} + \frac{\lambda(p^2)}{p^{2s}} + \ldots\right)\left(1 - \frac{\lambda(p)}{p^s} + \frac{p^{k-1}}{p^{2s}}\right) = 1$ by exercise 15.9.2(ii).

Now if $c(0) \neq 0$ then $c(n) = \lambda(n) = \sigma_{k-1}(n)$ for all n, and so

$$f(z) = c(0) + \sum_{n\geq 1} \sigma_{k-1}(n)q^n = \frac{(-1)^{k/2}B_k}{2k}\, E_k(\tau) + c'$$

for some constant c'. In this case $\lambda(p) = p^{k-1} + 1$ so that $1 - \lambda(p)t + p^{k-1}t^2 = (1-t)(1 - p^{k-1}t)$, and therefore

$$\sum_{n\geq 1} \frac{\lambda(n)}{n^s} = \prod_{p \text{ prime}} \left(\left(1 - \frac{1}{p^s}\right)\left(1 - \frac{p^{k-1}}{p^s}\right)\right)^{-1} = \zeta(s)\zeta(s-k+1).$$

We are more interested in the cases where $c(0) = 0$, the *eigencusp forms*.

One can show that there is a basis of eigenforms for each cusp space (and that these eigenforms have real coefficients). We sketch a proof: By exercise 15.3.1(iv) we know that if $z = x + iy$ then $f(z)y^{k/2-1}dy$ is invariant under the action of $\mathrm{PSL}(2,\mathbb{Z})$; moreover this is bounded on $\mathcal{H}$. We define the *Petersson inner product* of two cusp forms of weight k as

$$\langle f, g\rangle = \int_D f(z)\overline{g(z)}y^{2k-2}dxdy.$$

This is a hermitian scalar product on the cusp forms, and is both positive and non-degenerate. Moreover

$$\langle T(n)f, g\rangle = \langle f, T(n)g\rangle$$

which means that the $T(n)$ are hermitian operators with respect to the Petersson inner product. Since the $T(n)$ commute with one another one can show that there is an orthogonal basis for the cusp space, made up of eigenvectors of the $T(n)$. Moreover the eigenvalues, $\lambda(n)$, of the $T(n)$ must all be real numbers.

One can say more about the numbers $\lambda(n)$: One sees that if f has integer coefficients then so does $T(n)f$, by (15.5), for all $n \geq 1$. This implies that the coefficients of the characteristic polynomial of $T(n)$ (acting on the modular forms of weight k) are integers, and hence the eigenvalues, $\lambda(n)$, of the $T(n)$ must all be algebraic integers.

We saw above that the term in the Euler product for the Dirichlet series formed by the $\lambda(n)$, for the prime p, is $\left(1- \frac{\lambda(p)}{p^s} + \frac{p^{k-1}}{p^{2s}}\right)$. Taking $T = p^{-s}$ we can write this as $1 - \lambda(p)T + p^{k-1}T^2 = (1 - \alpha_p T)(1 - \beta_p T)$ for some algebraic numbers α_p, β_p In 1973, Deligne proved that

$$|\alpha_p| = |\beta_p| = p^{\frac{k-1}{2}}$$

(a weak form of this had been conjectured by Ramanujan, for the coefficients of $(2\pi)^{-12}\Delta(\tau)$).

15.10. The Mellin transform and the construction of L-functions. It is interesting to make a power series out of the coefficients of a given Dirichlet series. Hence for $\zeta(s)$, whose coefficients are all 1's we obtain $\sum_{n\geq 1} t^n = t/(1-t)$. For a Dirichlet L-function $L(s,\chi)$ we have, using the periodicity of $\chi \pmod q$, and so writing $n = qr + m$

$$\sum_{n\geq 1} \chi(n)t^n = \sum_{m=1}^{q} \sum_{r\geq 0} \chi(m)t^{qr+m} = \frac{\sum_{m=1}^{q} \chi(m)t^m}{1-t^q},$$

a rational function. This is not quite as ad hoc a procedure as it seems at first sight since by defining[34]

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}dt$$

for $\mathrm{Re}(s) > 0$ we have, by changing variable $t \to nt$, $\Gamma(s) = n^s \int_0^\infty e^{-nt}t^{s-1}dt$ and so

$$\Gamma(s)L(s,\chi) = \sum_{n\geq 1} \chi(n) \int_0^\infty e^{-nt}t^{s-1}dt = \int_0^\infty \frac{\sum_{m=1}^{q} \chi(m)e^{-mt}}{1-e^{-qt}}\, t^{s-1}dt.$$

`Exercise` 15.10.1. Use this expression to provide an analytic continuation for $L(s,\chi)$ for all $\mathrm{Re}(s) > 0$.

Now consider the Fourier expansion of a cusp form, say $f(\tau) = \sum_{n\geq 1} c_n q^n$. We define

$$L(f,s) := \sum_{n\geq 1} \frac{c_n}{n^s}\ .$$

[34] $\Gamma(s)$ is the function that extrapolates $n!$, so that $\Gamma(n+1) = n!$. Because of this it is involved in many beautiful combinatorial formulas many of which stem from

$$\frac{1}{s\Gamma(s)} = e^{\gamma s} \prod_{n=1}^{\infty} \left(1+ \frac{s}{n}\right) e^{-s/n}.$$

Then, following the above, we have

$$\Gamma(s)L(f,s) = \sum_{n\geq 1} c_n \int_0^\infty e^{-nt}t^{s-1}dt = (2\pi)^s \int_0^\infty \sum_{n\geq 1} c_n e^{-2\pi nt}t^{s-1}dt$$
$$= (2\pi)^s \int_0^\infty f(it)t^{s-1}dt,$$

as $f(it) = \sum_{n\geq 1} c_n e^{-2\pi nt}$. This converges for all s and so provides an analytic continuation for $\Gamma(s)L(f,s)$, and hence for $L(f,s)$.

In this case, though we can go further: Change variables $t = 1/u$, and notice that $i/u = -1/(iu)$ so that $f(i/u) = f(-1/(iu)) = (iu)^k f(iu)$, so that

$$(2\pi)^{-s}\Gamma(s)L(f,s) = \int_0^\infty f(i/u)u^{-1-s}du = \int_0^\infty (iu)^k f(iu)u^{-1-s}du$$
$$= i^k \int_0^\infty f(iu)u^{k-s-1}du = (-1)^{k/2}(2\pi)^{-(k-s)}\Gamma(k-s)L(f,k-s).$$

This is the *functional equation* for the *completed* L-function $\Lambda(f,s) := (2\pi)^{-s}\Gamma(s)L(f,s)$; that is $\Lambda(f,s) = (-1)^{k/2}\Lambda(f,k-s)$, a "symmetry" about the line $s = k/2$.

It is useful to know when the sum defining $L(f,s)$ is absolutely convergent. In the case of the Eisenstein forms this is not difficult: First recall that the coefficients are multiplicative, and note that $\sigma_k(p^a) = p^{ak} + p^{(a-1)k} + \ldots + 1$ so that $1 \leq \sigma_k(p^a)/p^{ak} \leq 1 + 1/p^k + 1/p^{2k} + \ldots$. Hence $1 \leq \sigma_k(n)/n^k \leq \zeta(k)$. We deduce that $\sum_{n\geq 1} \sigma_{k-1}(n)/n^s$ is absolutely convergent provided $\mathrm{Re}(s) > k$.

