MATHEMATICS 6110, FALL 2013
INTRODUCTORY NUMBER THEORY

KATHERINE E. STANGE

Contents

Date: Last revised: December 18, 2013.

1. Preface

This document will be updated throughout the semester.

2. Problems in Number Theory

Number theory may be loosely defined as the study of the integers: in particular, the interaction between their additive and multiplicative structures. However, modern number theory is often described as the study of such objects as algebraic number fields and elliptic curves, which we have invented in order to answer elementary questions about the integers. Therefore, an argument can be made that the best way to define number theory is to exhibit some of these motivational problems.

2.1. **Are there infinitely many primes?** Yes, and you are invited to invent your own proofs of this fact (there are many). More generally, we study how many primes there are up to x (call this number $\pi(x)$).

The Prime Number Theorem (Hadamard and De La Vallée Poussin, 1896) famously states that $\pi(x) \sim x/\log x$, or actually the slightly better approximation $\pi(x) \sim \mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}$. The $\sim$ notation indicates that the ratio of the two functions tends to 1 in the limit. This growth rate, as a conjecture, goes back to Dirichlet and Gauss around 1800.

This doesn't answer the question completely, however: it is a never-ending problem to analyse the error term in closer and closer detail. The famous unproven Riemann Hypothesis, in one form, states that

$$\pi(x) = \mathrm{Li}(x) + O(x^{1/2} \log x).$$

This is "big O" notation, and it means that $\pi(x) - \mathrm{Li}(x)$ is eventually bounded above by a constant multiple of $x^{1/2} \log x$.

You've probably heard of the Riemann Hypothesis in another form, concerning the location of zeroes of the zeta function on the complex plane. (In fact, proofs of the prime number theorem all depended on complex analysis until a proof of Selberg and Erdös in 1949!) More on the mathematical details later, but it is worth mentioning that this is considered one of the premier unsolved problems in modern mathematics, and that most mathematicians both firmly believe the hypothesis and yet don't believe it will be proven in our lifetimes. It has so many powerful consequences in number theory, that the result must lie very deep. There are a great many research papers which prove results

conditional on various forms of the Riemann Hypothesis; hundreds of results will suddenly be unconditionally true when a proof is eventually found.

2.2. **Is there a closed formula for the n-th prime?** Believe it or not, there are some contenders, but they are not simple. Willans gives the formula

$$p_n = 1 + \sum_{m=1}^{2^n} \left\lfloor \sqrt[n]{n} \left(\sum_{x=1}^{m} \left\lfloor \cos^2 \pi \frac{(x-1)! + 1}{x} \right\rfloor \right)^{-1/n} \right\rfloor,$$

which is certainly a closed formula in some sense. It is a sort of obfuscation of the relationship between p_n and $\pi(x)$, using Wilson's theorem.

Theorem 2.1 (Wilson's Theorem)**.** *p is prime or 1 if and only if $(p-1)! \equiv -1 \pmod{p}$.*

It is, however, not particularly useful for computation, so in some sense it is not a satisfactory answer.

2.3. **Is there a (possibly multivariate) polynomial that gives only primes when evaluated on all integer inputs?** No. However, there are multivariate polynomials whose *positive* values are exactly all the primes, as the variables range over *natural* numbers. Such a polynomial in 26 variables, due to Jones, Sato, Wada and Wiens, is

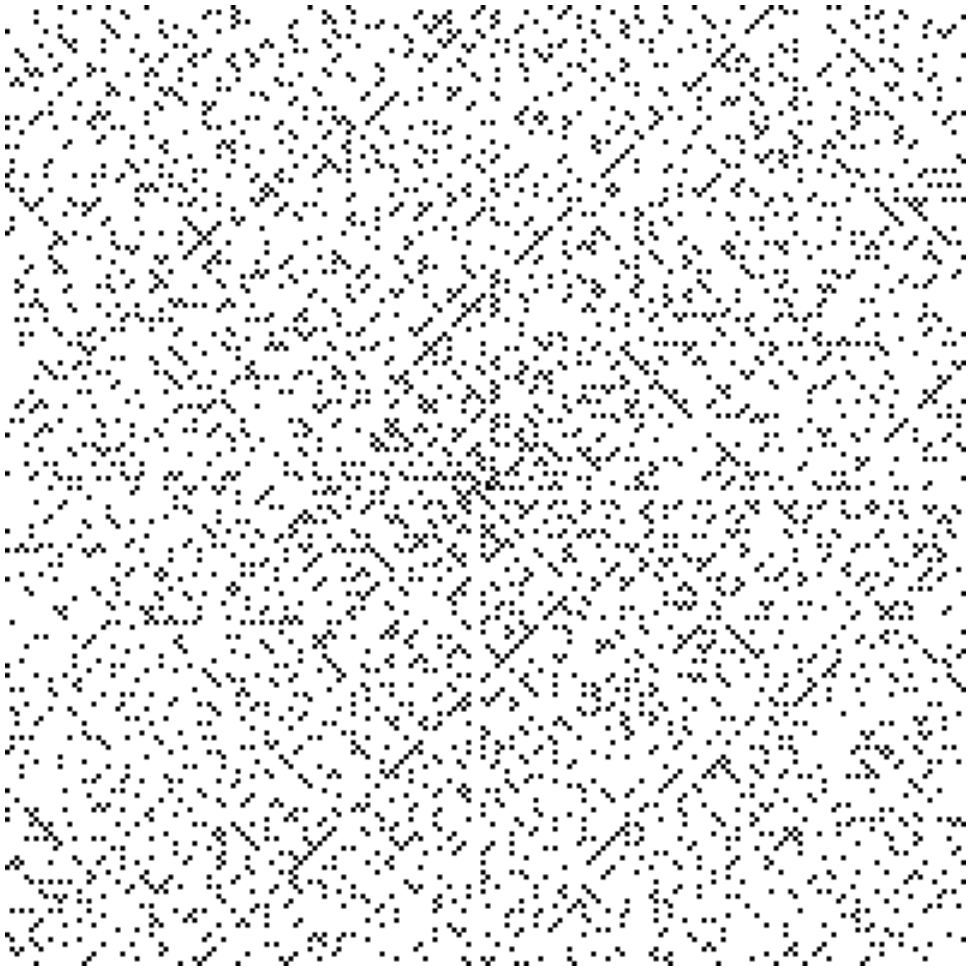
$$\begin{aligned}
&(k+2)(1-[wz+h+j-q]^2-[(gk+2g+k+1)(h+j)+h-z]^2- \\
&\quad [16(k+1)^3(k+2)(n+1)^2+1-f^2]^2-[2n+p+q+z-e]^2- \\
&\quad\quad [e^3(e+2)(a+1)^2+1-o^2]^2-[(a^2-1)y^2+1-x^2]^2- \\
&\quad\quad\quad [16r^2y^4(a^2-1)+1-u^2]^2-[n+l+v-y]^2- \\
&\quad\quad\quad [(a^2-1)l^2+1-m^2]^2-[ai+k+1-l-i]^2- \\
&\quad\quad [((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2]^2- \\
&\quad\quad [p+l(a-n-1)+b(2an+2a-n^2-2n-2)-m]^2- \\
&\quad\quad [q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2- \\
&\quad\quad\quad\quad [z+pl(a-p)+t(2ap-p^2-1)-pm]^2)
\end{aligned}$$

If you prefer fewer variables, you can get down to 10 variables if you let the degree go up to 15905. The proof is based on the logical notion of a Diophantine set.

2.4. **Are there infinitely many primes of the form $4n + 1$?** Yes. More generally, there are infinitely many primes of the form $an + b$ for any coprime a and b. This is Dirichlet's celebrated theorem on arithmetic progressions (1837). We won't cover the proof; it's more typically done in your analytic number theory course. However, there is a sense in which this question lies very much in the realm of algebraic number theory, and we will touch on related topics. In general, we expect about half of all primes to be congruent to 1 modulo 4 and the other half to be congruent to 3 mod 4. However, there are more in the former category, in the sense that, counting up to N, the former category is usually larger. See 'Prime Number Races' by Granville and Martin (American Mathematical Monthly).

2.5. **Are there infinitely many primes of the form $n^2 + 1$?** The more general question is a variation on the polynomial that produces only primes – but we now require only that it produce infinitely many primes. It is unknown for any quadratic polynomial. Iwaniec has shown that there are infinitely many n for which $n^2 + 1$ is the product of at most two primes.

Ulam noticed that if you draw the primes in a spiral around the origin on a square grid, it looks far from random. The integers which are values of quadratic polynomials eventually head out along diagonal lines. The most visible diagonal on his spiral is $n^2 + n + 41$, which is prime for $0 \leq n < 40$ (but not for $n = 40$).



2.6. **Are there infinitely many primes p for which $p+2$ is prime?** This is the famous twin primes conjecture (in the affirmative). It is still unsolved. More generally, one can ask how often $f_1(x)$ and $f_2(x)$ are simultaneously prime for some polynomials f_1 and f_2. Of course, there are infinitely many pairs of integers a, d such that a and $a+d$ are prime; we just don't know if we can take $d = 2$ infinitely often. In fact, van der Corput showed there are infinitely many 3-term arithmetic progressions in 1929. In 2004, Green and Tao showed there are infinitely many length k arithmetic progressions for all k.

Why do we think the answer is yes? This is an example of a pervasive heuristic argument in number theory. Using the Prime Number Theorem, we can guess that the 'probability' of a number x between 1 and N being prime is about $1/\log N$. Therefore, we expect the chance

that both x and $x+2$ are prime is about $1/\log^2 N$: there will be about $N/\log^2 N$ twin prime pairs below N.

But wait, this also predicts that there are infinitely many primes p such that $p+1$ is prime! Refine the model: odd numbers between 1 and N have a $2/\log N$ chance of being prime, while even ones have a 0 chance. Refine it for multiples of 3, of 5, etc. and eventually we obtain a count of twin primes that is

$$2 \prod_{p \text{ odd prime}} \left(1 - \frac{1}{(p-1)^2}\right) \frac{N}{\log^2 N}.$$

Not seeing any obvious reasons this is wrong, we conjecture this as the growth rate of twin primes. This relies on the oft-used heuristic that, having identified the 'obvious' ways in which primes are not random (congruence conditions, like most even numbers are not prime), they are otherwise *entirely random*!

Chen has shown in the 70's that there are infinitely many primes p such that $p+2$ is a product of at most two primes. This uses 'sieve methods'.

2.7. **Are there any quadratic forms with integer coefficients which represent all positive integers?** The answer is no, for binary and ternary forms. You will see in an introductory algebraic number theory course a classification of which integers are the sum of two squares; this fundamental result goes back to Fermat in 1640, but an elementary proof is not very easy.

Lagrange showed in 1770 that every positive integer is the sum of four squares. For quaternary and higher, it has been proven by Bhargava and Hanke (the '290 theorem' in 2005), that to determine if a form is 'universal' in this manner, it suffices to determine if it represents $1, 2, \ldots, 290$. (This came after the '15 theorem' of Conway and Schneeberger which applies to so called 'matrix-integral' forms; i.e. forms whose non-diagonal coefficients are even.)

2.8. **Is there an algorithm to determine if a given polynomial equation in any number of variables has an integer solution?** This is Hilbert's 10th Problem. Actually, he asked the audience to devise such a process, as it came as quite a surprise that the answer would be NO. This is a celebrated result of Davis, Matiyasevich, Putnam and Robinson. The existence of a polynomial whose positive values on natural numbers are all the primes is a corollary. The proof lies in the

realm of logic, and uses facts about the Fibonacci numbers in an essential way. For a wonderful read, see the book "Hilbert's Tenth Problem," by Matiyasevich.

The same question for rationals, in place of integers, is an open problem.

2.9. **Does there exist a deterministic polynomial-time algorithm (in the number of digits) to determine if n is prime?** A first method would be to check all divisors up to $\sqrt{n}$; this takes $O(\sqrt{n})$ time. At first glance, Fermat's Little Theorem seems a promising criterion.

Theorem 2.2 (Fermat's Little Theorem)**.** *For any prime p and a coprime to p,*

$$a^{p-1} \equiv 1 \pmod{p}.$$

However, many composite n also satisfy this equation for some a. If a composite n satisfies this equation for all coprime a, then it is called a *Carmichael number*, about which there are many questions (there are infinitely many such numbers).

In 1975, a deterministic polynomial time-algorithm was given by Miller, but this is only assuming the *Extended Riemann Hypothesis*[1]. Around the same time some randomized polynomial-time algorithms were discovered (meaning it can return NO when it should return YES, but with random probability $< 1/2$), and many more have appeared since. Finally, in 2002, Agrawal, Kayal and Saxena found the desired algorithm, running in $O(\log^{15/2} n)$ time.

2.10. **Can we factor numbers in deterministic polynomial time?** A good reference on this extensive subject is the book "Prime Numbers: A Computational Perspective," by Crandall and Pomerance. The quick answer is that there are sub-exponential algorithms known since the 70's, but no polynomial time algorithms, even under various generalised Riemann Hypotheses. However, there does not seem to be any evidence indicating that it is not possible, besides the fact that we have tried and failed, especially since the 70's. However, there are a great many very interesting algorithms, some of which we will meet in this

[1]A note on the Extended Riemann Hypothesis. The terminology on the various extensions of the Riemann Hypothesis is confusing; see the book "The Riemann hypothesis: a resource for the afficionado and virtuoso alike," by Peter Borwein, Stephen Choi, Brendan Rooney and Andrea Weirathmueller. The version used here is the usual critical-strip statement, applied to some particular Dirichlet L-functions (but not all).

class. With current methods, we can factor integers up to about 232 decimal digits (it took two years / 2000 computing years in 2009).

What complexity class is it? $\mathcal{P}$ refers to problems for which there are polynomial time algorithms. $\mathcal{NP}$ refers to problems for which a correct answer can be verified in polynomial time. Because of the AKS primality testing algorithm of 2005 (see above), factoring is in $\mathcal{NP}$. Famously, we do not know if $\mathcal{P} = \mathcal{NP}$.

2.11. **Up to N, are there always more natural numbers with an odd number of prime factors than with an even number of prime factors?** This is known as the Pólya Conjecture, and it seems heuristically reasonable that 'most' integers have an odd number of prime factors. It has important consequences in number theory and was widely believed between 1919 (when the conjecture was made) and 1958, when Haselgrove showed that it is false for infinitely many N. It is true until $N = 906,150,257$, when it fails. Never trust numerical evidence.

2.12. **Does $x^2 - 1141y^2 = 1$ have infinitely many solutions? (Note: if you ask the computer to check up to 25 digits, it will find none.)** As another example of misleading numerical evidence, the first solution to $x^2 - 1141y^2 = 1$ has y of 26 digits; there are infinitely many solutions. This is an example of a Pell equation (another topic we could see in this course). For more examples of 'The Strong Law of Small Numbers' (don't trust them), see Richard Guy's article by the same name.

2.13. **For any irrational number α, are there infinitely many rational numbers p/q such that $|\alpha - p/q| < 1/q^2$?** True for all irrational α; this is an application of the pigeonhole principle due to Dirichlet. It is Fields Medal work that for any algebraic α, there are only finitely many p/q such that $|\alpha - p/q| < 1/q^{2+\epsilon}$ (Roth's Theorem, 1955). This is the fundamental question of the area called *Diophantine approximation*.

2.14. **Given n, if it is even, divide by 2 and if it is odd, return $3n + 1$; if we iterate this rule, must we eventually reach the loop $1 \to 4 \to 2 \to 1$?** This is known as the Collatz Conjecture, and it is famous for driving mathematicians crazy in every mathematical discipline. It is an open question, and it is not clear which methods will resolve it.

2.15. **Given a real number $\alpha > 0$, if it is > 1, then subtract 1 and if it is < 1, then invert it; if we iterate this rule, must we eventually reach a loop?** The process above expresses any α as a *continued fraction*:

$$\alpha = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cdots}}}$$

Every real number has a continued fraction expansion. The a_i are eventually periodic (corresponding to a loop in the dynamical system of the question), if and only if α is rational or quadratic. A few famous continued fraction expansions are:

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{6 + \cdots}}}}}}}}$$

which has the pattern $2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots$, and

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cdots}}}}}}}}$$

which has no discernable pattern. Chopping off the fraction at any finite point, we obtain good rational approximations to α; in fact, these are the best rational approximations of Dirichlet's theorem above!

3. Some numerical experiments and Sage

We played with Sage in class; worksheets are on the website. Our goal was to collect some data on some of the motivating questions for this class, in a homework assignment.

- **Which integers are the sums of two squares?** This question (which can be viewed as an instance of a Diophantine equation) is among the simplest of questions relating the multiplicative structure of the integers to its additive structure. (A *Diophantine equation* is a polynomial equation for which we seek integer solutions. You could ask about the Diophantine equation $z = x^2 + y^2$ in three variables, or about the family of Diophantine equations $c = x^2 + y^2$ in two variables as c varies, for example.) It is also a key to all sorts of interesting number theory, particularly to algebraic number theory.

- **What is the distribution of the prime numbers?** This is the first and most natural question leading to analytic number theory. It leads into all sorts of other 'statistical' questions, like: How many divisors do integers have on average?
- **Which numbers are squares modulo a prime?** To understand $\mathbb{Z}$, it often makes sense to look 'at one prime at a time.' The ring $\mathbb{Z}/p\mathbb{Z}$ is not as simple as you think. One of the 'golden theorems' of number theory provides a sort of answer to this question.
- **How hard is it to find primes and factor numbers?** This question, in its most concrete sense, is hugely important for modern cryptographic methods. If someone finds a fast factoring algorithm tomorrow, they could access all sorts of secret data kept by individuals and governments.

We computed and made conjectures concerning some of these questions (homework).

4. Multiplicative arithmetic functions and Möbius inversion

Definition 4.1. *Define the following two functions $v, \sigma : \mathbb{Z}^+ \to \mathbb{Z}^+$, by*

$$v(n) = \text{ the number of positive divisors of } n,$$
$$\sigma(n) = \text{ the sum of the positive divisors of } n.$$

If one begins to experiment here, one may notice some patterns. You may be able to convince yourself that if n and m are coprime, then $v(nm) = v(n)v(m)$. After all, to form a divisor of nm, we can multiply a divisor of n by a divisor of m. (Try to make this into a rigorous argument.)