Now suppose that $f(z)$ is a cusp form of weight k. By exercise 15.3.1(v) we know that $|y^{k/2}f(z)|$ is invariant under $\mathrm{PSL}(2,\mathbb{Z})$. From the Fourier expansion $f(z) = \sum_{n\geq 1} a_n q^n$ with $q = e^{2i\pi z}$ we know that if $z = x+iy$ then $|f(z)| \ll |q| = e^{-2\pi y}$, and hence $|y^{k/2}f(z)|$ is bounded throughout D (and thus throughout $\mathcal{H}$), say with maximum M. Now to pick out the coefficient of q^n from a power series, there is a standard trick from complex analysis:

$$a_m = \frac{1}{2i\pi}\int_{|q|=r} q^{-m-1}\sum_{n\geq 1} a_n q^n dq$$

provided $\sum_{n\geq 1} a_n q^n$ is analytic on the whole of $|q| \leq r$. To create such a circle we let x vary from 0 to 1, for y satisfying $r = e^{-2\pi y}$, and thus

$$a_m = \int_{x=0}^1 q^{-m} f(z)dx$$

since $dq = 2i\pi q dx$, Hence $|a_m| \leq \max_{0\leq x\leq 1}|q^{-m}f(z)| \leq e^{2\pi my}My^{-k/2}$, for any $y > 0$. We select $y = 1/m$ to obtain $|a_n| \leq e^{2\pi}Mn^{k/2} \ll n^{k/2}$ for all n. Notice that this is much smaller than the coefficients of the Eisenstein series of the same weight. We deduce that $L(f,s)$ is absolutely convergent for $s > k/2 + 1$.

Deligne (1973) has improved this bound to $a_n \ll n^{k-1/2}\sigma_0(n)$. Hecke showed that each such $L(f,s)$ may be extended analyticually to a meromorphic function on all of $\mathbb{C}$, and is analytic if f is a cusp form. We can use the above functional equation to determine $L(f,s)$ when $\mathrm{Re}(s) < k/2$, from the values of $L(f,s)$ when $\mathrm{Re}(s) > k/2$

15.11. Congruence subgroups. We define the subgroup $\Gamma(N)$ of $\mathrm{PSL}(2,\mathbb{Z})$ to be the set of matrices in $\mathrm{PSL}(2,\mathbb{Z})$ that are $\equiv I \pmod N$. The *congruence subgroups* G of level N are those for which $\Gamma(N) \subset G \subset \mathrm{PSL}(2,\mathbb{Z})$, and that this is not true for any smaller N.[35] The most useful is

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z};\quad N|c \ \text{ and } \ ad-bc=1 \right\}.$$

The *modular forms of level N and weight k* are those analytic f which satisfy the functional equation

$$f\left(\frac{az+b}{cz+d}\right) = (\det M)^{-k/2}(cz+d)^k f(z)$$

for all $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$. We need to f to be analytic at all of its cusps (i.e. at all of the images of ∞ under $\mathrm{PSL}(2,\mathbb{Z})/\Gamma_0(N)$ which is a finite group).

Now let $\Theta(\tau) = \sum_{n\in\mathbb{Z}} q^{n^2}$ where $q = e^{2i\pi n\tau}$, so that $\Theta(\tau) = \Theta(\tau+1)$. Now $\Theta(\tau) = \theta(2\tau)$ and so, by (10.2) we have $\Theta(-1/4\tau)^8 = (4)^{-2}(4\tau)^4\Theta(\tau)^8$. Hence $\Theta(\tau)^8$ is a modular form of level 4 and weight 4.[36]

Now in (10.3) we saw the example $f(\tau) = \eta(4\tau)^2\eta(8\tau)^2$ which, it turns out, is a modular form of level 32 and weight 2. Actually it is more, it is an eigenform (which explains why the coefficients are multiplicative) and a cusp form (which allows us to involve the Mellin transform). Using what we did there we see that we can construct the L function for the elliptic curve $E : y^2 = x^3 - x$, that is $L(E,s) = \sum_{n\ge1} a_n/n^s = L(f,s)$ where the a_n are given by (10.3), by

$$\Gamma(s)L(E,s) = \int_0^\infty e^{-t} \prod_{n\ge1} (1-e^{-4nt})^2(1-e^{-8nt})^2 t^{s-1} dt.$$

One can proceed for congruence subgroups in a similar way to what we did for $\mathrm{PSL}(2,\mathbb{Z})$. There are additional complications that are best understood in a course on elliptic curves. In general the completed L-function for an eigencusp form is $\Lambda(f,s) := N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s)$, with functional equation $\Lambda(f,s) = \pm(-1)^{k/2}\Lambda(f,k-s)$, where the *root number* ±1 depends on aspects of the eigenspace.

Eichler and Shimura showed that for any eigencusp form f of weight 2 and level N, there exists an elliptic curve E, of conductor N,[37] such that $L(E,s) = L(f,s)$

The *modularity conjecture*[38] suggests that the converse is true: For any elliptic curve E defined over $\mathbb{Q}$ there exists an eigencusp form f_E of weight 2 and level N_E (where N_E

[35] It is worth remarking that there are subgroups of finite index of $\mathrm{PSL}(2,\mathbb{Z})$ which are not congruence subgroups.

[36] There is also a definition by which we can call $\Theta(\tau)$ itself a modular form of weight 1/2 though more care must be taken with the multiplier.

[37] The *conductor* is an invariant that divides the discriminant, as will be discussed in a footnote on the next page.

[38] First posed by Taniyama in 1955 as a vague question, and better formulated subsequently by Shimura and Weil (independently) as people better understood elliptic curves.

is the conductor of E) such that $L(f_E, s) = L(E, s)$. From what we saw in the previous section this allows us to analytically continue $L(E, s)$ to the whole complex plane, which is very useful. The modularity conjecture is now a theorem due to Andrew Wiles,[39] and this has many interesting consequences (like Fermat's Last Theorem, as we shall discuss).

As we have seen, the modular forms of a given weight and level form a vector space, and are fairly straightforward to construct. Hence if we are given an elliptic curve, we can compute its conductor, and then search the eigencusp forms of weight 2 and level N_E to find one for which $L(f, s) = L(E, s)$. This is a finite calculation and thus makes the modularity conjecture easily checkable.[40]

16-17. More stuff to add on elliptic curves, etc

16.1. The Birch Swinnerton-Dyer conjecture. We seek to develop the heuristic of section 6.1 in the context of elliptic curves: We want to count the rational points on $E: \ y^2 = x^3 + ax + b$ up to a given height. In exercise 12.1.1 we saw that any solution may be written as $x = m/n^2, \ y = \ell/n^3$ with $(m, n) = 1$, so that

$$m^3 + amn^4 + bn^6 = \ell^2.$$

Let $S(N) = \#\{m, n \in \mathbb{Z} : \ |m^3 + amn^4 + bn^6| \le N\}$. Then

$$S(N) \approx \mathrm{Vol}\{(m, n) \in \mathbb{R}^2 : \ 1 \le m^3 + amn^4 + bn^6 \le N\} = \kappa_E N^{1/2}$$
$$\text{where } \kappa_E := \mathrm{Vol}\{(m, n) \in \mathbb{R}^2 : \ 0 < m^3 + amn^4 + bn^6 \le 1\}$$

using the change of variable $m \to N^{1/3}m, \ n \to N^{1/6}n$. A random integer close to x is a square with probability around $1/\sqrt{x}$. So we guesstimate that the number of solutions to $\ell^2 = m^3 + amn^4 + bn^6$ with $|m^3 + amn^4 + bn^6| \le N$ is roughly

$$2 \sum_{\substack{1 \le m^3+amn^4+bn^6 \le N \\ (m,n) \ne (0,0)}} \frac{1}{\sqrt{m^3 + amn^4 + bn^6}} = \ 2 \int_1^N \frac{dS(t)}{\sqrt{t}} \approx \ \kappa_E \int_1^N \frac{dt}{t} \approx \ \kappa_E \log N.$$

`Exercise` 16.1.1. Writing $m = n^2 t$ or otherwise, prove that

$$\kappa_E = \frac{2}{3} \Omega(E) \text{ where } \Omega(E) := \int_{t \in \mathbb{R}: \ t^3+at+b \ge 0} \frac{dt}{\sqrt{t^3 + at + b}} = \int_{(x,y) \in E(\mathbb{R})} \frac{dx}{|y|}.$$

---

[39] Although Wiles made the key breakthrough, and thus proved the conjecture in a wide variety of cases, the whole conjecture was subsequently resolved by ....

[40] We do need to discuss what the conductor is. One explanation: Out of all possible Weierstrass equations, there is a 'best' one (not unique) called a global minimal Weierstrass equation in which the discriminant is minimized. We say that E has *good reduction* at p if E over $\mathbb{F}_p$ is non-singular (that is $p \not| \Delta_E$). Also $E: y^2 = f(x)$ has *multiplicative reduction* at p if $f(x)$ has just two distinct roots in $\overline{\mathbb{F}}_p$; and *additive reduction* at p if $f(x)$ has just one distinct root in $\overline{\mathbb{F}}_p$. It can be that an elliptic curve has bad reduction over $\mathbb{Q}$, but good reduction over some field extension, in which case it has *potential good reduction*. If $p \ne 2, 3$ then the power of p dividing N_E is given by $3 - \#\{$ distinct roots of $f(x)$ in $\overline{\mathbb{F}}_p\}$. To understand what happens at the primes 2 and 3 one has to be more careful than we have been in defining the minimal model.

We need to take into account p-divisibility just as in section 6.1. So instead of the implicit assumption that there p^2 solutions $(\ell, m, n) \pmod p$ to $\ell^2 \equiv m^3 + amn^4 + bn^6 \pmod p$ with $p \not| (m, n)$ (so that $(m, n) = 1$), we must count them correctly: If $n \equiv 0 \pmod p$ then $\ell^2 \equiv m^3 \pmod p$ with $m \not\equiv 0 \pmod p$ so the solutions are $(t^3, t^2, 0) : \ 1 \leq t \leq p-1$. If $n \not\equiv 0 \pmod p$, let $y \equiv \ell/n^3, \ x \equiv m/n^2 \pmod p$ so we get $p-1$ times the number of solutions to $y^2 \equiv x^3 + ax + b \pmod p$ (that is $(n^3 y, n^2 x, n) : \ 1 \leq n \leq p-1$), which is $(p-1)(\#E(\mathbb{F}_p) - 1)$ (the "-1" since we must subtract off the point at ∞). Hence the total number of solutions is $(p-1)\#E(\mathbb{F}_p)$, and hence, in analogy with section 6.1, we must correct our guesstimate by the product of the local factors:

$$\prod_p \left(1 - \frac{1}{p}\right)\left(1 - \frac{a_p}{p} + \frac{p}{p^2}\right)$$

using the results of section 14.7 with $a_p = \alpha_p + \overline{\alpha}_p$. However we know that a_p is twice the real part of an algebraic integer of size $\sqrt{p}$, so we might expect that it typically has size about a constant times $\sqrt{p}$. Hence the above product is not likely to be convergent unless there is an enormous amount of cancelation (that is, the $\sum_{x<p<2x} a_p \log p$ is typically around size x whereas $\sum_{x<p<2x} |a_p| \log p$ is around size $x^{3/2}$). So in order to not assume that the product converges we really must truncate the product at N, that is work only with the primes $p \leq N$. Hence, since $\prod_p \left(1 - \frac{1}{p}\right) \asymp \frac{1}{\log N}$, we expect that the number of points on our elliptic curve up to N is

$$\asymp \Omega(E) \prod_{p \leq N} \frac{\#E(\mathbb{F}_p)}{p} \ .$$

If we replace p by p^s in the above (complete) product then we have

$$\prod_p \left(1 - \frac{a_p}{p^s} + \frac{p}{p^{2s}}\right) \ = \ \frac{1}{L(E, s)} \ .$$

It is "well known" that the value of an Euler product at $s = 1$ truncated at $p \leq N$, is well-approximated by the value of the associated zeta-function at $1 + \frac{1}{\log N}$. Hence we need to understand the order of vanishing of $L(E, s)$ at $s = 1$: Suppose that we have the Taylor series

$$L(E, s) = c_E (s-1)^r + c'_E (s-1)^{r+1} + \dots$$

at $s = 1$, where r is chosen so that $c_E \neq 0$ (and therefore $c_E = L^{(r)}(E, 1)/r!$) so that the expected number of points on our elliptic curve is

$$\asymp \frac{\Omega(E)}{c_E} \cdot (\log N)^r.$$

Based on such a heuristic and comparing this with the square of the estimate obtained in Proposition 13.9,[41] and various data, Birch and Swinnerton-Dyer made two conjectures. The first is:

[41] I do not see why we should take the square except that it works beautifully!

Birch Swinnerton-Dyer conjecture, I. *The order of the zero at $s = 1$ of $L(E, s)$ equals the rank of $E(\mathbb{Q})$.*

We define the *algebraic rank* of E to be the rank of $E(\mathbb{Q})$, and the *analytic rank* of E to be the order of the zero at $s = 1$ of $L(E, s)$. The Birch Swinnerton-Dyer conjecture, part I, predicts that these are equal.

Notice that if we compare two elliptic curves E and E', for which $\#E(\mathbb{F}_p)$ is one larger on average than $\#E'(\mathbb{F}_p)$; then $a_p(E)$ is one smaller on average than $a_p(E')$, and so the order of vanishing of $L(E, s)$ is one larger than the order of vanishing of $L(E', s)$, hence rank$(E(\mathbb{Q}))$ =rank$(E'(\mathbb{Q})) + 1$. That is if there are more points on average modulo all primes, then there are more points over the rationals. If one uses some deeper ideas of analysis one can come up with the following conjecture that is more-or-less equivalent to the Birch Swinnerton-Dyer conjecture, I:

$$\text{rank}(E(\mathbb{Q})) = \text{Mean}_p\ (\#E(\mathbb{F}_p) - p - 1) - \frac{1}{2}\,. \tag{16.1.1}$$

(To be precise we need to say what we mean by "Mean$_p$".) Further comparison of our heuristic here with the square of Proposition 13.9 leads one to surmise that

$$\frac{L^{(r)}(E,1)}{r!} \approx_r \frac{\Omega(E)|\text{Regulator}(E(\mathbb{Q}))|}{|\text{Torsion}(E(\mathbb{Q}))|^2}\,.$$

As in section 6.1, we again need to take account of the appropriate class group (called the *Tate-Shafarevic group* and denoted Ш(E)), and various Tamagawa numbers $c_p(E)$.

Birch Swinnerton-Dyer conjecture, II. *If the order of the zero at $s = 1$ of $L(E, s)$ is r then*

$$\frac{L^{(r)}(E,1)}{r!} = \prod_p c_p(E) \cdot \frac{|Ш(E)|\Omega(E)|\text{Regulator}(E(\mathbb{Q}))|}{|\text{Torsion}(E(\mathbb{Q}))|^2}\,.$$

The $c_p(E), \Omega(E)$ and Torsion$(E(\mathbb{Q}))$ can be easily determined, and therefore $\delta(E) := \Omega(E)\prod_p c_p(E)/|\text{Torsion}(E(\mathbb{Q}))|^2$. So one can view this formula as linking an L-function value with $|Ш(E)| \cdot |\text{Regulator}(E(\mathbb{Q}))|$ just like Dirichlet's class number formula. Moreover when the rank is zero, then the regulator equals 1, and so we have $L(E, 1) = \delta(E)|Ш(E)|$ (analogous to imaginary quadratic fields).