Definition 4.2. *A function $f : \mathbb{Z}^+ \to \mathbb{C}$ is* multiplicative *if $f(1) = 1$ and $f(nm) = f(n)f(m)$ whenever n and m are coprime.*

In other contexts, the term 'multiplicative' doesn't require coprimality in the definition (e.g. when we discuss the Gaussian norm later in the course). However, in the context of arithmetic functions, a function which satisfies $f(nm) = f(n)f(m)$ for all n and m is called *completely multiplicative* to differentiate it from the definition above.

A completely multiplicative function is multiplicative.

Examples of completely multiplicative functions include $f(n) = 1$, $f(n) = n$ and $f(n) = n^2$. Some multiplicative functions are not completely multiplicative, as we shall shortly see.

Theorem 4.3. *If $f(n)$ is multiplicative, so is $F(n) = \sum_{d|n} f(d)$.*

Proof. Suppose that m and n are coprime. Then

$$\begin{aligned} F(mn) &= \sum_{d|mn} f(d) \\ &= \sum_{r|m,s|n} f(rs) \\ &= \sum_{r|m,s|n} f(r)f(s) \\ &= \sum_{r|m} f(r) \sum_{s|n} f(s) \\ &= F(m)F(n). \end{aligned}$$

□

Note that we used coprimality in the step which introduces r and s above.

Corollary 4.4. *The functions $v(n)$ and $\sigma(n)$ are multiplicative.*

Proof. This follows from the multiplicativity of $f(n) = n$ and $f(n) = 1$, since

$$v(n) = \sum_{d|n} 1, \quad \sigma(n) = \sum_{d|n} d.$$

□

The functions $v(n)$ and $\sigma(n)$ are not completely multiplicative, however. For example,

$$v(2) = 2 \text{ but } v(4) = 3,$$

and

$$\sigma(2) = 3 \text{ but } \sigma(4) = 7.$$

As a corollary to the multiplicativity of v and σ, we obtain formulae in terms of the prime factorisation of n. Once we know a function is multiplicative, then to obtain such a formula only requires that we understand its values on prime powers.

Let $n = p^\alpha$. Then the positive divisors of n are $1, p, p^2, \ldots, p^\alpha$. Therefore

$$v(n) = \alpha + 1, \quad \sigma(n) = 1 + \cdots + p^\alpha = \frac{p^{\alpha+1} - 1}{p - 1}.$$

Now suppose that $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_\ell^{\alpha_\ell}$. Then

$$v(n) = \prod_{i=1}^{\ell} (\alpha_i + 1), \quad \sigma(n) = \prod_{i=1}^{\ell} \left(\frac{p_\ell^{\alpha_\ell + 1} - 1}{p_\ell - 1} \right).$$

In Theorem 4.3 above, we obtain $F(n)$ from $f(n)$. Can we go backward, i.e. can we obtain $f(n)$ from knowledge of $F(n)$? This will be answered below.

Definition 4.5. *The Dirichlet product of $f, g : \mathbb{Z}^+ \to \mathbb{C}$ is*

$$f \circ g : \mathbb{Z}^+ \to \mathbb{C}$$

given by

$$(f \circ g)(n) = \sum_{\substack{d_1 d_2 = n \\ d_i \in \mathbb{Z}^+}} f(d_1) g(d_2) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right).$$

I was careful to specify that the sum is over positive integers d_i, but I will not be careful henceforth; it is always the case in such sums.

Proposition 4.6. *The Dirichlet product is commutative and associative.*

Proof. Commutativity is clear from the definition. It is associative since

$$f \circ (g \circ h) = \sum_{d_1 d_2 d_3 = n} f(d_1) g(d_2) h(d_3) = (f \circ g) \circ h.$$

□

Definition 4.7. *Define functions $\mathbb{1}, I : \mathbb{Z}^+ \to \mathbb{C}$ by*

$$\mathbb{1}(n) = \begin{cases} 1 & n = 1 \\ 0 & n > 1 \end{cases},$$
$$I(n) = 1.$$

Proposition 4.8. *For any $f : \mathbb{Z}^+ \to \mathbb{C}$,*

(1) $\mathbb{1} \circ f = f \circ \mathbb{1} = f$,
(2) $I \circ f = f \circ I = \sum_{d|n} f(d)$.

At this point we have seen that the Dirichlet product is commutative and associative, and has an identity, $\mathbb{1}$. What about inverses? It turns out that inverses exist for many but not all functions $f : \mathbb{Z}^+ \to \mathbb{C}$. In the next proposition, we discover the inverse of I.

Definition 4.9. *The Möbius function $\mu : \mathbb{Z}^+ \to \{-1, 0, 1\}$ is defined by*

$$\mu(n) = \begin{cases} 1 & n = 1 \\ 0 & b^2 \mid n, \text{ for } b > 1 \\ (-1)^\ell & n = p_1 \cdots p_\ell, \text{ for } p_i \text{ prime.} \end{cases}$$

The Möbius function is easily verified to be multiplicative.

Proposition 4.10. *$I \circ \mu = \mu \circ I = \mathbb{1}$.*

Proof. We have $\mu \circ I(1) = \mu(1)I(1) = 1$. Now let $n > 1$ and write $n = p_1^{a_1} \cdots p_\ell^{a_\ell}$ for p_i prime and $\ell \geq 1$. Then,

$$\begin{aligned} \mu \circ I(n) &= \sum_{d|n} \mu(d) \\ &= \sum_{\substack{(e_1,\ldots,e_\ell) \\ e_i \in \{0,1\}}} \mu(p_1^{e_1} \cdots p_\ell^{e_\ell}). \end{aligned}$$

But there is a bijection between those tuples $(e_1, \ldots, e_\ell)$ which have $e_1 + \cdots + e_\ell \equiv 0 \bmod 2$ (equivalently, $\mu(\prod p_i^{e_i}) = 1$) and those which have $e_1 + \cdots + e_\ell \equiv 1 \bmod 2$ (equivalently, $\mu(\prod p_i^{e_i}) = -1$). The bijection is given by changing the last digit e_ℓ from 1 to 0 or vice versa. This bijection shows that the sum above vanishes. Hence $\mu \circ I(n) = 0$ for $n > 1$. □

Theorem 4.11 (Möbius Inversion Theorem)**.** *Let $F(n) = \sum_{d|n} f(d)$. Then*

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right).$$

In other words, if $F = f \circ I$, then $f = F \circ \mu$.

Proof. $F \circ \mu = (f \circ I) \circ \mu = f \circ (I \circ \mu) = f \circ \mathbb{1} = f$. □

Earlier, we have a theorem (Theorem 4.3) that whenever f is multiplicative, then $f \circ I$ is multiplicative. Actually, an almost identical proof shows that

Theorem 4.12. *If $f, g : \mathbb{Z}^+ \to \mathbb{C}$ are multiplicative functions, then $f \circ g$ is multiplicative.*

Proof. Exercise. □

Definition 4.13. *The* Euler phi function*, $\phi : \mathbb{Z}^+ \to \mathbb{C}$, is given by*

$\phi(n) =$ *the number of integers $x \in [1, n]$ which are relatively prime to n.*

Proposition 4.14.

$$\sum_{d|n} \phi(d) = n.$$

Proof. Consider the n rational numbers

$$\frac{1}{n} \quad \frac{2}{n} \quad \frac{3}{n} \quad \cdots \quad \frac{n-1}{n} \quad \frac{n}{n}$$

in lowest terms. Let $d \mid n$. Then k/d appears in this list as lowest terms for exactly those k which are relatively prime to d. That is, we

see denominator d exactly $\phi(d)$ times in the list. This holds for each $d \mid n$, and as d ranges through divisors of n, each fraction is accounted for once. Therefore,

$$n = \sum_{d|n} \phi(d).$$

□

Proposition 4.15. *$\phi(n)$ is multiplicative.*

Proof. From the last proposition, $\phi \circ I = g$, where $g : \mathbb{Z}^+ \to \mathbb{C}$ is the function $g(n) = n$. By Möbius inversion, $\phi = g \circ \mu$. Since g and μ are multiplicative, ϕ must be multiplicative. □

The proof actually shows that whenever $f \circ I$ is multiplicative, f must be.

Now we may obtain a formula for the Euler ϕ function. For a prime p, $\phi(p) = p - 1$. For a prime power, the only integers below p^α which are not coprime to p^α are the powers of p. Hence $\phi(p^\alpha) = \frac{p-1}{p} p^\alpha = p^{\alpha-1}(p-1)$. Then, by multiplicativity, for $n = p_1^{a_1} \cdots p_\ell^{a_\ell}$,

$$\phi(n) = n \prod_{i=1}^{\ell} \left(1 - \frac{1}{p_i}\right).$$

5. The zeta function

Consider the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

For $s > 1$, this series converges (by the p-test). However, in the limit $s \to 1^+$, it diverges.

The series is called the *Riemann zeta function*. By unique factorisation, each n is expressed uniquely as a product of primes, and as n ranges through the integers, all possible products appear. Therefore,

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right).$$

This is the analytic version of the statement of unique prime factorisation in the integers! Summing the convergent geometric series on the right gives the *Euler product formula*:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

If there were only finitely many primes, then the product on the right would be a finite product, which would imply the zeta function converges for all s. Since this is not true, this serves as a proof that there are infinitely many primes.

Series like the zeta function, i.e. of the form $\sum_{i=1}^{\infty} \frac{a_n}{n^s}$ for some sequence a_n, are called *Dirichlet series*.

Dirichlet series serve as *generating functions* for arithmetic functions. Generating functions are used in combinatorics to obtain identities between counting functions of various sorts (there's an example on your homework; see also the introduction to Newman's book Analytic Number Theory). Similarly, functions like the Riemann zeta function can be used to obtain identities between arithmetic functions. The Euler product identity is the analytic form of the statement of unique factorisation. In analytic number theory, one extends the definition of $\zeta(s)$ to complex s and studies its properties in order to discover facts about primes (most noteably, the Prime Number Theorem describing the growth of primes).

We will have occasion to use the following fact in the near future:

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}. \tag{1}$$

We'll see momentarily that this is a special case of the Möbius inversion formula interpreted as a statement about Dirichlet series. In fact, the Dirichlet product structure on arithmetic functions is simpler to understand in the language of Dirichlet series.

Theorem 5.1. *Suppose that $g = f_1 \circ f_2$. Let $G(s), F_1(s)$ and $F_2(s)$ be the Dirichlet series associated to g, f_1 and f_2 respectively, i.e.*

$$G(s) = \sum_{n=1}^{\infty} \frac{g(n)}{n^s}, \quad F_1(s) = \sum_{n=1}^{\infty} \frac{f_1(n)}{n^s}, \quad F_2(s) = \sum_{n=1}^{\infty} \frac{f_2(n)}{n^s}.$$

Then, $G(s) = F_1(s)F_2(s)$.

Proof.

$$\begin{aligned} F_1(s)F_2(s) &= \sum_{m=1}^{\infty} \frac{f_1(m)}{m^s} \sum_{k=1}^{\infty} \frac{f_2(k)}{k^s} \\ &= \sum_{n=1}^{\infty} \frac{1}{n^s} \sum_{m|n} f_1(m) f_2\left(\frac{n}{m}\right) \\ &= \sum_{n=1}^{\infty} \frac{g(n)}{n^s} \\ &= G(s). \end{aligned}$$

□

From this, one can see that multiplication is associative and commutative, and that the function $\mathbb{1}$ is the identity. One can also see that inverses are more complicated: one needs the quotient of two Dirichlet series to come in the form of a Dirichlet series.

In particular, one obtains (1) by taking $g = \mathbb{1}$, and $f_1 = I$. Then $G(s) = 1$ and $F_1(s) = \zeta(s)$. Then, $F_2(s) = 1/\zeta(s)$, while $f_2 = \mu$.

6. The Basel Problem

What is

$$\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2}?$$

Converges slowly; before computers, only a couple of decimal places could be computed. This was part of the inspiration for Riemann to define $\zeta(s)$.

Euler gave an argument (1735) whose validity needed to wait until a century later when analysis was rigorized (is that a word?). Here's the "argument" which is not a proof.

It depends on the expansion of $\sin(x)$ a la Calculus:

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

But Euler also expanded it as a product, as if it were a polynomial:

$$\sin(x) = \prod_{n=-\infty}^{\infty} \left(1 - \frac{x}{\pi n}\right) = x\left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\cdots$$

He thought this was reasonable, because the zeroes of $\sin(\pi x)$ are all the integers; so each side vanishes at the right places. But it's pretty suspicious, really: how do we know an infinite product converges? How do we know the constant is right? But let's run with it.

Expanding the product then, into a series in powers of x, we obtain

$$\begin{aligned}\sin(x) = x(1) & \\ &+ x^3\left(-\frac{1}{\pi^2} - \frac{1}{4\pi^2} - \frac{1}{9\pi^2} - \cdots\right) \\ &+ \pi x^5\left(\frac{1}{\pi^2\cdot 4\pi^2} + \frac{1}{\pi^2\cdot 9\pi^2} + \frac{1}{4\pi^2\cdot 9\pi^2} + \cdots\right) \\ &+ \cdots\end{aligned}$$

Then we compare coefficients. The coefficient of x^3 tells us that

$$-\frac{1}{3!} = -\sum_{n=1}^{\infty}\frac{1}{n^2\pi^2}.$$

or, in other words

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

The justification needed to make this into a proof (that we can treat $\sin(x)$ as if it were a polynomial and factor it by its zeroes) is the Weierstrass Factorization Theorem.

A closed form for $\zeta(3)$ is not known; it was only proven to be irrational in 1979 by Apery.

7. Counting Squarefree Integers, using Möbius inversion

First, we need to show a different inversion formula, similar to the regular Möbius inversion. It is a consequence of regular Möbius inversion, i.e. the fact that $\mu \circ I = \mathbb{1}$.

Theorem 7.1. *Let f be defined for all positive real numbers. For positive real x, define*

$$F(x) = \sum_{n=1}^{\lfloor x\rfloor} f\left(\frac{x}{n}\right).$$

Then

$$f(x) = \sum_{n=1}^{\lfloor x\rfloor}\mu(n)F\left(\frac{x}{n}\right).$$

Proof.

$$\begin{aligned}
&\sum_{n=1}^{\lfloor x \rfloor} \mu(n) F\left(\frac{x}{n}\right) \\
&= \sum_{n=1}^{\lfloor x \rfloor} \mu(n) \sum_{m=1}^{\left\lfloor \frac{x}{n} \right\rfloor} f\left(\frac{x}{mn}\right) \\
&= \sum_{n=1}^{\lfloor x \rfloor} \sum_{m=1}^{\left\lfloor \frac{x}{n} \right\rfloor} \mu(n) f\left(\frac{x}{mn}\right) \\
&= \sum_{k=1}^{\lfloor x \rfloor} \sum_{n|k} \mu(n) f\left(\frac{x}{k}\right) \\
&= \sum_{k=1}^{\lfloor x \rfloor} f\left(\frac{x}{k}\right) \sum_{n|k} \mu(n) \\
&= \sum_{k=1}^{\lfloor x \rfloor} f\left(\frac{x}{k}\right) (\mu \circ I)(k) \\
&= \sum_{k=1}^{\lfloor x \rfloor} f\left(\frac{x}{k}\right) \mathbb{1}(k) \\
&= f(x)
\end{aligned}$$

□

This allows us to count the squarefree numbers. Define

$$Q(x) = \#\{1 \leq n \leq x : n \text{ is squarefree}\}.$$

Theorem 7.2.

$$Q(x) = \frac{6x}{\pi^2} + O(\sqrt{x}).$$

Proof. Let y be a positive integer. Define sets S_i, $i = 1, 2, \ldots$ by

$$S_i = \{n \leq y^2 : \text{ the largest square factor of } n \text{ is } i^2\}.$$

Then

$$\#S_i = Q\left(\frac{y^2}{i^2}\right).$$

In particular, if $i > y$, then S_i is empty. Since each positive integer $\leq y^2$ belongs to exactly one set, we have

$$y^2 = \sum_{i \leq y} Q\left(\frac{y^2}{i^2}\right).$$

Applying Theorem 7.1,

$$\begin{aligned} Q(y^2) &= \sum_{i \leq y} \mu(i) \left\lfloor \frac{y^2}{i^2} \right\rfloor \\ &= \sum_{i \leq y} \mu(i) \left(\frac{y^2}{i^2} + O(1) \right) \end{aligned}$$

Because $\sum_{i \leq y} \mu(i) \leq y$, we have

$$\begin{aligned} Q(y^2) &= y^2 \sum_{i \leq y} \frac{\mu(i)}{i^2} + O(y) \\ &= y^2 \sum_{i=1}^{\infty} \frac{\mu(i)}{i^2} + O\left(y^2 \sum_{i > y} \frac{1}{i^2} \right) + O(y) \end{aligned}$$

Because $\sum_{i>y} \frac{1}{i^2} \approx \int_y^\infty \frac{di}{i^2} = \frac{1}{y}$, we have

$$Q(y^2) = y^2 \sum_{i=1}^{\infty} \frac{\mu(i)}{i^2} + O(y)$$

Below, we will see that $\sum_{i=1}^{\infty} \frac{\mu(i)}{i^s} = \frac{1}{\zeta(s)}$. Also, $\zeta(2) = \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$. Hence

$$\begin{aligned} Q(y^2) &= \frac{y^2}{\zeta(2)} + O(y) \\ &= \frac{6y^2}{\pi^2} + O(y) \end{aligned}$$

We obtain the theorem by replacing y^2 by x.

□

8. The growth of $\pi(x)$

Counting the primes is one of the central tasks of analytic number theory. A good first result is the following, which gives a hint as to the importance of the zeta function.

Theorem 8.1. *The sum*

$$\sum_{p \ prime} \frac{1}{p}$$

diverges.

Proof. Let us 'approximate' $\zeta(1)$, which diverges, by taking only the primes of size at most N:

$$\zeta_N(1) = \prod_{p \leq N} \left(1 - \frac{1}{p}\right)^{-1}.$$

(All products over p are products over prime p in this proof.) Then $\zeta_N(1) \to \infty$ as $N \to \infty$. Let us compute

$$\begin{aligned}\log \zeta_N(1) &= -\sum_{p \leq N} \log\left(1 - \frac{1}{p}\right) \\ &= \sum_{p \leq N} \sum_{m=1}^{\infty} \frac{1}{mp^m} \\ &= \sum_{p \leq N} \frac{1}{p} + \sum_{p \leq N} \sum_{m=2}^{\infty} \frac{1}{mp^m}\end{aligned}$$

We can bound the second sum:

$$\sum_{m=2}^{\infty} \frac{1}{mp^m} < \sum_{m=2}^{\infty} \frac{1}{p^m} = \frac{1}{p^2}\left(\frac{1}{1 - \frac{1}{p}}\right) \leq \frac{2}{p^2}.$$

Therefore,

$$\log \zeta_N(1) \leq \sum_{p \leq N} \frac{1}{p} + 2\sum_{p \leq N} \frac{1}{p^2}.$$

The left hand side diverges as $N \to \infty$, but the right-most sum converges (since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ is finite, for example). This implies that $\sum_{p \leq N} \frac{1}{p}$ diverges as $N \to \infty$. □

This tells us that, for example, the primes are more numerous than the squares, since $\sum_{n=1}^{\infty} \frac{1}{n^2}$ converges.