The Birch Swinnerton-Dyer conjectures have been proved only in certain special cases:

• In 1976 Coates and Wiles proved that if E has complex multiplication and analytic rank zero, then it has algebraic rank zero. In 1991 Rubin went on to show that the p-parts of the predicted formula $L(E, 1) = \delta(E)|Ш(E)|$ are then correct for all primes $p \geq 11$.

• In 1983 Gross and Zagier showed[42] that if E has analytic rank one then its algebraic rank is ≥ 1. In 1990 Kolyvagin improved this by showing that if E has analytic rank 0 or 1, then its algebraic rank is the same.

• In 2011 Bhargava and Shankar showed that a positive proportion of elliptic curves have analytic rank zero, and hence, by Kolyvagin's result, part I of the Birch Swinnerton-Dyer conjecture holds for at least a positive proportion of elliptic curves.

[42]They showed this under the assumption that E is modular, which we now know to be true for all curves over $\mathbb{Q}$.

16.2. More easy modularity. As usual define $\eta(\tau) = q^{1/24}\prod_{n\geq 1}(1-q^n)$. There are twelve elliptic curves for which $f_E(\tau)$ is a product and quotient of η-functions (as shown by Martin and Ono). They are

Conductor	η-quotient	Elliptic curve
11	$\eta^2(\tau)\eta^2(11\tau)$	$y^2+y=x^3-x^2-10x-20$
14	$\eta(\tau)\eta(2\tau)\eta(7\tau)\eta(14\tau)$	$y^2+xy+y=x^3+4x-6$
15	$\eta(\tau)\eta(3\tau)\eta(5\tau)\eta(15\tau)$	$y^2+xy+y=x^3+x^2-10x-10$
20	$\eta^2(2\tau)\eta^2(10\tau)$	$y^2=x^3+x^2+4x+4$
24	$\eta(2\tau)\eta(4\tau)\eta(6\tau)\eta(12\tau)$	$y^2=x^3-x^2-4x+4$
27	$\eta^2(3\tau)\eta^2(9\tau)$	$y^2+y=x^3-7$
32	$\eta^2(4\tau)\eta^2(8\tau)$	$y^2=x^3+4x$
36	$\eta^4(6\tau)$	$y^2=x^3+1$
48	$\frac{\eta^4(4\tau)\eta^4(12\tau)}{\eta(2\tau)\eta(6\tau)\eta(8\tau)\eta(24\tau)}$	$y^2=x^3-4x-4$
64	$\frac{\eta^8(8\tau)}{\eta^2(4\tau)\eta^2(16\tau)}$	$y^2=x^3-4x$
80	$\frac{\eta^6(4\tau)\eta^6(20\tau)}{\eta^2(2\tau)\eta^2(8\tau)\eta^2(10\tau)\eta^2(40\tau)}$	$y^2=x^3-x^2+4x-4$
144	$\frac{\eta^{12}(12\tau)}{\eta^4(6\tau)\eta^4(24\tau)}$	$y^2=x^3--1$

TABLE . The only η-products giving L-functions of elliptic curves.

We can use various identities for η-quotients. In sections 10.5 and 10.6 we saw (in the order we found them) that

$$\frac{\eta^5(2\tau)}{\eta^2(\tau)\eta^2(4\tau)} = \sum_{m\in\mathbb{Z}} q^{m^2};$$

$$\frac{\eta^2(16\tau)}{\eta(8\tau)} = \sum_{m\geq 0} q^{(2m+1)^2}.$$

$$\frac{\eta^2(\tau)}{\eta(2\tau)} = \sum_{m\in\mathbb{Z}} (-1)^m q^{m^2};$$

$$\eta(24\tau) = \sum_{m\geq 1} \left(\frac{12}{m}\right) q^{m^2}.$$

$$\eta^3(8\tau) = \sum_{\substack{a\geq 1\\ a \text{ odd}}} (-1)^{\frac{a-1}{2}} a q^{a^2}.$$

and we add

$$\frac{\eta^5(6\tau)}{\eta^2(3\tau)} = \sum_{m\geq 1} (-1)^{m-1}\left(\frac{m}{3}\right) m q^{m^2}$$

`Exercise` Prove that

$$\frac{\eta^2(\tau)}{\eta(2\tau)} + 4\,\frac{\eta^2(16\tau)}{\eta(8\tau)} = \frac{\eta^5(2\tau)}{\eta^2(\tau)\eta^2(4\tau)}.$$

Exercise Establish that $\frac{1}{2i\pi}\frac{\eta'(\tau)}{\eta(\tau)} = \frac{1}{24} - \sum_{n\geq 1}\frac{nq^n}{1-q^n}$ by using the chain rule. Replacing τ by 4τ, and multiplying through by 4, and then subtracting from the above, and using exercise 7.5.4, deduce that

$$-1- \frac{4}{i\pi}\left(\frac{\eta'(\tau)}{\eta(\tau)} - \frac{\eta'(4\tau)}{\eta(4\tau)}\right) = 8\sum_{\substack{n\geq 1\\ 4\not| n}}\frac{nq^n}{1-q^n} = \left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4$$

Hence prove that

$$\frac{\eta^{20}(2\tau)}{\eta^8(\tau)\eta^8(4\tau)} = \frac{4i}{\pi}\left(\frac{\eta'(\tau)}{\eta(\tau)} - \frac{\eta'(4\tau)}{\eta(4\tau)}\right) - 1$$

Exercise Show the following: If $p \equiv 3 \pmod 4$ then $a_{32}(p), a_{64}(p) = 0$. If $p \equiv 1 \pmod 4$ we write $p = a^2 + b^2$ where $a, b \geq 1$ and a is odd. Then $a_{32}(p) = 2(-1)^{\frac{a+b-1}{2}}a$ and $a_{64}(p) = 2(-1)^{\frac{a-1}{2}}a$

Exercise Show the following: If $p \equiv 2 \pmod 3$ then $a_{27}(p), a_{36}(p), a_{144}(p) = 0$. If $p \equiv 1 \pmod 3$ we write $p = m^2 + 3n^2$ where, if $3\not| n$ we choose the signs of m and n so that $m \equiv n \equiv 2 \pmod 3$ and then $a_{27}(p) = m + 3n$. If $3|n$ we choose the sign of m so that $m \equiv 1 \pmod 3$ and then $a_{27}(p) = 2m$. We also have $a_{36}(p) = 2\left(\frac{m}{3}\right)m$, $a_{144}(p) = 2(-1)^{m-1}\left(\frac{m}{3}\right)m$.

In section 15.10 we saw that if f is a cusp form then $\Gamma(s)L(f,s) = (2\pi)^s\int_0^\infty f(it)t^{s-1}dt$, and so, in particular $L(f,1) = 2\pi\int_0^\infty f(it)dt$. Now $\eta(it) = e^{-2\pi t/24}\prod_{n\geq 1}(1-e^{-2\pi nt}) \in (0,1)$ for $t \in (0,\infty)$, and so if f is an η-quotient then $f(it) \in (0,1)$ for all $t \in (0,\infty)$. Hence $L(f,1) = 2\pi\int_0^\infty f(it)dt$ is a positive real number, and is therefore non-zero. By the Birch-Swinnerton Dyer conjecture we deduce that the associated elliptic curve has rank 0.

Congruences for modular forms and FLT
The Sato-Tate conjecture. Taylor's Theorem
The Congruent number problem
Deuring's lifting lemma – the number of elliptic curves with p+1-t points
Singular moduli and Gross-Zagier
Heegner points – of proof of class number 1 via Diophantine problem
Gross-Zagier fomula; Kolyvagin
Taniyama-Shimura conjecture
Comparing quadratic twists of a given elliptic curve and $\#E(\mathbb{F}_p)$ with $\#E_d(\mathbb{F}_p)$
Goldfeld's conjecture... and rank distribution

19. Integral points on elliptic curves

We change models of elliptic curves by linear changes of variables, which allows us to keep consistent our notion of rational points (as long as we are careful about points at infinity). However such transformations do not preserve integer points, making such questions a little more ad hoc, in the sense that the question depends on the choice of model for the elliptic curve.

Siegel showed that there are only finitely many integral points on any model of an elliptic curve, and indeed on any model of any curve that is not transformable to a linear equation. The proof is a little beyond us here, but note that if we had $x(x-1)(x+1) = y^2$ in integers, then either $x-1$ and x are squares, or $(x-1)/2$ and $(x+1)/2$ are squares, the only solution to consecutive squares being 0 and 1, and hence $x = 1, y = 0$. This proof generalizes but in an example like $x(2x+1)(3x+1) = y^2$ we see that there exist integers u, v, w such that $x = \pm u^2$, $2x+1 = \pm v^2$, $3x+1 = \pm w^2$. Hence $v^2 - 2u^2 = 3v^2 - 2w^2 = \pm 1$. Squaring the second solution gives $(3v^2 + 2w^2)^2 - 6(2vw)^2 = 1$, and so we get solutions to a *simultaneous Pell equation*. Since the solutions to one Pell equation are so sparse, it seems likely that there are few co-incidences between two.[43]

19.1. Taxicab numbers and other diagonal surfaces. When Ramanujan lay ill from pneumonia in an English hospital he was visited by G.H. Hardy, his friend and co-author. Struggling for conversation, Hardy remarked that the number, 1729, of the taxicab in which he had ridden from the train station to the hospital was extremely dull. Ramanujan contradicted him noting that it is the smallest number which is the sum of two cubes in two different ways:

$$1^3 + 12^3 = 9^3 + 10^3 = 1729.$$

(Ramanujan might also have mentioned that it is the third smallest Carmichael number!). There are many other such identities; indeed Euler showed that all solutions to

$$a^3 + b^3 = c^3 + d^3$$

can be obtained by scaling

$$a = r^4 + (p-3q)(p^2+3q^2)r, \;\; b = (p+3q)r^3 + (p^2+3q^2)^2,$$
$$c = r^4 + (p+3q)(p^2+3q^2)r, \;\; d = (p-3q)r^3 + (p^2+3q^2)^2.$$

How about $a^4 + b^4 + c^4 = d^4$? Euler conjectured that there are no non-trivial solutions, but in 1986 Elkies showed that there are infinitely many, the smallest of which is

$$95800^4 + 217519^4 + 414560^4 = 422481^4.$$

(It was rather lucky that this is just large enough to have avoided direct computer searches to that time, since Elkies was inspired to give his beautiful solution to this problem). Euler had even conjectured that there is no non-trivial solution to the sum of $n-1$ powers equalling an nth power, but that had already been disproved via the example

$$27^5 + 84^5 + 110^5 + 133^5 = 144^5.$$

[43] Bennett et. al. proved that there are never more than two solutions to $x^2 - az^2 = y^2 - bz^2 = 1$ for given integers $a > b \geq 1$. This cannot be improved since for any z we can select integers x and y such that $x^2 \equiv y^2 \equiv 1 \pmod{z^2}$ and then take $a = (x^2-1)/z^2$, $b = (y^2-1)/z^2$.

19.2. Sums of two cubes. Suppose that are studying rational solutions of $a^3 + b^3 = k$ ($\neq 0$). Writing $u = a + b$, $v = a - b$ and then $y = 36kv/u$, $x = 12k/u$ we get $y^2 = x^3 - 3(12k)^2$.

`Exercise 19.2.1.` Show that from every rational solution x, y to $y^2 = x^3 - 3(12k)^2$ we can obtain a rational solution a, b to $a^3 + b^3 = k$.

Hence we see that studying the sum of two cubes is also a problem about elliptic curves. We have seen that 1729 is the smallest integer that can be represented in two ways. Are there integers that can be represented in three ways, or four ways, or...? Actually this is not difficult to answer: $1^3 + 12^3 = 9^3 + 10^3 = 1729$. Using the doubling process on the cubic curve $a^3 + b^3 = 1729$:

$$\text{If } P = (a, b) \text{ then } 2P = (A, B) \text{ then } A = a\ \frac{a^3 - 3458}{1729 - 2a^3} \text{ and } B = b\ \frac{a^3 + 1729}{1729 - 2a^3}\ .$$

So, starting from the solution $(12, 1)$, we get further solutions $(20760/1727, -3457/1727)$, $(184026330892850640/15522982448334911, 61717391872243199/15522982448334911)$, and the next solution is pointless to write down since each ordinate has seventy digits! The main point is that there are infinitely many different solutions, let us write them as $(u_i/w_i, v_i/w_i)$, $i = 1, 2, \ldots$ with $w_1|w_2|\ldots$.[44] Hence we have N solutions to $a^3 + b^3 = 1729w_N^3$ taking $a = u_i(w_N/w_i)$ and $b = v_i(w_N/w_i)$.

This scaling up of rational points seems like a bit of a cheat, so let's ask whether there exists an integer m that can be written in N ways as the sum of two cubes of coprime integers? People have found examples for $N = 3$ and 4 but not beyond, and this remains an open question.

[44] As we saw when discussing the proof of the Lutz-Nagell Theorem.

21. Diophantine equations in polynomials

We have already seen that Diophantine equations do not easily have solutions in polynomials once the degree is large enough. The key result in section 11.1, obtained using elementary calculus, was:

The *abc* Theorem for Polynomials. *If $a(t), b(t), c(t) \in \mathbb{C}[t]$ do not have any common roots and provide a genuine polynomial solution to $a(t) + b(t) = c(t)$, then the maximum of the degrees of $a(t), b(t), c(t)$ is less than the number of distinct roots of $a(t)b(t)c(t) = 0$.*

`Exercise 21.1.` Deduce that if $x(t), y(t), z(t) \in \mathbb{C}[t]$ do not have any common roots, but have degrees ≥ 1, and provide a solution to to $x(t)^n + y(t)^n = z(t)^n$ then $n = 1$ or 2. Exhibit solutions for $n = 1$ and 2.

`Exercise 21.2.` Deduce that if $x(t), y(t), z(t) \in \mathbb{C}[t]$ do not have any common roots, but have degrees ≥ 1, and provide a solution to to $x(t)^p + y(t)^q = z(t)^r$ then $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$.

Lemma 21.1. *There do not exist polynomials $x(t), y(t) \in \mathbb{C}[t]$ of degree ≥ 1, such that four non-proportional linear combinations of $x(t)$ and $y(t)$ are each squares.*

Proof. Suppose there are such polynomials, and take an example of smallest degree. Suppose that two of the linear combinations are $u(t)^2$ and $v(t)^2$; evidently they cannot share any roots, else we could divide through by their common factor and have an example of smaller degree. Since all constants are a square in $\mathbb{C}[t]$, we can multiply u and v by constants, so that $u(t)^2,\ v(t)^2,\ u(t)^2 - v(t)^2$ and $u(t)^2 - m^2v(t)^2$ are all squares in $\mathbb{C}[t]$, for some $m \in \mathbb{C}$.

Now $(u(t)-v(t), u(t)+v(t))_{\mathbb{C}[t]} = (u(t)-mv(t), u(t)+mv(t))_{\mathbb{C}[t]} = 1$ since any common factors would also divide $(u(t), v(t)) = 1$. Hence each of $u(t) - v(t),\ u(t) + v(t),\ u(t) - mv(t),\ u(t)+mv(t)$ are squares in $\mathbb{C}[t]$, which yields a contradiction since these have lower degree than the original example.