9. The Sieve of Eratosthenes

Write out a list of primes. Cross off all even numbers save 2 ("sieve by 2"). Cross off all numbers divisible by 3 save 3 itself ("sieve by 3"). Cross off all numbers divisible by 5 save 5 itself ("sieve by 5"). At each stage, there is some first number in the list which is neither crossed off nor saved; this is the next prime. It should be saved, and its multiples crossed off. Should we continue this forever, we will have sieved by each prime, and a list of the primes and only the primes shall remain. This is the Sieve of Eratosthenes.

This gives us a chance to return to probabilistic heuristics. For example, sieving by 2 leaves about 1/2 the integers. Sieving by 3 leaves about 1/3 of those remaining, etc. So we expect after sieving by all primes up to $\sqrt{x}$ (which should suffice for eliminating non-primes below bound x), that we have

$$\pi(x) \approx x \prod_{p \leq \sqrt{x}} \left(1 - \frac{1}{p}\right).$$

(here I will use p to denote primes instead of specifying that the sum is over primes every time).

It turns out, however, that this is a good but not great estimate. The problem is that it assumes that, probabilistically speaking, whether a number is divisible by p_1 is independent of whether it is divisible by p_2, i.e. after we've removed multiples of p_1, the multiples of p_2 are still proportion $1/p_2$ of what remains. But this is not true once primes are big (relative to x)! After all, you can only fit so many big primes into the factorization of one number. It turns outwe get more like 8/9ths of the estimate above, in reality.

10. The Prime Number Theorem

Grandly named because it is regarded as probably the most important result in number theory.

Theorem 10.1. *We have*

$$\pi(x) \sim \frac{x}{\log x} \sim \mathrm{Li}(x)$$

These estimates are not as mysterious as they seem. Starting from Gauss' 1792 observation that the probability of a number of size approximately x of being prime seems to be $1/\log x$, one could roughly estimate that proportion $1/\log x$ of numbers below x are prime. Better yet, one could take into account that the proportion is changing as the numbers grow, to guess

$$\sum_{n=2}^{x} \frac{1}{\log n} \approx \int_2^x \frac{dt}{\log t},$$

which is called $\mathrm{Li}(x)$. For this reason, $\frac{x}{\log x}$ is actually an underestimate, and $\mathrm{Li}(x)$ is a much better guess. However, the ratio of both guesses is 1 in the limit.

The estimate was conjectured by Gauss in 1792 (Legendre was also making conjectures in this general direction).

If $\mathrm{Li}(x)$ really is a good guess, then we should concentrate on the error term of this guess. In 1848-50, Chebyshev acheived some partial results (he showed that if the limit $\lim_{x\to\infty} \frac{\pi(x)\log(x)}{x}$ exists, it must be 1). In 1859, Riemann studied $\zeta(s)$ in a very famous paper. Riemann extended $\zeta(s)$ moromorphically to the entire complex plane (it has exactly one pole, at $s = 1$) by analytic continuation. We find that $\zeta(s)$ has so-called 'trivial' zeroes at negative even integers.

But the full PNT was proven for the first time in 1896 independently by Hadamard and de la Vallée Poussin. Their method was essentially analytic. The key to the proof is to show that all other zeroes lie in the *critical strip* $0 < \Re(s) < 1$. The tricky part is showing that there are no zeroes having $\Re(s) = 1$.

The Hadamard-de la Vallé Poussin proof gives an error estimate, in fact:

Theorem 10.2. *There exist positive constants C and a such that*

$$|\pi(x) - \mathrm{Li}(x)| \leq Cxe^{-a\sqrt{\log x}}.$$

However, it is not a great error estimate.

Further progress on the prime number theorem consists of improving the error bound. To do so, one proves larger and larger zero-free regions in the critical strip. What do we hope the final outcome will be?

Well, if the primes are really "random," then we can use the following probabilistic argument to guess at the error term.

Suppose that one writes a sequence of zeroes and ones, the nth one indicating whether n is prime (1) or not (0). Model this by a random sequence subject to the single constraint that the probability of the n-th digit being a 1 is $1/\log n$. Assume each digit is independent (an infinite sequence of independent random variables, in the language of probability). Then, to guess what happens with the sequence from the primes, we ask instead, what is true of probability 1 for this collection of random sequences? (There are a few more details missing here.) This is called the *Gauss-Cramér model*. It predicts that

$$\pi(x) = \mathrm{Li}(x) + O(\sqrt{x}\log x).$$

And in fact, if one could prove the Riemann Hypothesis, that all nontrivial zeroes lie on the line $\Re(s) = 1/2$, then the error term would become

Conjecture 10.3. *For $x \geq 3$,*

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

In 1949, Selberg and Erdös provided an elementary (no analysis) but very intricate proof of the PNT.

Questions about the growth of primes are equivalent to questions about the growth of the sum of the Möbius function.

Theorem 10.4. *The following are equivalent:*

(1) The Riemann Hypothesis
(2) For $x \geq 3$,

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

(3) For all $\epsilon > 0$, there exists C_ϵ such that

$$\left|\sum_{n=1}^{N} \mu(n)\right| \leq C_\epsilon N^{1/2+\epsilon}$$

The PNT is equivalent to the following theorem.

Theorem 10.5.

$$\lim_{N\to\infty} \frac{\sum_{n=1}^{N} \mu(n)}{N} = 0$$

A naïve approach to the question of the size of $\sum_{n=1}^{N} \mu(n)$, called the *Mertens function*, is to assume the Möbius function is essentialy randomly ± 1 on the squarefree n. A random walk of N steps leaves you at an expected distance of $\sqrt{N}$ from where you started. So it seems intuitive, perhaps, but it is not easy to prove.

11. Primes and Zeroes of the zeta function

We'd like to see how the zeroes of the zeta function actually relate to the primes. This is just a bit of calculus.

The first step is to consider the Mangoldt function

$$\Lambda(n) = \begin{cases} \log p & n \text{ is a prime power} \\ 0 & \text{otherwise} \end{cases}$$

Why this function? For one thing, weighting the primes by $\log p$ makes their sum 'density 1' everywhere. But more importantly, knowing the growth of this will give the growth of the primes. It turns out that Gauss' estimate is equivalent to estimating

$$\sum_{n\leq x} \Lambda(n) \sim x.$$

Here's why:

$$\begin{aligned}(\log x)\pi(x) &= \sum_{p\le x} \log x \\ &= \sum_{p\le x} \log p \left(\frac{\log x}{\log p}\right) \\ &= \sum_{p\le x} \log p \left\lfloor \frac{\log x}{\log p} \right\rfloor + O\left(\sum_{p\le x} \log p\right) \\ &= \sum_{p\le x} \log p \sum_{m:p^m\le x} 1 + O\left(\sum_{p\le x} \log p\right) \\ &= \sum_{n\le x} \Lambda(n) + O\left(\sum_{p\le x} \log p\right).\end{aligned}$$

Of course, we really want to ignore the big-O term, which means we expect

$$\sum_{n\le x} \Lambda(n) \sim x.$$

In fact, if this is true, then

$$\lim_{x\to\infty} \frac{\sum_{p\le x} \log p}{x} \le 1.$$

So we discover that

$$(\log x)\pi(x) \sim x$$

which is the statement of the Prime Number Theorem. For this reason, we can estimate this Mangoldt sum instead of $\pi(x)$, and perhaps it's simpler (since the estimate is just x, not $\mathrm{Li}(x)$). The other reason is

that it relates nicely to the zeta function, using Euler's product formula:

$$\begin{aligned}
-\frac{\zeta'(s)}{\zeta(s)} &= -\frac{d}{ds}\log\zeta(s) \\
&= -\frac{d}{ds}\log\prod_{p \text{ prime}}\left(1-p^{-s}\right)^{-1} \\
&= -\frac{d}{ds}\sum_{p \text{ prime}} -\log\left(1-p^{-s}\right) \\
&= \sum_{p \text{ prime}} \frac{d}{ds}\log\left(1-p^{-s}\right) \\
&= \sum_{p \text{ prime}} \frac{p^{-s}\log p}{1-p^{-s}} \\
&= \sum_{p \text{ prime}} \left(p^{-s}\log p\right)\left(1+\frac{1}{p^s}+\frac{1}{p^{2s}}+\cdots\right) \\
&= \sum_{p \text{ prime}}\sum_{n=1}^{\infty}\frac{\log p}{p^{ns}} \\
&= \sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s}
\end{aligned}$$

(Here we follow Andrew Granville's article in the Princeton Companion to Mathematics.) So we actually want to take the denominator out and cut off this function by requiring $p^m = n \leq x$. We can use a step function to do this. Define

$$f(y) = \begin{cases} 0 & 0 < y < 1 \\ 1/2 & y = 1 \\ 1 & y > 1 \end{cases}$$

Perron's formula in analysis gives $f(y)$ as a path integral in the complex plane (along a vertical line):

$$f(y) = \frac{1}{2\pi i}\int_{s:Re(s)=c}\frac{y^s}{s}\,ds.$$

Using $y = x/p^m$ (assume x is not a prime power), and taking c large enough so that everything converges absolutely,

$$\begin{aligned}\sum_{n\leq x}\Lambda(n) &= \sum_{p,m\geq 1}(\log p)f\left(\frac{x}{p^m}\right)\\ &= \frac{1}{2\pi i}\sum_{p,m\geq 1}\log p\int_{s:Re(s)=c}\left(\frac{x}{p^m}\right)^s\frac{ds}{s}\\ &= \frac{1}{2\pi i}\int_{s:Re(s)=c}\left(\sum_{p,m\geq 1}\frac{\log p}{p^{ms}}\right)x^s\frac{ds}{s}\\ &= \frac{1}{2\pi i}\int_{s:Re(s)=c}\left(\sum_{n\geq 1}\frac{\Lambda(n)}{n^s}\right)x^s\frac{ds}{s}\\ &= -\frac{1}{2\pi i}\int_{s:Re(s)=c}\frac{\zeta'(s)}{\zeta(s)}\frac{x^s}{s}ds.\end{aligned}$$

Now this should look familiar if you've taken complex analysis. We won't assume complex analysis, except for a few basic facts. The main facts here are that for an analytic function f (except at finitely many points),

(1) The poles of $f'(z)/f(z)$ have order 1 and represent the zeroes and poles of f; each residue is the order of that zero or pole (pole = negative residue).
(2) Cauchy's residue theorem says that for an appropriate contour C

$$\frac{1}{2\pi i}\int_C f(z)dz = \sum_{z:\text{poles and zeroes inside } C}\text{Res}(f,z)$$

So, applying this to ζ (analytically continued), which has a unique pole at 1, and whose zeroes are all simple,

$$\sum_{n\geq 1}\Lambda(n) = x - \sum_{\zeta(\rho)=0}\frac{x^\rho}{\rho} - \frac{\zeta'(0)}{\zeta(0)}$$

The details are left for those who know some complex analysis. The point is that we have an explicit formula giving the error term in our estimate for the Mangoldt sum, which is governed by the zeroes of the zeta function. Knowing these zeroes, then, will let us pin down the error term and estimate the growth of $\pi(x)$. That story will have to wait for a class in analytic number theory.

12. Number Rings and Prime Factorization

In $\mathbb{Z}$, everything factors uniquely as a product of primes and one copy of 1 or -1 (the two *units*). This property is so hugely fundamental to everything that everyone assumed it held in other rings. Here are a few of the other "number rings" people wanted to study:

Definition 12.1. *The* Gaussian integers *are the ring*

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\} \subset \mathbb{C}.$$

Addition is as for the complex numbers, i.e.

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$
$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

This confirms that the result of an addition or multiplication of two Gaussian integers is again a Gaussian integer. The elements 0 and 1 are also Gaussian integers, so that $\mathbb{Z}[i]$ is a subring of $\mathbb{C}$. Furthermore, since $\mathbb{C}$ has no zero divisors, $\mathbb{Z}[i]$ is an integral domain.

We can visualise the Gaussian integers as the lattice of points in the complex plane which have integer coordinates. Multiplication of two complex numbers adds their angles and multiplies their lengths.

The Gaussian integers are equipped with a particularly useful function.

Definition 12.2. *The norm on the Gaussian integers is the function* $N : \mathbb{Z}[i] \to \mathbb{Z}^+$ *given by* $N(a + bi) = a^2 + b^2$ *(i.e. the square of the length of the vector from the origin to* $a + bi$*).*

The norm satisfies the following useful properties:

(1) $N(a + bi) = 0 \iff a + bi = 0$
(2) N is multiplicative, i.e. for $\alpha, \beta \in \mathbb{Z}[i]$,
$$N(\alpha\beta) = N(\alpha)N(\beta).$$
(3) $N(a + bi) = (a + bi)(a - bi)$, or in other words, $N(\alpha) = \alpha\overline{\alpha}$, i.e. the norm is the product of an element with its complex conjugate.

The Gaussian integers of shortest length are those of length one: $1, -1, i, -i$; all others have longer length. Since lengths multiply, this shows that whenever $uv = 1$, it must be that u and v are chosen from the list $\{1, -1, i, -i\}$.

Definition 12.3. *The* Eisenstein Integers *are the ring* $\mathbb{Z}[\omega]$*, where* $\omega = \frac{-1+\sqrt{-3}}{2}$*, a primitive cube root of unity. In other words,*

$$\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\} \subset \mathbb{C}.$$

The primitive cube root of unity ω has minimal polynomial x^2+x+1.

To verify that this is a ring, let's see that addition and multiplication of Eisenstein integers give Eisenstein integers:

$$(a+b\omega)+(c+d\omega)=(a+c)+(b+d)\omega.$$

Multiplication is a bit longer to compute:

$$\begin{aligned}(a+b\omega)(c+d\omega) &= ac+ad\omega+bc\omega+bd\omega^2\\ &= ac+ad\omega+bc\omega+bd(-\omega-1)\\ &= (ac-bd)+(ad+bc-bd)\omega\end{aligned}$$

One more ring that's of interest to us is $\mathbb{Z}\left[\frac{1+\sqrt{-23}}{2}\right]$. It was a long time before people understood the importance of this, but in fact,

$$\frac{1+\sqrt{-23}}{2}\cdot\frac{1-\sqrt{-23}}{2}=2\cdot 3.$$

That's two different factorisations. If we saw something like this in the integers, we would assume these weren't both *prime* factorisations, i.e., that there were smaller things we could multiply to make 2 and 3.

This ring has a Norm function like the Gaussian integers. It is

$$N(a+b\sqrt{-23})=a^2+23b^2$$

The norm is clearly a non-negative number and zero only for the element 0. You can check that this norm is multiplicative, just as for the Gaussian integers. It also has the form

$$N\left(c+d\frac{1+\sqrt{-23}}{2}\right)=(c+d/2)^2+23d^2/4=c^2+cd+6d^2$$

so that it takes values in $\mathbb{Z}$.

We have $N(2)=4$ and $N(3)=9$, whereas

$$N\left(\frac{1\pm\sqrt{-23}}{2}\right)=6$$

So if there are some primes inside these factorisations (ways to break the factorisation down further), they would involve elements of the ring having norm 1, 2,or 3. It's easy to check that there *are no* elements of norm 2 or 3.

Let's consider elements of norm 1. It turns out the only such are ± 1. Obviously, these aren't good primes because any factorisation can be written in many different ways in terms of ± 1, by which I mean that

$$1\cdot x=1^2\cdot x=1^3\cdot x\ldots$$

So it would seem that there's no good way to define prime numbers for this ring. And that leaves us crying, because primes are the starting point to understanding the multiplicative structure of $\mathbb{Z}$.

Somehow, we really want there to be *something* with norm 2 or 3, so we could further factorize our equation above. There aren't any *elements* of the ring with norm 2 or 3, so Dedekind invented the idea of an *ideal number* – not a number in the ring, but something that could play this role.

13. Rings and Ideals

All the rings in this course will be commutative.

Definition 13.1 (Reminder)**.** *An ideal is a subset I of a commutative ring R such that*

(1) I is an additive subgroup of R,
(2) $rI \subset I$ for all $r \in R$,

In general, any ideal is of the form

$$I = (a_1, a_2, \ldots) = \left\{ \text{ finite sums } \sum a_i r_i : r_i \in R \right\}$$

for some $a_i \in R$, called *generators*. (Frequently, the rings we work in will have finitely generated ideals; but a general ring can have ideals not finitely generated.)

We will write (a) for the ideal generated by a; an ideal with exactly one generator is called *principal*. It is the case that for the ring $\mathbb{Z}$, every ideal has exactly one generator. In other words, the ideals of $\mathbb{Z}$ are just

$$\begin{aligned} (1) &= (-1) = \mathbb{Z}, \\ (2) &= (-2) = \text{even integers}, \\ (3) &= (-3), \\ (4) &= (-4), \\ &\cdots . \end{aligned}$$

We'll prove this later, but as an example, the $(10, 15) = (5)$ since $5 = 15 - 10$ (so that $(5) \subset (10, 15)$) and 15 and 10 are multiples of 5 (so that $(10, 15) \subset (5)$).

When we study $\mathbb{Z}$, we talk about primes, about the *units* 1 and -1, about divisibility, etc. We'd like to examine these ideas in other rings, so we'll set some terminology. All of our terminology has an interpretation in terms of ideals.

- $b \mid a$ *(b divides a)* if $a, b \in R$, $b \neq 0$, $a = bc$ for some $c \in R$. Equivalently, $(a) \subset (b)$.
- u *is a unit* if $u \mid 1$. Equivalently, $(u) = R$.
- a *and* b *are associates* if $a = bu$ for some unit $u \in R$. Equivalently, $(a) = (b)$.
- p *is irreducible* if whenever $a \mid p$, it must be that a is a unit or an associate of p. Equivalently, $(p) \subset (a) \implies (a) = (p)$ or $(a) = R$.
- p *is prime* if it is not a unit, nonzero, and whenever $p \mid ab$, then $p \mid a$ or $p \mid b$. Equivalently, $ab \in (p) \implies a \in (p)$ or $b \in (p)$.