Theorem 21.2. *Let $f(X) \in \mathbb{C}[X]$ have degree $d \geq 3$, and have no repeated roots. There are no solutions in non-constant rational functions $x(t), y(t) \in \mathbb{C}(t)$ to $y^2 = f(x)$.*

Proof. Suppose that we have a solution with $x(t) = a(t)/b(t)$ and $y(t) = u(t)/v(t)$, where $(a(t), b(t))_{\mathbb{C}[t]} = (u(t), v(t))_{\mathbb{C}[t]} = 1$. Writing $f(x) = c\prod_{i=1}^{d}(x - \alpha_i)$, we have

$$u(t)^2b(t)^d = cv(t)^2\prod_{i=1}^{d}(a(t) - \alpha_i b(t)).$$

Now $v(t)^2|u(t)^2b(t)^d$ and $(v, u) = 1$ so that $v(t)^2|b(t)^d$. On the other hand $b(t)^d|v(t)^2\prod_{i=1}^{d}(a(t)-\alpha_i b(t))$. Now $(b(t), a(t) - \alpha_i b(t)) = (b(t), a(t)) = 1$ for each i and so $b(t)^d|v(t)^2$. Putting these two criterion together we deduce that

$$v(t)^2 = \kappa b(t)^d \tag{21.1}$$

for some $\kappa \in \mathbb{C}$, and

$$u(t)^2 = c\kappa\prod_{i=1}^{d}(a(t) - \alpha_i b(t)). \tag{21.2}$$

Now $(a(t) - \alpha_i b(t), a(t) - \alpha_j b(t))$ contains $(\alpha_i - \alpha_j)b(t)$ and $(\alpha_i - \alpha_j)a(t)$, and therefore $\alpha_i - \alpha_j$. Hence the terms the product on the right side of (21.2) are pairwise coprime over $\mathbb{C}[t]$, and so, in particular,

$$a(t) - \alpha_i b(t) \in \mathbb{C}[t]^2 \quad \text{for} \quad i = 1, 2, \ldots, d.$$

This contradicts Lemma 21.1 when $d \geq 4$. When $d = 3$ we get the additional equation $b(t) = (v(t)/b(t))^2$ from (21.1), and so again contradict Lemma 21.1.

22. Fermat's Last Theorem

For a very long time Fermat's Last Theorem was the best known and most sought after open question in number theory. It inspired the development of much great mathematics, in many different directions.

Proposition 22.1. *If Fermat's Last Theorem is false then there exists an odd prime p and pairwise coprime non-zero integers x, y, z such that*

$$x^p + y^p + z^p = 0.$$

Hence, to prove Fermat's Last Theorem, one can restrict attention to odd prime exponents.

Proof. Suppose that $x^n + y^n = z^n$ with $x, y, z > 0$ and $n \geq 3$. If two of x, y have a common factor then it must divide the third and so we can divide out the common factor. Hence we may assume that x, y, z are pairwise coprime positive integers. Now any integer $n \geq 3$ has a factor m which is either $= 4$ or is an odd prime. Hence, if $n = dm$ then $(x^d)^m + (y^d)^m = (z^d)^m$, so we get a solution to Fermat's Last Theorem with exponent m. We can rule out $m = 4$ by Corollary 3.2. If $m = p$ is prime and we are given a solution to $a^p + b^p = c^p$ then $a^p + b^p + (-c)^p = 0$ as desired.

22.2. FLT and Sophie Germain. The first strong result on FLT was due to Sophie Germain:

Lemma 22.2. *Suppose that p is an odd prime for which $q = 2p + 1$ is also prime. If a, b, c are coprime integers for which $a^p + b^p + c^p \equiv 0 \pmod q$ then q divides at least one of a, b, c.*

Proof. Since $p = \frac{q-1}{2}$ we know that $t^p \equiv -1$ or $1 \pmod q$ for any integer t that is not divisible by q, and so if q does not divide abc then $a^p + b^p + c^p \equiv -3, -1, 1$ or $3 \pmod q$. This is impossible as $q = 2p + 1 > 3$.

Sophie Germain's Theorem. *Suppose that p is an odd prime for which $q = 2p + 1$ is also prime. There do not exist integers x, y, z for which p does not divide x, y, z and $x^p + y^p + z^p = 0$.*

Proof. Assume that there is a solution. As we saw in Proposition 22.1 we may assume that x, y, z are pairwise coprime so, by Lemma 22.2, exactly one of x, y, z is divisible by q: Let us suppose that q divides x, without loss of generality since we may re-arrange x, y and z as we please.

Now $(-z)^p = x^p + y^p = (x+y)(\frac{x^p+y^p}{x+y})$ and $(\frac{x^p+y^p}{x+y}, x+y) = (py^{p-1}, x+y) = (p, x+y) = 1$ so that $x + y$ and $\frac{x^p+y^p}{x+y}$ are both pth powers. Proceeding like this with $y + z$ and with $z + x$ we deduce that there exist integers a, b, c, d for which

$$x + y = a^p,\ z + x = b^p,\ y + z = c^p \text{ and } \frac{y^p + z^p}{y + z} = d^p \ \ (\text{as } p \not| xyz), \text{ where } x = -cd.$$

Now $a^p = z + y \equiv (z + x) + (x + y) \equiv b^p + c^p \pmod q$ as $q|x$, and so q divides at least one of a, b, c by Lemma 22.2. However since $(q, b)|(x, z + x) = (x, z) = 1$ as $q|x$ and $b|z + x$

and so q does not divide b, as well as a, analogously. Hence q divides c, that is $-z \equiv y \pmod q$. But then

$$d^p = \frac{y^p + z^p}{y+z} = \sum_{j=0}^{p-1} (-z)^{p-1-j} y^j \equiv \sum_{j=0}^{p-1} y^{p-1-j} y^j = p y^{p-1} \pmod q.$$

Therefore, as $y \equiv x + y = c^p \pmod q$ and $q - 1 = 2p$, we deduce that

$$4 \equiv 4d^{2p} = (2d^p)^2 \equiv (2py^{p-1})^2 = (-1)^2 (c^{2p})^{p-1} \equiv 1 \pmod q,$$

which is impossible as $q > 3$.

Hence if one can show that there are infinitely many pairs of primes $p, q = 2p+1$ then there are infinitely many primes p for which there do not exist integers x, y, z for which p does not divide x, y, z and $x^p + y^p + z^p = 0$.

After Sophie Germain's Theorem, the study of Fermat's Last Theorem was split into two cases:

I) Where $p \not| xyz$; and II) where $p | xyz$.

One can easily develop Germain's idea to show that if $m \equiv 2$ or $4 \pmod 6$ then there exists a constant $N_m \neq 0$ such that if p and $q = mp + 1$ are primes for which $q \not| N_m$ then FLTI is true for exponent p. This was used by Adleman, Fouvry and Heath-Brown to show that FLTI is true for infinitely many prime exponents.