Quick test of terminology:

(1) What are the associates of 3 in $\mathbb{Z}$?
(2) What are the units in $\mathbb{Q}[x]$?
(3) Is $x^2 + 1$ irreducible in $\mathbb{Q}[x]$?
(4) Is $x^2 + 1$ prime in $\mathbb{Q}[x]$?
(5) What are the associates of 1 in R, by another name?

In both $\mathbb{Z}$ and $k[x]$, the notions of *prime* and *irreducible* coincide. This isn't true in every ring. If you are familiar with prime and maximal ideals, compare that terminology with the terminology of prime and irreducible elements.

Example 13.2. *As we saw, the Gaussian integers of length one are* $1, -1, i, -i$*; all others have longer length. Since lengths multiply, this shows that whenever* $uv = 1$*, it must be that* u *and* v *are chosen from the list* $\{1, -1, i, -i\}$*. Hence this is the full list of units in the Gaussian integers.*

If one draws the elements of an ideal such as $(2 + i)$ in the Gaussian integers, one finds a lattice. As another example, draw the elements of the ideal $(3 + 3i, 2)$ and discover that it is a principal ideal. Which one?

14. Principal Ideal Domains

This follows Ireland and Rosen's book 'A classical introduction to modern number theory' and references are given as 'IR'. That book does what we're about to do for the integers in the first chapter, as a warmup. Then it does it in the generality here, doing it all again. The proof is the same, just in different levels of generality, so take a look at that if you'd like.

All the rings we study in this section will be integral domains. A reminder: an integral domain is a ring with no zero divisors, i.e. no nonzero elements x and y that have product $xy = 0$.

Definition 14.1. *R is a* principal ideal domain *(PID) if every ideal I is* principal*, i.e. $I = (a)$ for some $a \in R$.*

Proposition 14.2. *Let R be a PID. Then p is irreducible if and only if p is a prime.*

Proof. First, suppose that p is prime. Suppose that $a \mid p$. Let $b = p/a \mid p$. Then $ab = p$, so either $p \mid a$, in which case a is an associate of p, or $p \mid b$, in which case b is an associate of p and a is therefore a unit. So p is irreducible.

Now suppose that p is irreducible. Suppose that $p \mid ab$, i.e. $ab \in (p)$. Then, since $(p) \subset (a, p)$, one of two things happens (by the irreducibility property; note that to use this, we need to know (a, p) is principal):

(1) $(a, p) = (p)$, in which case $p \mid a$; or
(2) $(a, p) = R$, in which case $(b) = (ab, pb)$, which is a subset of (p) since $ab, pb \in (p)$. But then $(b) \subset (p)$, i.e. $p \mid b$.

Therefore p is prime. □

Now we can state our goal, which is to show unique prime factorisation in a PID (i.e. a PID is a *unique factorisation domain* or UFD).

Theorem 14.3 (PID $\implies$ UFD; IR Theorem 3, p. 12)**.** *Let R be a PID. Let S be a set of prime such that every prime has a unique associate in S. Then any $a \in R$, $a \neq 0$ can be written uniquely in the form*

$$a = u \prod_{p \in S} p^{e(p)}$$

where u is a unit and $e(p) \in \mathbb{Z}^{\geq 0}$.

A ring having the property described in this theorem is called a *unique factorization domain*, abbreviated UFD.

The proof requires a sequence of lemmas.

Lemma 14.4 (Ascending Chain Lemma)**.** *Let R be a PID. Let $(a_1) \subset (a_2) \subset (a_3) \subset \cdots$ (called an ascending chain of ideals). Then there exists a k such that $(a_k) = (a_{k+\ell})$ for $\ell = 0, 1, 2, \ldots$ (i.e. the chain stabilizes).*

Proof. Let

$$I = \bigcup_{i=1}^{\infty} (a_i).$$

Then I is an ideal (exercise), so $I = (a)$. But then $a \in (a_k)$ for some k, which implies that $I = (a) \subset (a_k)$. So we have stabilized:

$$(a_k) = (a_{k+1}) = \cdots$$

□

Lemma 14.5. *Let R be a PID. Every nonzero nonunit $a \in R$ is a product of primes.*

Proof. **First step: show that some prime divides a.**

- If a is prime, we are done.
- If not, then $a = a_1 b_1$ where a_1, b_1 are nonzero nonunits.
  - If a_1 is prime, we are done.
  - If not, then $a_1 = a_2 b_2$ where a_2, b_2 are nonzero nonunits.
    - If a_2 is prime, we are done.
    - If not, then ...

For as long as we continue this process, we generate an ascending chain of ideals

$$(a) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots$$

The chain cannot go on forever (from the Ascending Chain Lemma), so the process above must have terminated somewhere, and there is a prime dividing a.

Second step: show that a is a product of primes.

- If a is prime, we are done.
- If not, then $a = p_1 c_1$ where p_1 is prime (by the First Step)
  - If c_1 is a unit, we are done.
  - If not, then $c_1 = p_2 c_2$ where p_2 is prime (by the First Step).
    - If c_2 is a unit, we are done.
    - If not, then ...

For as long as we continue this process, we generate an ascending chain of ideals

$$(a) \subsetneq (c_1) \subsetneq (c_2) \subsetneq \cdots$$

The chain cannot go on forever (from the Ascending Chain Lemma), so the process above must have terminated somewhere, and a has the form

$$a = p_1 p_2 \cdots p_k c_k, \quad c_k \text{ a unit.}$$

Since p_1, p_2, $\ldots p_k c_k$ are primes, a is a product of primes. □

Lemma 14.6. *Let R be a PID. Let p be a prime, and let $a \neq 0$. Then there exists an n such that $p^n \mid a$ but $p^{n+1} \nmid a$.*

Proof. For each $m > 0$, if there exist b_m such that $a = p^m b_m$, then $p b_{m+1} = b_m$ (they're both equal to a/p^m). So for as many $m = 1, 2, 3, \ldots$ as this continues to happen for, we get a chain

$$(b_1) \subsetneq (b_2) \subsetneq (b_3) \subsetneq \cdots \subsetneq (b_m)$$

This chain cannot go on forever, so there must be some smallest m for which there was no b_m with that property. □

Definition 14.7. *The integer n in the previous lemma is uniquely determined by p and a, so we can write*

$$n = \mathrm{ord}_p(a)$$

Lemma 14.8. *Let R be a PID. If $a, b \in R$, $a, b \neq 0$, then*

$$\mathrm{ord}_p(ab) = \mathrm{ord}_p(a) + \mathrm{ord}_p(b).$$

Proof. Write

$$\alpha = \mathrm{ord}_p(a), \quad \beta = \mathrm{ord}_p(b).$$

Then

$$a = p^\alpha c, p \nmid c, \quad b = p^\beta d, p \nmid d.$$

Then $ab = p^{\alpha+\beta}cd$. But $p \nmid cd$. □

Proof that PID $\implies$ UFD. We already have the existence of the form of a, namely,

$$a = u \prod_{p \in S} p^{e(p)}. \tag{2}$$

and need only establish uniqueness.

Let q be a prime in S. Apply ord_q to (2), obtaining

$$\mathrm{ord}_q(a) = \mathrm{ord}_q(u) + \sum_{p \in S} e(p) \mathrm{ord}_q(p).$$

But $\mathrm{ord}_q(p)$ is 1 when $q = p$ and 0 otherwise. And $\mathrm{ord}_q(u) = 0$. So,

$$\mathrm{ord}_q(a) = e(q).$$

This shows that $e(q)$ is uniquely determined, so $u = a/\prod_{p \in S} p^{e(p)}$ is unique also. □

15. Euclidean domains

Definition 15.1. *R is a Euclidean domain if there is a function*

$$\lambda : R \setminus \{0\} \to \{0, 1, 2, \ldots\}$$

such that if $a, b \in R$, $b \neq 0$, then there exist $q, r \in R$ such that

(1) $a = bq + r$
(2) $r = 0$ or $\lambda(r) < \lambda(b)$.

Proposition 15.2 ($\mathbb{Z}$ is a Euclidean domain)**.** *For any $a, b, \in \mathbb{Z}$, there exist $q, r \in \mathbb{Z}$ such that*

(1) $a = qb + r$
(2) $r = 0$ (i.e. $b \mid a$) or $|r| < |b|$.

Proof. The set of non-negative $a - qb$ has a least element. It is less than b. □

Proposition 15.3 ($k[x]$ is a Euclidean domain)**.** *For any $a(x)$, $b(x) \in k[x]$, there exist $q(x), r(x) \in k[x]$ such that*

(1) $a(x) = q(x)b(x) + r(x)$
(2) $r(x) = 0$ *or* $\deg r(x) < \deg b(x)$

Proof. Ireland and Rosen, p. 7 Lemma 2. Same idea as for q, but argue *least degree*. □

Proposition 15.4 (Euclidean domain $\implies$ PID)**.** *Every Euclidean domain is a PID.*

Proof. (See Ireland and Rosen Proposition 1.3.1). Let I be an ideal. The set

$$\{\lambda(b) : b \in I, b \neq 0\}$$

has a least element, $\lambda(a)$.

Claim: $I = (a)$. This has two parts:

(1) $(a) \subset I$: This is clear.
(2) $I \subset (a)$: For this, assume $b \in I$. Then there are some $q, r \in R$ such that $b = qa + r$ and $r = 0$ or $\lambda(r) < \lambda(a)$. Then $r = b - qa \in I$, so $\lambda(r) \geq \lambda(a)$ (by the definition of a). This implies $r = 0$ and hence $b = qa \subset (a)$.

□

Proposition 15.5. $\mathbb{Z}[i]$ *is a Euclidean domain.*

Proof. Suppose we try to divide $\alpha = a + bi$ by $\beta = c + di \neq 0$. Then in $\mathbb{C}$, the result is

$$\alpha/\beta = r + si$$

for some $r, s \in \mathbb{R}$ (actually, they are in $\mathbb{Q}$).

A good guess for a 'close' Gaussian integer is $\delta = m + ni$ chosen such that

$$|r - m| \leq \frac{1}{2}$$
$$|s - n| \leq \frac{1}{2}$$

Set $\rho = \alpha - \beta\delta$ (this is the 'remainder'). If $\rho = 0$, we are done. Otherwise, we can verify that

$$\begin{aligned} N(\rho) &= N(\alpha - \beta\delta) \\ &= N\left(\beta\left(\frac{\alpha}{\beta} - \delta\right)\right) \\ &= N(\beta)N\left(\frac{\alpha}{\beta} - \delta\right) \\ &\leq \frac{1}{2}N(\beta) \\ &< N(\beta) \end{aligned}$$

In the first inequality, we are using the fact that $\frac{\alpha}{\beta} - \delta$ has coordinates not exceeding $1/2$, by our choice of δ.

In the last inequality, we use the fact that $\beta \neq 0$.

Thus we have verified that $\mathbb{Z}[i]$ has a Euclidean algorithm, where the function λ in the definition of a Euclidean domain is, in this case, the norm. □

Ireland and Rosen prove that $\mathbb{Z}[\omega]$ is a Euclidean domain. For your homework, find the units, etc.

16. The GCD and linear Diophantine equations

We will give two definitions of a greatest common divisor, and show that they are equivalent in a PID.

Definition 16.1. *An element $d \in R$ is a* greatest common divisor *of $a, b \in R$ if*

(1) $d \mid a$, $d \mid b$
(2) If $d' \mid a$, $d' \mid b$, then $d' \mid d$.

Any two gcd's of $a, b \in R$ are associate.

Definition 16.2. *In a PID, $(a, b) = (d)$ for some d. This d is called a* greatest common divisor

Proposition 16.3. *These two definitions are equivalent.*

Proof. Suppose that $(a, b) = (d)$. Then $(a) \subset (d)$ and $(b) \subset (d)$, i.e. $d \mid a$ and $d \mid b$. If here is some d' such that $d' \mid a$ and $d' \mid b$, then

$$(d') \supset (a), (b) \implies (d') \supset (a, b) = (d) \implies d' \mid d.$$

Now suppose that $d \mid a$, $d \mid b$ and whenever $d' \mid a, d' \mid b$, then $d' \mid d$. Then $(a) \subset (d)$, $(b) \subset (d)$, so $(a, b) \subset (d)$. Now $(a, b) = (d')$ for some

d' since we are in a PID. As above, this implies that $d' \mid a$ and $d' \mid b$, so that $d' \mid d$. Hence $(d) \subset (d') = (a, b)$. Therefore $(a, b) = (d)$. $\square$

In particular, if $d = \gcd(a, b)$, then $(d) = (a, b)$, so there exist some x_0 and y_0 so that

$$ax_0 + by_0 = d.$$

Also, for any x, y, we have

$$d \mid ax + by.$$

This allows us to describe the solutions in x, y to any equation $ax+by = m$. In particular, by the last remark, if $d \nmid m$, then there are no solutions.

Now suppose that $d \mid m$. In this case we may write $m = kd$, and $(x, y) = (kx_0, ky_0)$ is one solution. If (x', y') is another solution, then it must be that

$$a(kx_0 - x') + b(ky_0 - y') = kd - kd = 0$$

Hence $a(kx_0 - x') = -b(ky_0 - y')$, which implies that,

$$(kx_0 - x', ky_0 - y') = s\left(\frac{-b}{d}, \frac{a}{d}\right)$$

for some $s \in R$. In conclusion, then, the full set of solutions to $ax+by = m$ is

$$\left\{k(x_0, y_0) + s\left(\frac{-b}{d}, \frac{a}{d}\right) : s \in R\right\}$$

where $d = \gcd(a, b)$, $k = m/d$ and (x_0, y_0) is any one solution. In other words, we have parametrised the full set of solutions in terms of one single solution. Therefore, in a practical situation, the task is to find a single solution. The Euclidean algorithm serves this purpose.

17. Congruences Classes and $\mathbb{Z}/m\mathbb{Z}$

By a *congruence class modulo* m, we mean the following set:

$$\begin{aligned}\overline{a} &= \{n \in \mathbb{Z} : n \equiv a \pmod{m}\} \\ &= \{a + km : k \in \mathbb{Z}\}.\end{aligned}$$

In other words, the collection of all integers with a given remainder when divided by m.

Then $\mathbb{Z}/m\mathbb{Z}$ is the ring of congruence classes modulo m. It has cardinality m, since, for example, a complete set of representatives (one representative for each class) is $0, 1, \ldots, m - 1$.

If $\overline{a} = \overline{b}$, we write

$$a \equiv b \pmod{m}.$$

There is a map

$$\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}, \quad a \mapsto \overline{a}.$$

This map is a ring homomorphism, and checking this is a good exercise.

We will henceforth generally omit the line and write a for $\overline{a}$.

18. Solving Linear Congruences

Suppose we'd like to solve the congruence $ax \equiv b \pmod{m}$ in the variable x, i.e. find all congruence classes x which solve the equation. This is equivalent to finding x and y in the integers such that

$$ax + ym = b,$$

which we've already discussed a couple of sections ago. There are no solutions if $\gcd(a, m) \nmid b$. Otherwise,

$$(x, m) = k(x_0, y_0) + \frac{s}{d}(-m, a)$$

for $s \in \mathbb{Z}$, where $d = \gcd(a, m)$, $k = b/d$, and $ax_0 + y_0 m = b$.

How many solutions are there in $\mathbb{Z}/m\mathbb{Z}$? The possible x are

$$x \equiv \frac{bx_0}{d} - \frac{sm}{d} \pmod{m}$$

as s ranges through the integers. As s ranges, the value $\frac{sm}{d}$ hits d different values in $\mathbb{Z}/m\mathbb{Z}$. Therefore there are exactly d solutions, when $d = \gcd(a, m) \mid b$.

Corollary 18.1. *If $\gcd(a, m) = 1$, then $ax \equiv b \pmod{m}$ has one solution. In particular, $ax \equiv 1 \pmod{m}$ has one solution, so a is invertible.*

The corollary implies that $\mathbb{Z}/m\mathbb{Z}$ has $\phi(m)$ units (i.e. the units are exactly those $1 \leq a \leq m$ such that $\gcd(a, m) = 1$).

If $m = p$ is prime, then $\mathbb{Z}/p\mathbb{Z}$ has $p - 1$ units, i.e. everything except 0 is a unit. Therefore $\mathbb{Z}/p\mathbb{Z}$ is a field.

Otherwise, if m is not a prime, then $m = nk$ for some n and k nonzero, so $nk \equiv 0 \pmod{m}$, whereas $n \not\equiv 0 \pmod{m}$ and $k \not\equiv 0 \pmod{m}$. In other words, $\mathbb{Z}/m\mathbb{Z}$ has zero divisors!

We will use the notation

$$(\mathbb{Z}/m\mathbb{Z})^* = U(\mathbb{Z}/m\mathbb{Z})$$

for the group of units of $\mathbb{Z}/m\mathbb{Z}$.

Corollary 18.2 (Euler's Theorem)**.** *If $\gcd(a, m) = 1$, then*

$$a^{\phi(m)} \equiv 1 \pmod{m}.$$

Proof. The units of $\mathbb{Z}/m\mathbb{Z}$ form a group of order $\phi(m)$. □

An immediate corollary of Euler's Theorem is the following.

Corollary 18.3 (Fermat's Little Theorem)**.** *If p is prime, and $p \nmid a$, then*

$$a^{p-1} \equiv 1 \pmod{p}.$$

19. Chinese Remainder Theorem

This theorem dates as far back as Sun Tsu in the 1st century, according to some sources, and may be older still. Here's the form you may have seen it in:

Theorem 19.1 (Chinese Remainder Theorem)**.** *If $b_i \in \mathbb{Z}$ and $m_i \in \mathbb{Z}^{>0}$ for $i = 1, \ldots, t$, then the system of equations*

$$\begin{aligned} x &\equiv b_1 \pmod{m_1} \\ x &\equiv b_2 \pmod{m_2} \\ &\vdots \\ x &\equiv b_t \pmod{m_t} \end{aligned}$$

always has a solution and any two solutions are congruenct modulo the product $m_1 m_2 \cdots m_t$.

This could be restated as

Theorem 19.2. *Let $m_1, \ldots, m_t$ be coprime integers. Then we have a ring isomorphism*

$$\mathbb{Z}/m_1 \cdots m_t\mathbb{Z} \cong \mathbb{Z}/m_1\mathbb{Z} \times \cdots \times \mathbb{Z}/m_t\mathbb{Z}.$$

It's not much harder to state this for general commutative rings with identity.

Theorem 19.3 (Chinese Remainder Theorem for General Rings)**.** *Let $I_1, I_2, \ldots, I_k$ be ideals of a ring R. The following map is a ring homomorphism:*

$$\begin{aligned} \phi : R &\to R/I_1 \times R/I_2 \times \cdots \times R/I_k \\ \phi : r &\mapsto (r + I_1, r + I_2, \ldots, r + I_k). \end{aligned}$$

and the kernel of ϕ is $\cap_{i=1}^k I_i$. If the ideals are all comaximal (i.e. $I_i + I_j = R$ for all i, j)[2]*, then the map is surjective and the kernel is also equal to the product $\prod_{i=1}^k I_i$.*

[2]This is the generalization of coprimality

In particular, if they're all coprime,

$$R/\prod I_i \sim \prod_i R/I_i$$

Proof. (Standard; following, for example, Dummit and Foote, Section 7.6)

We will assume $k = 2$; the proof for larger k is the same (or you can do it by induction from the proof with $k = 2$). The proof for more general rings is also very similar.

First, ϕ is a ring homomorphism. This is because it is a homomorphism in each factor.

Second, the kernel of this map is

$$\{r : r \in I_1, r \in I_2\} = I_1 \cap I_2.$$

If $I_1 + I_2 = R$, then there are $r_1 \in I_1$, $r_2 \in I_2$ with $r_1 + r_2 = 1$ Therefore any $r \in I_1 \cap I_2$ satisfies $r = rr_1 + rr_2 \in I_1 I_2$. So $I_1 \cap I_2 \subset I_1 I_2$. But the reverse inclusion is always true. So the kernel in this case is $I_1 I_2$.

Third, the map is surjective. Let $(x, y) \in R/I_1 \times R/I_2$. Then

$$\begin{aligned} s &:= xr_1 + yr_2 \\ &= x(1 - r_2) + yr_2 \\ &\equiv x \pmod{I_2} \end{aligned}$$

and

$$\begin{aligned} s &= xr_1 + yr_2 \\ &= xr_1 + y(1 - r_1) \\ &\equiv y \pmod{I_1} \end{aligned}$$

So

$$s \mapsto (x \bmod m_1, y \bmod m_2)$$

and we have shown surjectivity. □

In particular, we have a group isomorphism

$$(\mathbb{Z}/m_1 \cdots m_t\mathbb{Z})^* \cong (\mathbb{Z}/m_1\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/m_t\mathbb{Z})^*.$$

In particular (more particular?), their cardinalities are equal, so

$$\phi(m_1 \cdots m_t) = \phi(m_1)\phi(m_2) \cdots \phi(m_t).$$

This constitutes another proof of the multiplicativity of ϕ.

20. The p-adic numbers

The p-adic numbers are motivated by a grand analogy, beginning with the observation that $\mathbb{Z}$ is a little like $\mathbb{C}[x]$, the ring of polynomials with complex coefficients. For example, both of them are unique factorization domains, so they have a list of primes. Here's a sort of chart of the analogy:

$\mathbb{Z}$	$\mathbb{C}[x]$
$\mathbb{Q}$	$\mathbb{C}(x)$
$q \in \mathbb{Q}$ has the form $\frac{a}{b}$, $a, b \in \mathbb{Z}$	$f(x) \in \mathbb{C}(x)$ has the form $\frac{p(x)}{q(x)}$, for $p(x), q(z) \in \mathbb{C}[x]$
$\mathbb{Z}$ is a UFD	$\mathbb{C}[x]$ is a UFD
The primes of $\mathbb{Z}$ are $2, 3, 5, 7, \ldots$	The primes of $\mathbb{C}[x]$ are $(x - \alpha)$ for $\alpha \in \mathbb{C}$
We can write a number base p	A rational function has a Taylor expansion at $x = \alpha$
$m = a_0 + a_1 p + \cdots + a_n p^n$	$p(x) = a_0 + a_1(x - \alpha) + a_2(x - \alpha)^2 + \cdots + a_n(x - \alpha)^n$
We can ask to what power does p divide m?	We can ask to what order is α a zero of $p(x)$?

Rational functions $f(x) \in \mathbb{C}(x)$ also have a Taylor expansion:

$$f(x) = \sum_{i=n}^{\infty} a_i (x - \alpha)^i$$

for some n, possibly negative. This converges in some small positive radius around α. This now begs the question: what about a rational number? Does it have a p-expansion? Specifically, can we write it in the form

$$a/b = \sum_{i=n}^{\infty} a_i p^i$$

for some n, possibly negative? (Note that this is different than the usual 'decimal expansion' of a rational, which may be infinite in the direction of negative powers of p, but always finite in the direction of positive powers of p. Here the series has a finite number of terms of negative powers of p, and possibly infinitely many terms of positive powers of p.)

Let's do some examples, and see if we can make sense, at least formally, of such an idea. Let's let $p = 2$.

Then,

$$1 = 1 \cdot 2^0 + 0 \cdot 2^1 + 0 \cdot 2^2 + \cdots$$

So far so good. Also,

$$1/2 = 1 \cdot 2^{-1} + 0 \cdot 2^0 + 0 \cdot 2^1 + \cdots$$

Another example:

$$3 = 1 \cdot 2^0 + 1 \cdot 2^1 + 0 \cdot 2^2 + \cdots$$

And we're still ok with

$$3/2 = 1 \cdot 2^{-1} + 1 \cdot 2^0 + 0 \cdot 2^1 + \cdots$$

But here's a trickier one: What about -1?

Well,

$$-1 = \frac{1}{1-2} = 1 + 2 + 2^2 + 2^3 + 2^4 + \cdots$$

This doesn't make a lot of sense in the reals, since this sequence diverges. But purely formally, this is 'correct' and consistent. For example,

$$\begin{aligned} 1 + (-1) &= 1 + (1 + 2 + 2^2 + 2^3 + \cdots) \\ &= 2 + 2 + 2^2 + 2^3 + \cdots \\ &= 2^2 + 2^2 + 2^3 + \cdots \\ &= 2^3 + 2^3 + 2^4 + \cdots \\ &= \cdots \\ &= 0. \end{aligned}$$

This is purely formal! (Although, in order to make this make sense, we can define an appropriate metric under which these sorts of series will converge.)

Let's do a few more examples. What about $1/3$? We could try

$$1/3 = 1/(1+2) = 1 - 2 + 2^2 - 2^3 + 2^4 - 2^5 + \cdots$$

But it would be really nice to have coefficients of the powers of p which lie strictly in the range $0 \leq a_i < p$. To get ride of those negative coefficients here is going to be a bit of a pain. Let's try something a bit different:

$$\begin{aligned} 1/3 &= 1 + \frac{2}{1-2^2} \\ &= 1 + 2(1 + 2^2 + 2^4 + 2^6 + \cdots) \\ &= 1 + 2 + 2^3 + 2^5 + 2^7 + \cdots \end{aligned}$$

Great! So it looks like we really can get a p-expansion for any rational number! Now, guided again by the analogy, note that we have an

injection

$$\mathbb{C}(x) \hookrightarrow \mathbb{C}((x-\alpha))$$

where the latter is the field of formal Laurent series, i.e. all things that look like

$$\sum_{i=n}^{\infty} a_i (x-\alpha)^i$$

for some n. If we define the collection of formal p-expansions, called *p-adics*, as

$$\mathbb{Q}_p = \left\{ \sum_{i=n}^{\infty} a_i p^i \middle| 0 \leq a_i \leq p-1, n \in \mathbb{Z} \right\}$$

then it turns out we get an injection

$$\mathbb{Q} \hookrightarrow \mathbb{Q}_p.$$

(We have not proven this yet, but keep reading...)

21. The projective limit definition of the p-adic integers.

Working with p-expansions, while natural from the motivational viewpoint given above, is a bit of a pain in practice. There's a much better, if apparently more abstract, way to define the p-adics. To do so, we will actually start by defining the *p-adic integers*, which are meant to be the p-expansions consisting only of non-negative powers of p.

Suppose we consider such an expansion:

$$\alpha = \sum_{i=0}^{\infty} a_i p^i.$$

Let's set some notation for the partial sums, which are integers:

$$S_n = \sum_{i=0}^{n-1} a_i p^i \in \mathbb{Z}$$

Let's consider the residue $\overline{s_n}$ of s_n modulo p^n. That is,

$$\overline{s_n} \in \mathbb{Z}/p^n\mathbb{Z}.$$

Then these partial sum residues $\overline{s_n}$ have a special property with respect to the following chain of ring homomorphisms:

$$\mathbb{Z}/p\mathbb{Z} \xleftarrow{\pi_1} \mathbb{Z}/p^2\mathbb{Z} \xleftarrow{\pi_2} \mathbb{Z}/p^3\mathbb{Z} \xleftarrow{\pi_3} \cdots$$

The homomorphism π_n is the map which considers a residue modulo p^{n+1} to be a residue modulo p^n, for example, if $p = 3$, then

$$\pi_1(10 \bmod 9) = 1 \bmod 3.$$

To wit[3], the special property is that

$$\pi_n(\overline{s_{n+1}}) = \overline{s_n}.$$

We will call this *coherence*. An example may clarify: if $\alpha = 1 + p^2 + 2p^3 + 3p^4 + \cdots$, then

$$\pi_4(1 + p^2 + 2p^3 + 3p^4) = 1 + p^2 + 2p^3 \pmod{p^4}.$$

This motivates us to start afresh with the following definition, which is an instance of a *projective limit* (also called an *inverse limit*).

Definition 21.1. *Let us define*

$$\varprojlim_n \mathbb{Z}/p^n\mathbb{Z} = \left\{ (x_n)_{n=1}^\infty \in \prod_{n=1}^\infty \mathbb{Z}/p^n\mathbb{Z} \middle| \pi_n(x_{n+1}) = x_n \text{ for } n = 1, 2, \ldots \right\}.$$

That is, we are defining a subset of an infinite product of rings, a subset which is characterised by the coherence property which the $\overline{s_n}$ satisfied. It is our hope that this subset is in fact a ring. But first, we will show that, as a set, it is in bijection with the p-expansions having only non-negative powers of p.

Proposition 21.2. *We have a bijection*

$$\left\{ \sum_{i=0}^\infty a_i p^i \middle| 0 \leq a_i \leq p-1 \right\} \leftrightarrow \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}.$$

Proof. Given a sum $\sum_{i=0}^\infty a_i p^i$, we will map it to the element

$$(\overline{s_n})_{n=1}^\infty,$$

where the $\overline{s_n}$ were defined at the beginning of this section. It is clear that

$$\pi_n(\overline{s_{n+1}}) = \overline{s_n}.$$

So that this element is indeed an element of $\varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$.

To show this is a bijection (both surjective and injective), we should start with an element $(x_n)_{n=1}^\infty$ of $\prod_{n=1}^\infty \mathbb{Z}/p^n\mathbb{Z}$ having the property that $\pi_i(x_{i+1}) = x_i$, and show that there is a unique series which has partial sums

$$s_n \equiv x_n \pmod{p^n}.$$

To be able to do so, it suffices to know that any residue class a modulo p^n can be uniquely represented in the form

$$a \equiv a_0 + a_1 p + \cdots a_{n-1} p^{n-1} \pmod{p^n},$$

[3]To woo?

where $0 \leq a_i < p$. For, in this case the coherence property implies that if

$$x_n \equiv a_0 + a_1 p + \cdots + a_{n-1} p^{n-1} \pmod{p^n}$$

then

$$x_{n-1} = \pi_{n-1}(x_n) \equiv a_0 + a_1 p + \cdots + a_{n-2} p^{n-2} \pmod{p^{n-1}},$$

i.e. the various expressions of the x_n are all partial sums of the *same* infinite series, and this series is unique. That is, we may define the series uniquely as

$$\sum_{i=0}^{\infty} a_i p^i.$$

This leaves us with the task of showing that a residue class a modulo p^n has a unique expression in the form

$$a \equiv a_0 + a_1 p + \cdots a_{n-1} p^{n-1} \pmod{p^n}.$$

We do so by induction. The case of $n = 1$ is easy: any a has a unique residue a_0 modulo p which satisfies $0 \leq a_i \leq p-1$. Now suppose it has been proven for $n-1$. Then, writing uniquely

$$a \equiv a_0 + a_1 p + a_2 p^2 + \cdots + a_{n-2} p^{n-2} \pmod{p^{n-1}}$$

we find that

$$a \equiv a_0 + a_1 p + a_2 p^2 + \cdots + a_{n-2} p^{n-2} + g p^{n-1} \pmod{p^n}$$

for some g (since, letting g range among integers, we obtain all possible lifts of a mod p^{n-1} to a residue modulo p^n). But in fact, each $0 \leq g \leq p-1$ gives a different residue class modulo p^n, only one of which is congruent to a. Hence g is determined uniquely in the range $0 \leq g \leq p-1$, which is what was required to prove. □

Proposition 21.3. $\varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$ *is a subring of* $\prod_{n=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z}$

Proof. $\varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$ inherits its ring operations from the product, i.e. addition and multiplication are coordinatewise:

$$(x_n)_{n=1}^{\infty} + (y_n)_{n=1}^{\infty} = (x_n + y_n)_{n=1}^{\infty}$$
$$(x_n)_{n=1}^{\infty} \times (y_n)_{n=1}^{\infty} = (x_n y_n)_{n=1}^{\infty}$$

Note that it does indeed contain the additive identity, which is

$$(0)_{n=1}^{\infty}$$

and the multiplicative identity,

$$(1)_{n=1}^{\infty}.$$

Since these have the property that $\pi_n(0) = 0$ and $\pi_n(1) = 1$. To see that it is a subring, it suffices to check that it is closed under the ring

operations. Since the π_n are ring homomorphisms, this just entails a brief check that the coherence property is preserved:

$$\pi_{n-1}(x_n + y_n) = \pi_{n-1}(x_n) + \pi_{n-1}(y_n) = x_{n-1} + y_{n-1}$$
$$\pi_{n-1}(x_n y_n) = \pi_{n-1}(x_n)\pi_{n-1}(y_n) = x_{n-1}y_{n-1}$$

□

At this point, it is worth checking (as an exercise) that these operations, applied to the series representation, are just addition and multiplication of series. Adding and multiplying series is a bit of a pain, because of all the carrying.

That will then justify writing

$$\mathbb{Z}_p = \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}.$$

That is, we can use the notation $\mathbb{Z}_p$ for the projective limit (it's an abuse of notation, as we already used it for series).

22. Back to the p-adic rationals

We originally wanted to consider the collection of all p-expansions, possibly including finitely many negative powers of p. We denoted this by $\mathbb{Q}_p$. We can extend the ring operations from $\mathbb{Z}_p$ (thought of as p-expansions) to $\mathbb{Q}_p$ by writing every element of $\mathbb{Q}_p$ in the form $p^{-m}g$ where $g \in \mathbb{Z}_p$ and $m \geq 0$. If we have two elements

$$p^{-m_1}g_1 \text{ and } p^{-m_2}g_2$$

then we can rewrite them in the form

$$p^{-N}g_1' \text{ and } p^{-N}g_2'$$

where $N \geq m_1, m_2$ (I'm just taking common denominators here). Then we can define addition

$$p^{-m_1}g_1 + p^{-m_2}g_2 = p^{-N}(g_1' + g_2')$$

and multiplication

$$p^{-m_1}g_1 p^{-m_2}g_2 = p^{-m_1-m_2}g_1g_2,$$

as extensions of the operations on $\mathbb{Z}_p$. If we define $\mathbb{Q}_p$ this way, then we obtain the following.

Proposition 22.1. *$\mathbb{Q}_p$ is a the field of fractions of $\mathbb{Z}_p$.*

Proof. We should check that the operations defined above make $\mathbb{Q}_p$ into a ring; this is straightforward. To see that it is a field, we need all non-zero elements to be invertible. This is a bit more interesting, and writing down inverses was given as a homework question. Finally, since $\mathbb{Z}_p$ injects into $\mathbb{Q}_p$, and all $f \in \mathbb{Q}_p$ are of the form $p^{-m}g$, $m \geq 0$ and $g \in \mathbb{Z}_p$ (i.e. quotients of elements of $\mathbb{Z}_p$), $\mathbb{Q}_p$ is the field of fractions of $\mathbb{Z}_p$. □

Proposition 22.2.

$$\mathbb{Z} \hookrightarrow \mathbb{Z}_p$$
$$\mathbb{Q} \hookrightarrow \mathbb{Q}_p$$

Proof. The first injection is given by

$$a \mapsto (a \bmod p, a \bmod p^2, a \bmod p^3, \ldots).$$

This is a ring homomorphism. Then, since $\mathbb{Q}$ is the field of fractions of $\mathbb{Z}$, it must inject into the field of fractions of $\mathbb{Z}_p$. □

23. Absolute Values

This perspective is well-developed in Gouvea's book "p-adic Numbers"; we will follow that for now.

Definition 23.1. *An* absolute value *on a field F is a function*

$$|\cdot| : F \to \mathbb{R}^{\geq 0}$$

satisfying

(1) $|x| = 0$ *if and only if* $x = 0$
(2) $|xy| = |x||y|$ *for all* $x, y \in F$
(3) $|x + y| \leq |x| + |y|$ *for all* $x, y \in F$

In addition, it is non-archimedean *if also*

$$|x + y| \leq \max\{|x|, |y|\} \text{ for all } x, y \in F.$$

(Otherwise, archimedean*.)*

Being non-archimedean implies item (3); it is stronger than the triangle inequality. On any field, you have the *trivial* absolute value which is 0 at 0 and 1 elsewhere. The absolute value you're used to is an example.

There's a 'logarithmic' version of this definition.

Definition 23.2. *A* valuation *on a field F is a function*

$$v : F \to \mathbb{Z}^{\geq 0} \cup \{\infty\}$$

such that

(1) $v(x) = \infty$ *if and only if* $x = 0$
(2) $v(xy) = v(x) + v(y)$ *for all* $x, y \in F$
(3) $v(x + y) \geq \min\{v(x), v(y)\}$ *for all* $x, y \in F$

For us, the important example is the following. We used the kind of dumb notation $\mathrm{ord}_p(n)$ when proving unique factorisation (probably because I was following Ireland and Rosen there), but let's extend it to $\mathbb{Q}$ and write that as $v_p(n)$ instead, and call it the *p-adic valuation*. Any $x \in \mathbb{Q}$ is of the form $x = p^{v_p(x)}\frac{a}{b}$ where $p \nmid ab$.