There were many early results on the first case of Fermat's Last Theorem, which showed that if there is a solution with $p \not| xyz$, then some extraordinary other things must happen. Here is a list of a few:

If there is a solution to FLTI then

i) We have $2^{p-1} \equiv 1 \pmod{p^2}$. This seems to happen rarely (in fact, only the two examples 1093 and 3511 are known). One also has $3^{p-1} \equiv 1 \pmod{p^2}, 5^{p-1} \equiv 1 \pmod{p^2}, \ldots, 113^{p-1} \equiv 1 \pmod{p^2}$. One can obtain as many criteria like this as one wishes after a finite amount of calculation.

ii) p divides the numerator of $B_{p-3}, B_{p-5}, \ldots, B_{p-r}$ for $r \leq (\log p)^{1/2-o(1)}$. And p divides the numerator of at least $\sqrt{p} - 2$ non-zero Bernoulli numbers $B_n,\ 2 \leq n \leq p-3$.

iii)

Let us try to prove Fermat's Last Theorem, ignoring many of the technical issues. Let ζ be a primitive p the root of unity. Then we can factor

$$x^p + y^p = (x+y)(x+\zeta y)(x+\zeta^2 y) \ldots (x+\zeta^{p-1} y).$$

Now we are working in the set $\mathbb{Z}[\zeta]$, and we see that $\gcd(x+\zeta^i y, x+\zeta^j y)$ divides $(x+\zeta^i y) - (x+\zeta^j y) = (\zeta^i - \zeta^j)y$ and $\zeta^j(x+\zeta^i y) - \zeta^i(x+\zeta^j y) = (\zeta^j - \zeta^i)x$, so that $\gcd(x+\zeta^i y, x+\zeta^j y)$ divides $(\zeta^i - \zeta^j)(x, y)$. Note that $\zeta^k,\ 1 \leq k \leq p-1$ are the roots of $x^{p-1} + x^{p-2} + \ldots + 1$ and so

$$\prod_{k=1}^{p-1} (1 - \zeta^k) = \prod_{k=1}^{p-1} (x - \zeta^k)\big|_{x=1} = x^{p-1} + x^{p-2} + \ldots + 1\big|_{x=1} = p;$$

therefore if $k = j - i$ then $\zeta^i - \zeta^j = \zeta^i(1 - \zeta^k)$ divides p. So now assume that we have a solution to FLTI, that is $x^p + y^p = z^p$ with $\gcd(x, y) = 1$ and $p \not| z$, and so $x+y,\ x+\zeta y,\ x+\zeta^2 y,\ x+\zeta^{p-1}y$ are pairwise coprime elements of $\mathbb{Z}[\zeta]$ whose product is a pth power. If this works like the regular integers then each $x + \zeta^j y$ is a pth power. So we have gone from three linearly independent pth powers to p linearly dependent pth powers! In particular if $x + \zeta^j y = u_j^p$ then

$$(\zeta^j-\zeta^k)u_i^p+(\zeta^k-\zeta^i)u_j^p+(\zeta^i-\zeta^j)u_k^p = (\zeta^j-\zeta^k)(x+\zeta^i y)+(\zeta^k-\zeta^i)(x+\zeta^j y)+(\zeta^i-\zeta^j)(x+\zeta^k y) = 0.$$

Ignoring for a moment two technical details: the coefficients and the fact that we are no longer working over the integers, we see that we have found another solution to FLTI, this time with pth powers which are divisors of the previous pth powers and hence are smaller. Thus this seems to have the makings of a plan to prove FLTI by a descent process.

In 1850 Kummer attempted to prove Fermat's Last Theorem, much along the lines of last subsection. He however resolved a lot of the technical issues that we have avoided, creating the theory for ideals for $\mathbb{Z}[\zeta]$, much as we saw it discussed earlier for quadratic fields. That such similar theories evolved for quite different situations suggested that there was probably a theory of ideals that worked in any number field. Indeed such results were proved by Dedekind and became the basis of algebraic number theory, and indeed much of the study of algebra. Kummer's exact criteria was to show that if p does not divide a certain class number (associated to $\mathbb{Z}[\zeta]$) then Fermat's Last Theorem is true for exponent p. He showed that p does not divide that certain class number, if and only if p does not divide the numerators of $B_2, B_4, B_6, \ldots, B_{p-3}$.

In 1994 Andrew Wiles finally proved Fermat's Last Theorem based on an extraordinary plan of Frey and Serre to bring in ideas from the theory of elliptic curves. In fact Fermat's Last Theorem falls as a consequence of Wiles' (partial) resolution of the modularity conjecture. The elliptic curve associated to a solution $a^p + b^p + c^p = 0$ is $y^2 = x(x + a^p)(x - b^p)$, because then the discriminant Δ, which is the product of the difference of the roots, squared, equals $(abc)^{2p}$.

Wiles' proof of Fermat's Last Theorem, is extraordinarily deep involving modular forms, a subject far removed from the original, and other of the most profound themes in arithmetic geometry. If the whole proof were written in the leisurely style of, say, this book, it would probably take a couple of thousand pages. This could not be the proof that Fermat believed that he had – could Fermat have been correct? Could there be a short, elementary, marvelous proof still waiting to be found? Such a proof came to Lisbeth Salander in *The girl who played with fire* just as she went into the final tense moments of that novel — can truth follow fiction, as it so often does, or will this always remain a mystery?

Another famous problem about powers in Catalan's 1844 conjecture that the only perfect powers that differ by 1 are 8 and 9 (that is if $x^p - y^q = 1$ then either $x = 0$, or $y = 0$, or $x = q = 3$ and $y = p = 2$. After Baker's Theorem it was known that there could be only finitely many such pairs, but the conjecture was only proved in 2004 by Mihailescu with a proof more along the lines of Kummer's work on Fermat's Last Theorem.

Now that the two key conjectures in this field have been resolved, one can ask about other Diophantine equations involving powers. A hybrid is the *Fermat-Catalan equation*

$$x^p + y^q = z^r.$$

However there can be many uninteresting solutions: For example if $q = p$ and $r = p + 1$ one has solutions $(a(a^p + b^p))^p + (b(a^p + b^p))^p = (a^p + b^p)^{p+1}$ for any integers a and b. This kind of solution can be ignored by assuming that x, y and z are pairwise coprime. However it is not hard to find some solutions:

$$1 + 2^3 = 3^2, \quad 2^5 + 7^2 = 3^4, \quad 7^3 + 13^2 = 2^9, \quad 2^7 + 17^3 = 71^2, \quad 3^5 + 11^4 = 122^2.$$

and with a little more searching one finds five surprisingly large solutions:

$$17^7 + 76271^3 = 21063928^2, \quad 1414^3 + 2213459^2 = 65^7, \quad 9262^3 + 15312283^2 = 113^7,$$

$$43^8 + 96222^3 = 30042907^2, \quad 33^8 + 1549034^2 = 15613^3.$$

The Fermat-Catalan conjecture. *There are only finitely many solutions to $x^p+y^q = z^r$ in coprime integers x, y, z, where $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.*

A stronger version of the conjecture states that there are only the ten solutions listed above. However this sort of conjecture is always a little feeble since if someone happens to find one more isolated example, then would we not believe that those eleven solution are all?

Several people have observed that all of the solutions above have at least one exponent 2, so that one can conjecture that there are only finitely many solutions to $x^p + y^q = z^r$ in coprime integers x, y, z, where $p, q, r \geq 3$.

The cases where $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \geq 1$ are fully understood. When $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$ we only have the solution $1^6 + 2^3 = 3^2$.

`Exercise 22.1.1.` Show that there are infinitely many coprime solutions to $x^2 + y^2 = z^r$ for any fixed r. (Hint: Use your understanding of what integers are the sum of two squares.)

For the other cases there are infinitely many solutions; for example the parametrization

$$A((A - 8B))^3 + B(4(A + B))^3 = (A^2 + 20AB - 8B^2)^2$$

with $A = a^3, B = b^3$ (due to Euler, 1756)

Proof of Fermat's Last Theorem for polynomials. Suppose that

$$x^n + y^n = z^n,$$

where x, y, z are polynomials without common zeros. If d is the largest of their degrees then the *abc*-theorem for polynomials (see chapter 11.1) implies that $nd < 3d$, as this is an upper bound on the number of zeros of xyz. Hence $n < 3$.