Then we get

Definition 23.3. *The* p-adic absolute value *on* $\mathbb{Q}$ *is*

$$|x|_p = \begin{cases} p^{-v_p(x)} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

Proposition 23.4. *This is a non-archimedean absolute value on* $\mathbb{Q}$.

Aside: you can have fun using the degree as a valuation on $\mathbb{C}(x)$

An absolute value induces a metric, or a measure of distance between two things.

Definition 23.5. *A metric on a set* X *is a function*

$$d : X \times X \to \mathbb{R}^{\geq 0}$$

such that

(1) $d(x, y) = 0$ *if and only if* $x = y$ *for* $x, y \in X$
(2) $d(x, y) = d(y, x)$ *for* $x, y \in X$
(3) $d(x, z) \leq d(x, y) + d(y, z)$ *for* $x, y, z \in X$

If we have an absolute value on F, then $d(x, y) = |x - y|$ gives a metric on F. If the absolute value is non-archimedean, then we also get the following property, called the *ultrametric inequality*:

$$d(x, y) \leq \max\{d(x, z), d(z, y)\}.$$

Our metric is then called an *ultrametric*.

Proposition 23.6. *If* $x, y \in F$ *and* $|\cdot|$ *is an ultrametric, then for* $|x| \neq |y|$,

$$|x + y| = \max\{|x|, |y|\}.$$

Another way to say this is that *all triangles are isosceles*. Let's do a p-adic example. In the 5-adics, look at a lengths of all the distances between 1, 1/5, 2/15 and 7/15. Note that is does not make sense to think of $\mathbb{Q}$ as a line anymore!

A notion of distance gives a *topology*. We won't do an intro course to toplogy here, but you will recognise it as we go if you have seen it.

To understand the topology, or the metric, it is helpful to explore the properties of open and closed balls:

$$B(a,r) = \{x \in F : |a-x| < r\}, \quad \overline{B}(a,r) = \{x \in F : |a-x| \leq r\}.$$

For homework, verify some of the crazy properties:

(1) Every point contained in an (open or closed) ball is a centre of that ball.
(2) Balls are both open and closed (except the empty ball).
(3) Two (both open or both closed) balls are either disjoint or one is contained in the other.

Throw your euclidean intuition out the window, please.

On $\mathbb{Q}$, we've seen the usual absolute value, the trivial absolute value and the p-adic absolute values. Two absolute values are considered equivalent if they generate the same topology, i.e. the same open sets. Another equivalent definition is that for any $x \in F$, $|x|_1 < 1$ if and only if $|x|_2 < 1$. It is a famous theorem of Ostrowski that up to equivalence, this is the whole collection of absolute values on $\mathbb{Q}$. All these absolute values are related by the *product formula*:

$$\prod_{p \leq \infty} |x|_p = 1, \text{ for all } x \in \mathbb{Q}^*.$$

where $|\cdot|_\infty$ is the archimedean, usual absolute value. We think of ∞ as the 'archimedean prime' or 'prime at infinity.'

We won't do it in this course, but it turns out that one obtains the p-adics $\mathbb{Q}_p$ as a completion of $\mathbb{Q}$ under the p-adic absolute value. Unfortunately, unlike the case for $\mathbb{C}$ as the completion under the archimedean absolute value, $\mathbb{Q}_p$ isn't algebraically closed. You have to close it, then complete it again, and then it's closed.

24. Solving non-linear congruences

Let f be a polynomial and consider the equation

$$f(x) \equiv 0 \pmod{m} \tag{3}$$

Write $m = p_1^{a_1} \cdots p_\ell^{a_\ell}$. Then, by the Chinese Remainder Theorem, (3) is solvable if and only if $f(x) \equiv 0 \pmod{p_i^{a_i}}$ is solvable for each $i = 1, \ldots, \ell$. To be more precise about using CRT, here's the argument:

Suppose we have a b_i which solves $f(x) \equiv 0 \pmod{p_i^{a_i}}$ for each i. Then, by the ring homomorphism of CRT, there exists some b such that $b \equiv b_i \pmod{p_i^{a_i}}$ for all i. But also by the ring homomorphism of CRT, $f(b) \equiv 0 \pmod{m}$. The other direction is clear.

Therefore, it suffices to look for solutions modulo prime powers.

Now we can reap the reward of learning about the p-adics. They "package" information modulo all powers of p into one object. Here's an example of the power of the language:

Proposition 24.1. *Let $f \in \mathbb{Z}[x]$, and let p be a prime. Then*

$$f(x) \equiv 0 \pmod{p^n}$$

is solvable for all $n = 1, 2, \ldots$ if and only if

$$f(x) = 0$$

is solvable in $\mathbb{Z}_p$.

Proof. In $\mathbb{Z}_p$, $f(x) = 0$ means that

$$(f(x) \bmod p, f(x) \bmod p^2, \ldots) = (0 \bmod p, 0 \bmod p^2, \ldots)$$

So if there's a p-adic integer solution x, then in particular, it gives a solution mod p^i for each $i = 1, 2, \ldots$, by looking at the appropriate coordinate.

Conversely, suppose that $f(x_i) \equiv 0 \pmod{p^i}$ for each $i = 1, 2, \ldots$. We may be lucky and find that the sequence

$$(x_i)_{i=1}^{\infty}$$

is already in $\varprojlim^n \mathbb{Z}/p^n\mathbb{Z}$, i.e. it has the coherence property. Then this would give the p-adic solution.

More likely than not, though, it is not coherent. So we have more work to do. Viewing (x_i) as a sequence in $\mathbb{Z}$, it has infinitely many terms which are congruent mod p (by the pigeonhole principle), say to y_1 modulo p. Then

$$f(y_1) \equiv 0 \pmod{p}.$$

(Why? Because $f(x) \equiv 0 \pmod{p^k}$ for some k, and some $x \equiv y_1 \pmod{p}$.)

These infinitely many terms form a subsequence. From among the terms in that subsequence, by pigeonhole principle once again, there are infinitely many which are congruent modulo p^2, say to y_2 modulo p^2. Then

$$\begin{aligned} f(y_2) &\equiv 0 \pmod{p^2} \\ y_2 &\equiv y_1 \pmod{p} \end{aligned}$$

Now we have a subsequence of a subsequence. From among these terms, find infinitely many congruent modulo p^3, ...

Repeating this argument, we construct y_i for $i = 1, 2, 3, \ldots$. Then we may define

$$(y_i)_{i=1}^{\infty} \in \prod_{i=1}^{\infty} \mathbb{Z}/p^n\mathbb{Z}.$$

By construction, $(y_i) \in \varprojlim_n \mathbb{Z}/p^n\mathbb{Z}$ (i.e. we have guaranteed the coherence property) and

$$f(y_i) \equiv 0 \pmod{p^i}$$

for all $i = 1, 2, \ldots$, so that this is a p-adic zero of f. □

Using essentially the same proof, we obtain the following proposition.

Proposition 24.2. *Let $f \in \mathbb{Z}[x_1, \ldots, x_m]$, and let p be a prime. Then*

$$f(x_1, \ldots, x_m) \equiv 0 \pmod{p^n}$$

is solvable for all $n = 1, 2, \ldots$ if and only if

$$f(x_1, \ldots, x_m) = 0$$

is solvable in $\mathbb{Z}_p$.

We'll refer to the following as *Hensel's Lemma*, although it is just one among many similar statements by that name (most more general than this).

Proposition 24.3 (Hensel's Lemma the First)**.** *Let $f(x) \in \mathbb{Z}[x]$. Suppose that $b \in \mathbb{Z}$ is such that*

$$f(b) \equiv 0 \pmod{p}, \text{ and}$$
$$f'(b) \not\equiv 0 \pmod{p}$$

Then there exists an $\alpha \in \mathbb{Z}_p$ such that $\alpha \equiv b \pmod{p}$ and $f(\alpha) = 0$.

Instead of proving this directly, we'll prove a more detailed statement (which we'll also call Hensel's Lemma), which implies this. This more detailed version will be particularly useful for computation.

Proposition 24.4 (Hensel's Lemma the Second)**.** *Let $f(x) \in \mathbb{Z}[x]$. Then*

(1) If $f(b) \equiv 0 \pmod{p^a}$ and $f'(b) \not\equiv 0 \pmod{p}$, then there exists b' such that

$$f(b') \equiv 0 \pmod{p^{a+1}}, \text{ and}$$
$$b' \equiv b \pmod{p^a}$$

Furthermore, $b' = b + tp^a$ where

$$tf'(b) \equiv \frac{-f(b)}{p^a} \pmod{p}.$$

(2) If $f(b) \equiv 0 \pmod{p}$ *and* $f'(b) \equiv 0 \pmod{p}$*, then either all* $b' \equiv b \pmod{p^a}$*, or none of them, satisfy*

$$f(b) \equiv 0 \pmod{p^{a+1}}.$$

The proof is exactly Newton's Method under the p-adic metric. To illustrate the relationship, we'll demonstrate Newton's Method first.

Newton's Method is a method to find a root of a polynomial $f(x) \in \mathbb{Q}[x]$ by starting with an approximation and refining it at each step. One can prove that, after a suitable first guess, the method converges to a root. Let b be our first guess. The key is the Taylor series expansion of $f(x)$ around b:

$$f(x) = f(b) + f'(b)(x-b) + \frac{f''(b)}{2}(x-b)^2 + \cdots$$

To improve our guess from b to $b' = b + \epsilon$, we aim for $f(b+\epsilon) = 0$, i.e.

$$0 = f(b') = f(b) + f'(b)\epsilon + \cdots$$

Since the remaining terms are higher order, we can *approximate* by choosing

$$\epsilon = -\frac{f(b)}{f'(b)}, \tag{4}$$

as long as $f'(b) \neq 0$. In other words,

$$b' = b - \frac{f(b)}{f'(b)}.$$

This doesn't actually get us $f(b') = 0$, but it does get us

$$f(b') = \frac{f''(b)}{2}\epsilon^2 + \cdots$$

If b was fairly close to the root to begin with, and $f'(b)$ isn't too close to zero, then ϵ is small, which in turn means $f(b')$ is an even smaller value (quadratically smaller in some sense), so b' is very close to the root. I leave further details to you. The details (of how much better your approximate root gets) are actually tidier in the case of Hensel's Lemma.

Proof of Hensel's Lemma the Second. Suppose we have a solution

$$f(x_0) \equiv 0 \pmod{p^a}.$$

In the p-adic metric, this means

$$|f(x_0)|_p < \frac{1}{p^a}.$$

In other words, we have a fairly good approximation to a root.

For $t \in \mathbb{Z}$,

$$f(x_0 + tp^a) = f(x_0) + tp^a f'(x_0) + \cdots + \left(\frac{(tp^a)^n}{n!}\right) f^{(n)}(x_0), \tag{5}$$

where n is the degree of f. Here, tp^a is the ϵ of Newton's method. It remains to choose a good t to make things work.

Proof of first part. A good t is one so that

$$f(x_0 + tp^a) \equiv 0 \pmod{p^{a+1}},$$

or in other words, we have a better approximate root, i.e.

$$|f(x_0 + tp^a)|_p < \frac{1}{p^{a+1}}.$$

But since the higher order terms of the Taylor expansion are divisible by p^{a+1}, this desired congruance has a solution in t if and only if

$$tp^a f'(x_0) \equiv -f(x_0) \pmod{p^{a+1}}$$

has a solution in t, just as in the Newton's Method equation (4), which in turn is true if and only if

$$tf'(x_0) \equiv \frac{-f(x_0)}{p^a} \pmod{p}$$

has a solution in t. (We may rest assured that $p^a \mid f(x_0)$ by assumption.) This has a solution if and only if $p \nmid f'(x_0)$.

Further, if these various equivalent statements hold, then the solution is unique, since $f'(x_0)$ must be invertible modulo p.

Proof of second part. From (5), if $p \mid f'(x_0)$, then

$$f(x_0 + tp^a) \equiv f(x_0) \pmod{p^{a+1}}$$

for all t. Applying this repeatedly, beginning with $a = 1$, we obtain the result.

□

Proof of Hensel's Lemma the First. To get an element $\alpha = (b_n)_{n=1}^{\infty}$ of $\mathbb{Z}_p$ such that $f(\alpha) = 0$, we must construct a sequence of $b_i \in \mathbb{Z}$ such that

$$f(b_i) \equiv 0 \pmod{p^i}, \text{ and}$$
$$b_i \equiv b_{i-1} \pmod{p^{i-1}}$$

To begin, choose $b_1 = b$. By Proposition 24.4, there exists some $b_2 \equiv b_1 \pmod{p}$ such that $f(b_2) \equiv 0 \pmod{p^2}$. Repeating, there exists $b_3 \equiv b_2 \pmod{p^2}$ such that $f(b_3) \equiv 0 \pmod{p^3}$, and so on and so forth until we have all the b_i. □

Example 24.5. *Let* $f(x) = x^3 - 2x^2 + 3x + 9$. *Then* $f'(x) = 3x^2 - 2x + 3$.

Solutions modulo 3. *Here we simply try each solution:* 0 *and* 2 *work.*

Solutions modulo 9**.**

Can we lift $x_0 \equiv 0 \pmod 3$*?*

$$0 \mid f(0) = 9, \quad 3 \mid f'(0) = 3.$$

Therefore $0, 3, 6$ *are all solutions modulo* 9.

Can we lift $x_0 \equiv 3 \pmod 3$*?*

$$9 \nmid f(2) = 15, \quad 3 \nmid f'(2) = 7$$

Therefore, there is a unique lift which works, and it is given by solving

$$7t \equiv \frac{-15}{3} \pmod 3$$

In other words, $t \equiv 1 \pmod 3$ *so* $2 + 1 \cdot 3 \equiv 5 \pmod 9$ *is the unique lift.*

The solutions modulo 9 *are therefore:* 0, 3, 5, 6.

Solutions modulo 27**.**

Can we lift $x_0 \equiv 0 \pmod 9$*?*

$$27 \nmid f(0) = 9, \quad 3 \mid f'(0) = 3.$$

Therefore there are no lifts which work.

Can we lift $x_0 \equiv 3 \pmod 9$*?*

$$27 \mid f(3) = 27, \quad 3 \mid f'(3) = 6.$$

Therefore, there all the lifts work: $3, 12, 21 \pmod 2 7$.

Can we lift $x_0 \equiv 5 \pmod 9$*?*

$$27 \nmid f(5) = 99, \quad 3 \nmid f'(5) = 58.$$

Therefore, there is a unique lift, given by solving

$$58t \equiv \frac{-99}{9} \pmod 3$$

In other words, $t \equiv 1 \pmod 3$ *so* $5 + 1 \cdot 9 \equiv 14 \pmod{27}$ *is the unique lift.*

Can we lift $x_0 \equiv 6 \pmod 9$*?*

$$27 \nmid f(6) = 171, \quad 3 \mid f'(6) = 99.$$

Therefore there are no lifts which work.

The solutions modulo 27 *are therefore:* 3, 12, 14, 21.

25. Local vs. Global

In analogy to Taylor series, we call information relating to a particular prime 'local' information, as opposed to 'global' information. So while $\mathbb{Z}$ or $\mathbb{Q}$ are 'global' objects, $\mathbb{Z}_p$ and $\mathbb{Q}_p$ are 'local' objects because they contain information 'local' to the prime p.

We have now seen the following implications:

$$f(x_1, \ldots, x_n) = 0 \text{ is solvable in } \mathbb{Z}$$
$$\Longrightarrow$$
$$f(x_1, \ldots, x_n) = 0 \text{ is solvable modulo } p^a \text{ for all } a \geq 1$$
$$\Longleftrightarrow$$
$$f(x_1, \ldots, x_n) = 0 \text{ is solvable in } \mathbb{Z}_p$$

This gives us a method to show something is *not* solvable in $\mathbb{Z}$, just by looking modulo prime powers. For example, consider the equation

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

First off, we can assume that x, y and z do not all share a common factor. For, if they did, then we could cancel that factor to reduce to the case that they do not.

Then, we look modulo 4, where the only squares are 0 and 1 modulo 4. That means any solution x, y and z to the original equation modulo 4 would give a solution $X = x^2$, $Y = y^2$ and $Z = z^2$ to

$$X + Y \equiv 3Z \pmod 4$$

where $X, Y, Z \in \{0, 1\}$. This implies $3Z \in \{0, 3\}$ and $X+Y \in \{0, 1, 2\}$. So the only possibility is that $3Z \equiv X + Y \equiv 0 \pmod 4$. But this entails that all of X, Y and Z are even. This implies that all of x, y and z are even, a contradiction to our assumption that they have no common divisor.

We can therefore conclude that $x^2+y^2 = 3z^2$ has no integer solutions!

It is natural to ask whether there is a converse to this sort of argument; that is, provided we can find solutions modulo all integers, or perhaps modulo all prime powers, then could we conclude that there must be a solution in the integers? In general, no, but for certain types of equations, we do obtain the following.

Hasse's Principle, which may or may not be true for a particular equation or family of equations, states:

If $f(x_1, \ldots, x_n) = 0$ has non-trivial solutions in $\mathbb{R}$ and in $\mathbb{Q}_p$ for all p, then it has non-trivial solutions in $\mathbb{Q}$.

By non-trivial we mean to exclude the solution $(0, 0, \ldots, 0)$ to a homogeneous equation.

In this context, $\mathbb{Q}_p$ and $\mathbb{R}$ are called *local fields* and $\mathbb{Q}$ is a *global field*, so that Hasse's principle can be paraphrased as saying that f has solutions globally if and only if it has solutions locally everywhere. (We will not show it in this class, but $\mathbb{Q}_p$ and $\mathbb{R}$ are the only completions of $\mathbb{Q}$ with respect to a metric; this is the sense of 'everywhere' used above).

It is an open question to classify those f for which the Hasse principle may hold. It is known to be true, for example, for homogeneous quadratics with integer squarefree coefficients, e.g. $X^2 + 3XY + 7Y^2$ (this is called the Hasse-Minkowski Theorem). It is known to be false for $3x^3+4y^3+5z^3 = 0$ (Selmer, 1951) and for $x^4 - 17y^4 = 2z^2$ (Landau and Reichardt).