That Fermat's Last Theorem is easy to prove for polynomials is an old result, going back certainly as far as Liouville (1851). We can extend the this analysis:

Proof of the Fermat-Catalan conjecture for polynomials. Let x, y, z be polynomials without common zeros, for which

$$x^p + y^q = z^r.$$

Let D be the largest of the degrees of x^p, y^q, z^r, so that the degree of xyz is $\leq \frac{D}{p} + \frac{D}{q} + \frac{D}{r}$. Hence the *abc*-theorem for polynomials implies that $D < \frac{D}{p} + \frac{D}{q} + \frac{D}{r}$. That is $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$.

If $p, q, r > 1$ and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 1$ then either $p = q = 2 \leq r$, or we have $(p, q, r) = (2, 3, 3), (2, 3, 4)$ or $(2, 3, 5)$. We gave parametric solutions for the first two cases, just above. We also note that

$$AB(16(A+B)(8B-A))^3 + (2A^2 + 40AB - 16B^2)^4 = (4(A^2 + 8B^2)(A^2 - 88AB - 8B^2))^2$$

with $A = a^3, B = b^3$. Much more interesting is the $(2, 3, 5)$ case that was addressed by Klein.

22.2. The *abc*-conjecture. Could there be a result for the integers that is analogous to the *abc*-theorem for polynomials, which would also imply Fermat's Last Theorem, and perhaps much more? The idea would be to bound the size of the integers involved (in place of the degree) in terms of their distinct prime factors (in place of the number of roots). A first guess at an analogous result might be if $a + b = c$ with a, b, c pairwise coprime positive integers then a, b and c are bounded in terms of the number of prime factors of a, b, c but if, as we believe, there are infinitely many pairs of twin primes $p, p + 2$ then we have just three prime factors involved in $p + 2 = q$, and they get arbitrarily large. It therefore seems sensible to include the size of the prime factors involved in such a bound so we might guess at $c \leq \prod_{p|abc} p$, but again a simple example excludes this possibility: Let $1 + (2^n - 1) = 2^n$. If we take any prime q and then $n = q(q-1)$ we have $q^2 | 2^n - 1$ and so $\prod_{p|(2^n-1)2^n} p \leq 2(2^n - 1)/q < 2^n$. In this case $q \approx \sqrt{n} \approx \sqrt{\log n}$ so even though our guess was wrong it is not too far out. This suggests that our guess is almost correct and could be made correct by fudging things a little bit:

The *abc*-conjecture. *For any fixed $\epsilon > 0$ there exists a constant κ_ϵ such that if a, b, c are pairwise coprime positive integers for which*

$$a + b = c$$

then

$$c \leq \kappa_\epsilon \left(\prod_{\substack{p \text{ prime} \\ p|abc}} p \right)^{1+\epsilon}.$$

In particular one might guess that $\kappa_1 = 1$; that is $c \le \left(\prod_{p|abc} p\right)^2$. We can apply the *abc*-conjecture to FLT: Let $a = x^p, b = y^p, c = z^p$ with $0 < x, y < z$ so that

$$\prod_{p|abc} p = \prod_{p|xyz} p \le xyz < z^3.$$

The *abc*-conjecture implies that $z^p \le \kappa_\epsilon (z^3)^{1+\epsilon}$.

`Exercise 22.2.1.` Deduce that if $p > 3$ then z is bounded independently of p.

`Exercise 22.2.2.` Suppose that p, q, r are positive integers for which $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.

(1) Show that $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} \le \frac{41}{42}$
(2) Show that the *abc*-conjecture implies that if $x^p + y^q = z^r$ with $(x, y, z) = 1$ then x, y, z are bounded independently of p, q, r.
(3) Deduce that if the *abc*-conjecture is true then the Fermat-Catalan conjecture is true.

22.3. Faltings' Theorem née Mordell's conjecture. Let $f(x, y) \in \mathbb{Z}[x, y]$ be an irreducible polynomial in two variables with integer coefficients. We are interested in finding rational numbers u and v for which $f(u, v) = 0$.

We have seen how to completely resolve this for f of degree 1 or 2: there are either no solutions, or an infinite of rational solutions , as a rational function of the variable t.

For f of degree 3 and sometimes 4 we can sometimes reduce the problem to an elliptic curve, and given one solution we can find another as a function of that solution, and thus get infinitely many solutions unless we hit on a torsion point (of which there are no more than 16).

Faltings' Theorem tells us that these are the only two ways in which an equation like $f(u, v) = 0$ have infinitely many rational solutions. That is, if put to one side all solutions of $f(x, y) = 0$ that come from the two methods above, then we are left with finitely many solutions. Therefore, for higher degree f, there are only finitely many "sporadic" solutions. It is even feasible that the number of rational points left over is bounded by a function of the degree of f. Faltings' extraordinary theorem has many wonderful consequences ... For any given $p \ge 4$ there are only finitely many positive coprime integer solutions to $x^p + y^p = z^p$. Similarly

$$x^4 + y^4 = 17z^4 \quad \text{and} \quad x^2 + y^3 = z^7$$

each have only finitely many coprime integer solutions. More generally there are only finitely many positive coprime integer solutions x, y, z to the Fermat-Catalan equation

$$ax^p + by^q = cz^r$$

for any positive coprime integers a, b, c whenever $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$.[45]

[45]Notice that this is not the full Fermat-Catalan conjecture, since here we have proved that there are only finitely many solutions for each *fixed* p, q, r (for which $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$), rather than there are only finitely many solutions, in total, over all possible p, q, r.

One important failing of Faltings' Theorem is that it does not give an upper bound on the size of the solutions, and so no "algorithm" for finding them all, even though we know there are only finitely many.

In 1991 Elkies showed that using an explicit version of the *abc*-conjecture (that is, with a value assigned to κ_ε for each ϵ), one can deduce an explicit version of Faltings' Theorem. The proof revolves around a careful study of the extreme cases in the *abc*-Theorem for polynomials.

Moret-Bailly, building on ideas of Szpiro, went a step further. He showed that if one could get *good* upper bounds for the size of the co-ordinates of the rational points on[46] $y^2 = x^5 - x$ in any number field[47] then the *abc*-conjecture follows. ("Good" bounds, in this case, are bounds that depend explicitly on the discriminant of the number field over which the points are rational). Therefore, in a certain sense, this problem and the *abc*-conjecture are equivalent.

[46] Or, for the initiated, on any other smooth algebraic curve of genus > 1.

[47] That is, a finite field extension of $\mathbb{Q}$.

Mahler measure, especially Jensen's formula?

References

[1] J.W.S. Cassels, *Rational quadratic forms*, Academic Press, 1968.
[2] J.W.S. Cassels, *Lecture on Elliptic curves*, London Mathematical Society, 1991.
[3] H. Davenport, *Multiplicative number theory*, Springer Verlag, New York, 1980.
[4] G.H. Hardy and E.M. Wright, *An introduction to the theory of numbers*, Oxford, 1938.
[5] Anthony W. Knapp, *Elliptic curves*, Princeton U, 1992.
[6] Neal Koblitz, *Introduction to elliptic curves and modular forms*, Springer Verlag, New York, 1993.
[7] L.J. Mordell, *Diophantine equations*, Academic Press, 1969.
[8] Jean-Pierre Serre, *A course in arithmetic*, Springer Verlag, New York, 1973.
[9] Joseph H. Silverman, *The arithmetic of elliptic curves*, Springer Verlag, New York, 1986.
[10] Joseph H. Silverman and John Tate, *Rational points on elliptic curves*, Springer Verlag, New York, 1992.
[St1] Harold M. Stark, *Automorphic functions of one variable I*, Springer Lecture Notes.
[11] John Tate, *Lectures on "Rational points on elliptic curves"*, Haverford lectures, 1961.
[12] Alf van der Poorten, *Notes on Fermat's Last Theorem*, Canadian Mathematical Society, 1996.