What about equations of one variable? Any equation $f(x) = 0$ of one variable can be 'homogenized' to a homogeneous equation of two variables. By this I mean that a polynomial

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

becomes

$$a_n x^n + a_{n-1}x^{n-1}y + \cdots + a_n x^2 y^{n-2} + a_1 x y^{n-1} + a_0 y^n$$

But notice that the first has solutions in $\mathbb{Q}$, $\mathbb{Q}_p$ or $\mathbb{R}$ if and only if the second one does (as long as we exclude the solution $(x, y) = (0, 0)$). For, if we have a solution x to the first, then $(x, 1)$ is a solution to the second. Conversely, if $(x, y) \neq (0, 0)$ is a solution to the second, then x/y is a solution to the first (this is safe because if $y = 0$, then $x = 0$). As a result of this comment, we see that quadratic equations in one variable satisfy the Hasse principle, by the Hasse-Minkowski Theorem. (We haven't proven this, so you can't use it on your homework!)

26. Motivation to study quadratic forms, unique factorisation in number rings, and quadratic residues

Lemma 26.1 (Fermat's Lemma)**.** *If $p \equiv 1 \pmod 4$, then $x^2 \equiv -1 \pmod p$ is solvable.*

We will not provide a proof right now, but we'll provide a corollary, a theorem we have been anticipating:

Theorem 26.2. *An odd prime p is a sum of 2 squares if and only if $p \equiv 1 \pmod 4$.*

Proof. On homework, we showed that if $p \equiv 3 \pmod 4$, then p is not the sum of two squares. (This was just a brief argument modulo 4.)

So suppose that $p \equiv 1 \pmod 4$. By Fermat's Lemma, $p \mid x^2 + 1$ for some $x \in \mathbb{Z}$. So $p \mid (x+i)(x-i)$ in the Gaussian integers, $\mathbb{Z}[i]$. If p were prime in $\mathbb{Z}[i]$, then $p \mid x+i$ or else $p \mid x-i$. But suppose one of these were true. Then,

$$\frac{x \pm i}{p} = \frac{x}{p} \pm \frac{1}{p} i$$

is not a Gaussian integer, since $1/p \notin \mathbb{Z}$. So it must be that p is not prime. So p can be factored in the Gaussian integers, as

$$p = (a+bi)(c+di)$$

where the two factors are non-units. Since the norm $N(p) = p^2$, and $N(a+bi), N(c+di) > 1$ as they are non-units, the only possibility is that

$$N(a+bi) = N(c+di) = p.$$

So $a^2 + b^2 = p = c^2 + d^2$. □

In fact, we've learned more from this proof.

Theorem 26.3. *Let p be an odd prime. Then the following are equivalent:*

(1) $p \equiv 1 \pmod 4$,
(2) $x^2 \equiv -1 \pmod p$ *is solvable*
(3) the quadratic form $x^2 + y^2$ (Gaussian norm) takes value p for some x and y,
(4) p has a nontrivial factorisation in the Gaussian integers.

So in particular, an odd rational prime p is prime in the Gaussian integers if and only if $p \equiv 3 \pmod 4$.

This relationship generalises to other number rings, and provides motivation for studying *quadratic residues* (numbers which are the square of something modulo p), *quadratic forms* (homogeneous degree two forms), and *unique factorisation*. These are three of our main goals in the course from now on.

27. Studying $(\mathbb{Z}/p\mathbb{Z})^*$

By the Chinese Remainder Theorem, to study $(\mathbb{Z}/n\mathbb{Z})^*$, it suffices to study $(\mathbb{Z}/p^a\mathbb{Z})^*$ for prime powers p^a. We begin by studying $(\mathbb{Z}/p\mathbb{Z})^*$. Since all non-zero elements are invertible modulo p, $\mathbb{Z}/p\mathbb{Z}$ is a finite field. Thus we begin with a few lemmas about fields.

Lemma 27.1. *Let $f(x) \in k[x]$, where k is a field, and suppose that $f(x)$ is not identically zero. Let $n = \deg f(x)$. Then f has at most n distinct roots in k.*

Proof. The proof proceeds by induction on n. The case of $n = 1$ is immediate. Suppose that this has been proven for $n-1$, and consider f of degree n. If $f(x)$ has no roots, then we are done. If $f(x)$ has a root α, then by the division algorithm in $k[x]$ (which is a Euclidean domain),

$$f(x) = q(x)(x - \alpha) + r$$

for a constant r and some $q(x)$ of degree $n-1$; however, since α is a root, we find $r = 0$.

If $\beta \neq \alpha$ is another root, then

$$0 = f(\beta) = q(\beta)(\beta - \alpha)$$

which implies that $q(\beta) = 0$ (since k has no zero-divisors). Therefore β is one of the at most $n-1$ roots of q. Hence $f(x)$ has at most n roots (α together with the roots of $q(x)$). □

Corollary 27.2. *Let $f(x), g(x) \in k[x]$, where k is a field. Suppose that $n = \deg f(x) = \deg g(x)$. If $f(\alpha_i) = g(\alpha_i)$ for $n+1$ distinct values*

$$\alpha_1, \alpha_2, \ldots, \alpha_{n+1},$$

then $f(x) = g(x)$.

Proof. Apply the Lemma to $f(x) - g(x)$. Then it has $n+1$ distinct roots, hence is identically zero. □

Proposition 27.3.

$$x^{p-1} - 1 \equiv (x-1)(x-2)\cdots(x-(p-1)) \pmod{p}$$

Proof. Let $f(x) = x^{p-1} - 1 - (x-1)(x-2)\cdots(x-(p-1))$. Then $\deg f(x) < p-1$ since the leading terms cancel. But it has $p-1$ distinct roots in $\mathbb{Z}/p\mathbb{Z}$ (all the nonzero elements). So $f(x)$ must be identically zero, by Lemma 27.1. □

Corollary 27.4.

$$(p-1)! \equiv -1 \pmod{p}.$$

Proof. Set $x = 0$ in the previous proposition. □

Theorem 27.5 (Wilson's Theorem)**.**

$$(n-1)! \equiv -1 \pmod{n}$$

if and only if n is prime or 1.

Proof. We have just seen that the equation holds for n prime. For $n = 1$, it holds also: $0! \equiv 1 \equiv -1 \pmod 1$.

If $n = 4$, then $3! \equiv 6 \not\equiv -1 \pmod 4$, so the equation does not hold.

If $n > 4$ is composite, then there exist some a, b such that $n = ab$, with $1 < a, b < n$.

If $a \neq b$, then $n = ab \mid (n-1)!$ so $(n-1)! \equiv 0 \pmod n$ and the equation does not hold.

If $a = b$, then a and $2a$ appear in $(n-1)!$, so $2n = 2a^2 \mid (n-1)!$. Therefore $(n-1)! \equiv 0 \pmod n$ and the equation does not hold. □

Proposition 27.6. *If $d \mid p-1$, then $x^d \equiv 1 \pmod p$ has exactly d solutions.*

Proof. Let $dd' = p - 1$. Then define

$$g(x) = \frac{x^{p-1} - 1}{x^d - 1} = \frac{(x^d)^{d'} - 1}{x^d - 1} = (x^d)^{d'-1} + (x^d)^{d'-2} + \cdots + x^d + 1.$$

Then

$$x^{p-1} - 1 = (x^d - 1)g(x).$$

If $x^d - 1$ had fewer than d roots, then by Lemma 27.1, $x^{p-1} - 1$ would have fewer than $p - 1$ roots. But it has roots $1, 2, \ldots, p-1 \in \mathbb{Z}/p\mathbb{Z}$. By this contradiction, we have proven the proposition. □

Note: this may fail when $d \nmid p-1$. For example, suppose $p = 5$, and suppose we are interested in the roots of $x^3 - 1$ modulo p. We can fill out a chart:

x	x^2	x^3
0	0	0
1	1	1
2	4	3
3	4	2
4	1	4

From this we see that there is only one root of the polynomial. In particular,

$$\frac{x^3 - 1}{x - 1} = x^2 + x + 1$$

has no roots in $\mathbb{Z}/5\mathbb{Z}$. This implies that $\mathbb{Z}/5\mathbb{Z}$ is not algebraically closed (meaning there are polynomials over $\mathbb{Z}/5\mathbb{Z}$ with no roots in $\mathbb{Z}/5\mathbb{Z}$). (Recall that $\mathbb{C}$ is algebraically closed but $\mathbb{R}$ is not, because $x^2 + 1$ has no solution.)

Here's a little exercise: show that $x^d \equiv 1 \pmod p$ can never have exactly $d - 1$ roots.

Theorem 27.7. *$(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic.*

Proof. In any finite commutative group G, there exists $y \in G$ whose order is the least common multiple of the orders of all the elements of G. In particular, if n is the order of y, then $x^n = 1$ for all $x \in G$.

So all elements of $(\mathbb{Z}/p\mathbb{Z})^*$ are roots of $x^n - 1$, but 0 is not, so it has exactly $p-1$ roots in $\mathbb{Z}/p\mathbb{Z}$.

But $x^n - 1$ has at most n roots in $\mathbb{Z}/p\mathbb{Z}$ (by Lemma 27.1).

So

$$p - 1 \leq n.$$

On the other hand, $1, y, y^2, \ldots, y^{n-1}$ are all distinct, by the fact that y has order n. Therefore, $x^n - 1$ has at least n roots in $\mathbb{Z}/p\mathbb{Z}$, so

$$p - 1 \geq n.$$

Hence $n = p - 1$ and

$$(\mathbb{Z}/p\mathbb{Z})^* = \{1, y, y^2, \ldots, y^{n-1}\} = \langle y \rangle.$$

□

28. Primitive Roots

Definition 28.1. *$a \in \mathbb{Z}$ is called a* primitive root modulo n *if a generates the group $(\mathbb{Z}/n\mathbb{Z})^*$.*

Equivalently, $\gcd(a, n) = 1$ and a has multiplicative order $\phi(n)$.

Example 28.2. *Let's consider $n = 5$. We can form a chart*

x	x^2	x^3	x^4
1	1	1	1
2	4	3	1
3	4	2	1
4	1	4	1

Therefore the primitive roots are 2 *and* 3.

If $(\mathbb{Z}/n\mathbb{Z})^*$ is not cyclic, then there do not exist elements of order $\phi(n) = |(\mathbb{Z}/n\mathbb{Z})^*|$, so there are no primitive roots.

Example 28.3. *Let consider $n = 8$. Here $\phi(n) = 4$. The chart of powers of elements of $(\mathbb{Z}/n\mathbb{Z})^*$ is:*

x	x^2	x^3	x^4
1	1	1	1
3	1	3	1
5	1	5	1
7	1	7	1

The orders of $3, 5$ *and* 7 *are* 2*. The order of* 1 *is, of course,* 1*. Nothing has order* 4*, so there are no primitive roots. Equivalently,* $(\mathbb{Z}/n\mathbb{Z})^*$ *is not cyclic.*

Our goal in studying primitive roots is to determine which n have primitive roots. There are plenty of other interesting questions we won't address, however.

For example, one could ask for which p (or n) is a a primitive root? We don't know much about the answer to this question. Artin has conjectured that a is a primitive root modulo infinitely many primes p, and in fact, that it is a primitive root for the following percentage of primes:

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p(p-1)}\right) = 37.3955\cdots\%$$

In the 80s, Gupta, Murty and Heath-Brown were able to show that there exist at most two a which are primitive roots for only finitely many primes. Of course, without specifying which two (presumably there are in fact none, not two).

We could also ask what is the smallest primitive root modulo any particular prime p? And we should ask how to find primitive roots effectively. These are both important problems.

29. Studying $(\mathbb{Z}/n\mathbb{Z})^*$

Our goal in this section is to classify those n for which $(\mathbb{Z}/n\mathbb{Z})^*$ is cyclic. By the Chinese Remainder Theorem, it suffices to ask the question for prime powers. Our main results are:

Theorem 29.1. $(\mathbb{Z}/p^a\mathbb{Z})^*$ *is cyclic if and only if* p *is odd or* $p = 2$ *and* $a = 1$ *or* 2*.*

Corollary 29.2. *A primitive root exists modulo* n *if and only if one of the following holds:*

(1) $n = 1, 2, 4$*,*
(2) $n = p^a$ *for* p *an odd prime and* $a \geq 1$*,*
(3) $n = 2p^a$ *for* p *an odd prime and* $a \geq 1$*.*

We'll show how to derive the Corollary, then turn to proving the Theorem.

Proof of Corollary. For the first two items, Theorem 29.1 says there is a primitive root, so we're done. For the last item,

$$(\mathbb{Z}/2p^a\mathbb{Z})^* \cong (\mathbb{Z}/2\mathbb{Z})^* \times (\mathbb{Z}/p^a\mathbb{Z})^* \cong (\mathbb{Z}/p^a\mathbb{Z})^*$$

because $(\mathbb{Z}/2\mathbb{Z})^*$ is the trivial group. So we have primitive roots in this case.

For the converse, suppose that none of the items hold. Then we may divide into three cases:

$8 \mid n$**.** Write $n = 2^a m$, where m is odd and $a \geq 3$. Then

$$(\mathbb{Z}/n\mathbb{Z})^* \cong (\mathbb{Z}/2^a\mathbb{Z})^* \times (\mathbb{Z}/m\mathbb{Z})^*$$

However, the first factor is a non-cyclic subgroup. If $(\mathbb{Z}/n\mathbb{Z})^*$ were cyclic, it would only have cyclic subgroups. So $(\mathbb{Z}/n\mathbb{Z})^*$ is not cyclic.

$pq \mid n$**,** $p \neq q$ **primes.** Write $n = p^a q^b m$, where m is coprime to p and q, and $a, b \geq 1$. Then

$$(\mathbb{Z}/n\mathbb{Z})^* \cong (\mathbb{Z}/p^a\mathbb{Z})^* \times (\mathbb{Z}/q^b\mathbb{Z})^* \times (\mathbb{Z}/m\mathbb{Z})^*$$

However, the first two factors have even order, i.e. non-coprime orders. By group theory, their product is not cyclic, so $(\mathbb{Z}/n\mathbb{Z})^*$ is not cyclic.

$n = 4p$ **for** p **an odd prime.**

Then

$$(\mathbb{Z}/n\mathbb{Z})^* \cong (\mathbb{Z}/4\mathbb{Z})^* \times (\mathbb{Z}/p\mathbb{Z})^*$$

However, both factors have even order, so as in the last case $(\mathbb{Z}/n\mathbb{Z})^*$ is not cyclic.

□

Lemma 29.3. *If* $x \equiv y \pmod{p^a}$*, then* $x^p \equiv y^p \pmod{p^{a+1}}$.

Proof. Suppose that

$$x \equiv y \pmod{p^a}.$$

Then

$$x = y + cp^a$$

and therefore

$$x^p = y^p + py^{p-1}cp^a + \binom{p}{2} y^{p-2} c^2 p^{2a} + \cdots + c^p p^{p^a}.$$

But every term besides the first is divisible by p^{a+1}. Therefore,

$$x^p \equiv y^p \pmod{p^{a+1}}.$$

□

Lemma 29.4. *Let* $a \geq 2$*, and let* $p \neq 2$*. Then*

$$(1+p)^{p^{a-2}} \equiv 1 + p^{a-1} \pmod{p^a}$$

Proof. Prove this by induction on a. For $a = 2$, this is trivial. Suppose it has been proven for a. Then

$$(1+p)^{p^{a-2}} \equiv 1 + p^{a-1} \pmod{p^a}$$

Applying Lemma 29.3, we obtain

$$(1+p)^{p^{a-1}} \equiv (1+p^{a-1})^p \pmod{p^{a+1}}$$

Now we expand the right hand side with Binomial Theorem and argue that all the terms are divisble by p^{a+1} except the first two. That is,

$$(1+p^{a-1})^p = 1 + p \cdot p^{a-1} + \binom{p}{2} p^{2(a-1)} + \cdots + p^{p(a-1)}$$

If $k \geq 3$ or $a \geq 3$, we have $k(a-1) \geq a+1$, so that

$$p^{a+1} \mid \binom{p}{k} p^{k(a-1)}.$$

Otherwise, $a = k = 2$, and the argument above doesn't work; but here $p \mid \binom{p}{2}$ since $p > 2$, so p^3 divides this term and we have

$$(1+p^{a-1})^p \equiv 1 + p^a \pmod{p^{a+1}}$$

and we are done. □

Now we will prove Theorem 29.1. For an odd prime, the outline is as follows. First, find an element of order $p-1$, by lifting a primitive root mod p using Hensel's Lemma. Then, find an element of order p^{a-1}. These can then be combined to form an element of order $p^{a-1}(p-1)$.

Proof of Theorem 29.1. Let p be an odd prime. Define

$$f(x) = x^{p-1} - 1 \in \mathbb{Z}[x].$$

Then we have

$$f'(x) = (p-1)x^{p-2} \in \mathbb{Z}[x].$$

Take g to be a primitive root modulo p. Then $f(g) \equiv 0 \bmod p$ and $f'(g) \not\equiv 0 \bmod p$, so by Hensel's Lemma, there exists a g_1 such that

$$f(g_1) \equiv 0 \pmod{p^a}$$
$$g_1 \equiv g \pmod{p}.$$

So g_1 has order dividing $p-1$ modulo p^a. But if g_1 had a smaller order $n < p-1$, then

$$g_1^n \equiv 1 \pmod{p^a} \implies g^n \equiv g_1^n \equiv 1 \pmod{p}$$

which contradicts the choice of g as a primitive root modulo p. So g_1 has order $p-1$ modulo p^a.

Suppose that we consider $b \equiv 1 \pmod{p}$. Then $b^{p^{a-1}} \equiv 1 \pmod{p}$. I claim that $b^{p^{a-1}} \equiv 1 \pmod{p^a}$. This is because $b^{p^{a-1}}$ is a solution modulo p^a to the equation $x^{p-1} \equiv 1 \pmod{p^a}$ (this uses the fact that $\phi(p^a) = p^{a-1}(p-1)$). However, by Hensel's Lemma, there is a unique lift of 1 modulo p^a which satisfies $x^{p-1} \equiv 1 \pmod{p^a}$. Since 1 is such a lift, it must be that $b^{p^{a-1}} \equiv 1 \pmod{p^a}$.

Now we must show that we can choose such a b having order equal to p^{a-1}. Let $b = 1 + p$. Then this is a consequence of Lemma 29.4, which tells us that

$$b^{p^{a-2}} \not\equiv 1 \pmod{p^a}.$$

Since p^{a-1} and $p-1$ are coprime, group theory tells us that the order of $g_1 b_1$ is $p^{a-1}(p-1) = \phi(p^a)$, and therefore $g_1 b_1$ is a primitive root modulo p^a.

We are left with the case of $p = 2$. One can check directly that $(\mathbb{Z}/2\mathbb{Z})^*$ and $(\mathbb{Z}/4\mathbb{Z})^*$ are cyclic. Now, in general for $a \geq 3$, we claim that

$$(\mathbb{Z}/2^a\mathbb{Z})^* \cong \langle -1 \rangle \times \langle 5 \rangle.$$

where the first factor has order 2, and the second factor has order 2^{a-2}. Once we have established this, it is immediate that $(\mathbb{Z}/2^a\mathbb{Z})^*$ is not cyclic.

In order to prove this, we show by induction that

$$5^{2^{k-3}} \equiv 1 + 2^{k-1} \pmod{2^k}$$

for all $k \geq 3$. For $k = 3$, one verifies this easily. Now suppose it is known for k. Then by Lemma 29.3,

$$\begin{aligned} 5^{2^{k-2}} &\equiv (1 + 2^{k-1})^2 \pmod{2^{k+1}} \\ &\equiv 1 + 2 \cdot 2^{k-1} + 2^{2k-2} \pmod{2^{k+1}} \\ &\equiv 1 + 2^k \pmod{2^{k+1}} \end{aligned}$$

which completes the induction. This tells us both that

$$5^{2^{k-3}} \not\equiv 1 \pmod{2^k}$$

and also that

$$5^{2^{k-2}} \equiv 1 \pmod{2^k}$$

In other words, 5 has order 2^{k-2} modulo 2^k. Finally, we must verify that $\langle -1 \rangle$ and $\langle 5 \rangle$ are disjoint subgroups of $(\mathbb{Z}/2^a\mathbb{Z})^*$). This is a consequence of the fact that

$$5^\ell + 1 \equiv 2 \pmod{4}$$

for all ℓ, so that $5^\ell \not\equiv -1 \pmod{4}$, implying

$$5^\ell \not\equiv -1 \pmod{2^k}.$$

Since the two subgroups are disjoint, we have proven the group isomorphism we needed. □

30. nTH POWER RESIDUES

Definition 30.1. *If $m, n \in \mathbb{Z}^+$ and $\gcd(a, m) = 1$, then a is an n-*th power residue *if $x^n \equiv a \pmod{m}$ is solvable.*

Proposition 30.2. *If $(\mathbb{Z}/m\mathbb{Z})^*$ is cyclic, and $\gcd(a, m) = 1$, then a is an n-th power residue modulo m if and only if*

$$a^{\phi(m)/d} \equiv 1 \pmod{m}$$

where $d = \gcd(n, \phi(m))$.

Proof. Let g be a primitive root modulo m. Let $a = g^b$ and $x = g^y$. Then

$$x^n \equiv a \pmod{m} \iff g^{ny} \equiv g^b \pmod{m} \iff ny \equiv b \pmod{\phi(m)}$$

This is solvable if and only if $d \mid b$. And if there exist solutions, then there are d solutions. If $d \mid b$, then

$$a^{\phi(m)/d} \equiv g^{b\phi(m)/d} \equiv 1 \pmod{m}.$$

Conversely,

$$a^{\phi(m)/d} \equiv 1 \pmod{m} \implies g^{b\phi(m)/d} \equiv 1 \pmod{m} \implies \phi(m) \mid \frac{b\phi(m)}{d} \implies d \mid b.$$

□

In fact, we've also learned the following from this proof:

Proposition 30.3. *The congruence*

$$x^n \equiv a \pmod{m}$$

has no solutions or else exactly $\gcd(n, \phi(m))$ solutions.

If $m = 2^e p_1^{e_1} \cdots p_\ell^{e_\ell}$, then $x^n \equiv a \pmod{m}$ is solvable if and only if

$$\begin{aligned} x^n &\equiv a \pmod{2^e} \\ x^n &\equiv a \pmod{p_1^{e_1}} \\ &\vdots \\ x^n &\equiv a \pmod{p_\ell^{e_\ell}} \end{aligned}$$

are all solvable.

So we are interested in prime powers.

Proposition 30.4. *Let p be an odd prime. Let $p \nmid a$ and $p \nmid n$. If*

$$x^n \equiv a \pmod{p}$$

is solvable, then so is

$$x^n \equiv a \pmod{p^e}$$

for all $e \geq 1$.

Furthermore, all these congruences have the same number of solutions.

Proof. This is an application of Hensel's Lemma. If

$$f(x) = x^n - a$$

then $f'(x) = nx^{n-1}$. Each root α of $f(x)$ is nonzero, so $f'(\alpha) \not\equiv 0 \pmod{p}$. So there's a unique lift to α' with

$$f(\alpha') \equiv 0 \pmod{p^e}.$$

□

Proposition 30.5. *Let $2^k \mid\mid n$ (i.e. the exact power of 2 dividing n is 2^k), and let a be odd. Suppose that*

$$x^n \equiv a \pmod{2^{2k+1}}$$

is solvable. Then

$$x^n \equiv a \pmod{2^e}$$

is solvable for all $e \geq 1$. All these congruences have the same number of solutions.

Proof. Exercise. □

Example 30.6. $x^2 \equiv 5 \pmod{4}$ *is solvable ($x = 1$), but* $x^2 \equiv 5 \pmod{8}$ *is not solvable.*

Proposition 30.7. *Let a be odd, and let $e \geq 3$. Then*

1. *If n is odd, then $x^n \equiv a \pmod{2^e}$ has a unique solution.*
2. *If n is even, then $x^n \equiv a \pmod{2^e}$ has a solution if and only if the following things hold:*
   - *(a) $a \equiv 1 \pmod{4}$*
   - *(b) $a^{2^{e-2}/d} \equiv 1 \pmod{2^e}$ where $d = \gcd(n, 2^{e-2})$*

   And if there are solutions, then there are $2d$ of them.

Proof. Exercise. □

31. Quadratic residues

Definition 31.1. *The Legendre symbol is defined as follows. Let $a \in \mathbb{Z}$ and let p be an odd prime. Define*

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & a \equiv 0 \pmod{p} \\ 1 & a \text{ is a QR, i.e. } x^2 \equiv a \text{ is solvable} \\ -1 & a \text{ is a QNR, i.e. } x^2 \equiv a \text{ is not solvable} \end{cases}$$

Example 31.2.

$$\left(\frac{17}{5}\right) = \left(\frac{2}{5}\right) = -1$$

since the squares modulo 5 are 1 and 4.

$$\left(\frac{101}{97}\right) = \left(\frac{4}{97}\right) = 1$$

$$\left(\frac{-42}{61}\right) = 1$$

This last is because $10^{22} \equiv 19 \pmod{61}$, so $19 \equiv (10^{11})^2$, i.e. $-42 \equiv 18^2 \pmod{61}$.

Here are the most important first properties of the Legendre symbol.

Proposition 31.3. *Let p be an odd prime. Then*

(1) $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}$

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

(3) $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$

Proof. The second and third parts follow immediately from the first part.

Let g be a primitive root for $(\mathbb{Z}/p\mathbb{Z})^*$. The map $x \mapsto x^2$ is $2-to-1$ on $(\mathbb{Z}/p\mathbb{Z})^*$, since there is no residue equivalent to its negative, hence all square roots come in pairs if they come at all. In particular, there are $\frac{p-1}{2}$ quadratic residues and $\frac{p-1}{2}$ quadratic non-residues. The $\frac{p-1}{2}$ even powers of g are all quadratic residues, so the odd powers are non-residues.

We can write

$$a \equiv g^i \pmod{p}$$

for some i. Then we have two cases.

i **is even.** In this case

$$a \equiv (g^{i/2})^2 \pmod{p}$$

so a is a quadratic residue. Furthermore,

$$a^{\frac{p-1}{2}} \equiv g^{\frac{i(p-1)}{2}} \equiv 1 \pmod{p}$$

since $p-1 \mid \frac{i(p-1)}{2}$. So we have verified the statement.

i **is odd.** In this case $a \equiv g^i \pmod{p}$ is a non-residue. Furthermore,

$$a^{\frac{p-1}{2}} \equiv g^{\frac{i(p-1)}{2}} \not\equiv 1 \pmod{p}$$

since $p-1 \nmid \frac{i(p-1)}{2}$. But

$$(a^{\frac{p-1}{2}})^2 \equiv a^{p-1} \equiv 1 \pmod{p}$$

so

$$a^{\frac{p-1}{2}} \equiv -1 \pmod{p}.$$

So we have verified the statement.

□

Since $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$, we obtain the following corollary.

Corollary 31.4. *$x^2 \equiv -1 \pmod{p}$ is solvable if and only if $p \equiv 1 \pmod{4}$*

I.e., we have proven Fermat's Lemma!

We also get a nice result on the infinitude of primes congruent to 1 modulo 4. If you recall, the method for primes congruent to 3 modulo 4 used on homework didn't extend to this case.

Corollary 31.5. *There are infinitely many primes congruent to 1 modulo 4.*

Proof. Suppose that $2, p_1, \ldots, p_m$ are a finite set of primes. Let

$$N = 4p_1^2 p_2^2 \cdots p_m^2 + 1.$$

Then $N > 1$ must have some prime factors. Suppose that one of them is $p \mid N$. Then $\left(\frac{-1}{p}\right) = 1$ by the form of N. So $p \equiv 1 \pmod{4}$. But p cannot be among the finite list we began with, as $p \equiv 1 \pmod{p_i}$ for those p_i. Thus we have found, given any finite list of primes, a new prime congruent to 1 modulo 4. Hence there are infinitely many primes congruent to 1 modulo 4. □

32. Quadratic Reciprocity

You looked a little at data about quadratic residues on your first homework. You may have noticed that if you graph $\left(\frac{p}{q}\right)$ there is some kind of symmetry around the diagonal. By staring at tables of such things, Gauss came up with a conjecture. Here's the full statement.

Theorem 32.1. *Let p and q be odd primes. Then*

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

(2) $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$

(3) $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}$

In particular, if p and q are both 1 mod 4, then $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$. This is *really weird*; why should the existence of a square with residue p modulo q tell you anything about the existence of a square with residue q modulo p? It's just darn weird.

Example 32.2. *Let's demonstrate the use of Quadratic Reciprocity as motivation for proving it. It makes determining whether something is a quadratic residue very fast. By the multiplicativity of the Legendre symbol,*

$$\left(\frac{-42}{61}\right) = \left(\frac{-1}{61}\right)\left(\frac{2}{61}\right)\left(\frac{3}{61}\right)\left(\frac{7}{61}\right).$$

We can evaluate each of these symbols in turn:

$$\left(\frac{-1}{61}\right) = (-1)^{60/2} = 1$$

$$\left(\frac{2}{61}\right) = (-1)^{\frac{61^2-1}{8}} = 1$$

$$\left(\frac{3}{61}\right) = \left(\frac{61}{3}\right)(-1)^{\frac{2}{2}\cdot\frac{60}{2}} = \left(\frac{61}{3}\right) = \left(\frac{1}{3}\right) = 1$$

$$\left(\frac{7}{61}\right) = \left(\frac{61}{7}\right)(-1)^{\frac{6}{2}\cdot\frac{60}{2}} = \left(\frac{5}{7}\right) = \left(\frac{7}{5}\right)(-1)^{\frac{4}{2}\cdot\frac{6}{2}} = \left(\frac{7}{5}\right) = \left(\frac{2}{5}\right) = (-1)^{\frac{24}{8}} = -1$$

Therefore, taking the product,

$$\left(\frac{-42}{61}\right) = 1\cdot(-1)\cdot 1\cdot(-1) = 1$$

Alternatively, we could notice that

$$\left(\frac{-42}{61}\right) = \left(\frac{19}{61}\right) = \left(\frac{61}{19}\right)(-1)^{\frac{60}{2}\cdot\frac{18}{2}} = \left(\frac{4}{19}\right) = 1$$

which is faster because 19 *is a prime. Later, we will introduce the Jacobi symbol, which will remove the need to factor into primes in order to 'reverse' the symbol.*

Lemma 32.3 (Eisenstein's Lemma)**.** *Let p be an odd prime, and $a \in \mathbb{Z}$ coprime to p. Then*

$$\left(\frac{a}{p}\right) = (-1)^{\sum_{\substack{2 \leq n \leq p-1 \\ even}} \left\lfloor \frac{an}{p} \right\rfloor}.$$

The notation $\lfloor x \rfloor$ denotes the floor function: its value is the greatest integer less than or equal to x.

Proof. For even n in the interval $2 \leq n \leq p-1$, let $r(n)$ be the least positive residue of an modulo p (i.e. the smallest residue which is positive). Consider the following list of numbers:

$$(-1)^{r(2)} r(2), \quad (-1)^{r(4)} r(4), \quad \ldots, \quad (-1)^{r(p-1)} r(p-1).$$

First, these are all distinct modulo p. For, if

$$(-1)^{r(n_1)} r(n_1) \equiv (-1)^{r(n_2)} r(n_2) \pmod{p}$$

then

$$(-1)^{r(n_1)-r(n_2)} a n_1 \equiv a n_2 \pmod{p}$$

which implies that $n_1 \equiv \pm n_2 \pmod{p}$. But the n_i are all even, so this can only happen if $n_1 \equiv n_2 \pmod{p}$, which goes against our original assumptions.

Second, the list contains exactly $\frac{p-1}{2}$ numbers.

Third, the least positive residues of all these numbers are even. In fact,

$$(-1)^{r(n)} r(n) = \begin{cases} r(n) & \text{if } r(n) \text{ is even} \\ p - r(n) & \text{if } r(n) \text{ is odd} \end{cases}$$

Therefore, the list is a rearrangement of the list

$$2, 4, 6, \ldots, \frac{p-1}{2}.$$

Thus, taking the product of the two lists, we have

$$(-1)^{r(2)+r(4)+\cdots+r(p-1)} 2a \cdot 4a \cdot 6a \cdots (p-1)a \equiv 2 \cdot 4 \cdot 6 \cdots (p-1) \pmod{p}.$$

Since the product $2 \cdot 4 \cdot 6 \cdots (p-1)$ is coprime to p, we conclude that

$$(-1)^{r(2)+r(4)+\cdots+r(p-1)} \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

On the other hand,

$$\frac{an}{p} = \left\lfloor \frac{an}{p} \right\rfloor + \frac{r(n)}{p}.$$

So, since p is odd and n is even,

$$\left\lfloor \frac{an}{p} \right\rfloor = r(n) \pmod 2$$

This implies that

$$a^{\frac{p-1}{2}} \equiv (-1)^{\sum\limits_{\substack{2\le n\le p-1\\ \text{even}}} \left\lfloor \frac{an}{p} \right\rfloor}.$$

Finally, by Proposition 31.3,

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

Putting these two equations together proves the statement of the Lemma. □

Proposition 32.4.

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$$

Proof. We use Eisenstein's Lemma. In particular, define μ to be

$$\mu := \sum_{\substack{2\le n\le p-1\\ \text{even}}} \left\lfloor \frac{2n}{p} \right\rfloor = \left\lfloor \frac{p}{4} \right\rfloor - \left\lfloor \frac{p}{2} \right\rfloor = \#\{x \in \mathbb{Z} : p/4 < x < p/2\}.$$

Suppose that $p = 8k + r$. Then $r = 1, 3, 5$ or 7. The interval of interest becomes

$$2k + \frac{r}{4} < x < 4k + \frac{r}{2}.$$

We only care about the parity of the number of integers x in this interval, which is the same as the parity of the number of integers x in the interval

$$\frac{r}{4} < x < \frac{r}{2}.$$

For $r = 1$, there are no solutions, hence μ is even, but so is $\frac{p^2-1}{8}$. For $r = 3$ or $r = 5$, there is one solution, so μ is odd, but so is $\frac{p^2-1}{8}$. For $r = 7$, there are two solutions, so μ is even, but so is $\frac{p^2-1}{8}$. By Eisenstein's Lemma, we have proven the required statement. □

33. A proof of Quadratic Reciprocity

The proof consists of two parts: the first is Eisenstein's Lemma, which is similar to Gauss' Lemma. The second is the main proof, which relies on Eisenstein's Lemma and consists of a simple geometric argument.

Proof of Quadratic Reciprocity. I need to draw some pictures **********
□

34. The Jacobi Symbol

Definition 34.1. *Let b be a positive odd integer. Write $b = q_1^{e_1} \cdots q_s^{e_s}$ for its prime factorisation. Then for any a, we define the* Jacobi Symbol

$$\left(\frac{a}{b}\right) = \prod_{j=1}^{s} \left(\frac{a}{q_j}\right)^{e_j}.$$

We will use the same notation as the Legendre symbol because these two definitions agree whenever they are both defined.

Note that if $\gcd(a,b) > 1$, then $\left(\frac{a}{b}\right) = 0$. And if $\gcd(a,b) = 1$, then $\left(\frac{a}{b}\right) \in \{1, -1\}$.

Warning! It is very important to note that if b is not prime, then in general

$$\left(\frac{a}{b}\right) = 1 \implies a \text{ is a QR modulo } b.$$

For example,

$$\left(\frac{2}{15}\right) = \left(\frac{2}{5}\right)\left(\frac{2}{3}\right) = -1 \cdot -1 = 1$$

but $x^2 \equiv 2 \pmod{15}$ has no solutions (otherwise it would have solutions modulo 5 and 3, also).

However, it is still the case that

$$a \text{ a QR modulo } b \implies \left(\frac{a}{b}\right) = 1,$$

since whenever a is a QR modulo b, a must also be a QR for each prime factor of b.

Proposition 34.2. *Let b be an odd positive integer. Then a is a QR modulo b if and only if a is a QR modulo every prime $p \mid b$.*

Proof. Write $b = q_1^{e_1} \cdots q_s^{e_s}$ as the prime factorisation of b. By the Chinese Remainder Theorem, a is a QR modulo b if and only if a is a QR modulo every prime power $q_i^{e_i}$. By Hensel's Lemma (via Proposition 30.4), for q_i odd, a is a QR modulo $q_i^{e_i}$ if and only if a is a QR modulo q_i. □

The following is a catalogue of the properties of the Jacobi Symbol.

Proposition 34.3. *Let b, b_1, b_2 be positive odd integers. Then*

(1) $\left(\frac{a_1 a_2}{b}\right) = \left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right)$.
(2) $\left(\frac{a}{b}\right) = 0$ *if* $\gcd(a,b) > 1$.

(3) $\left(\frac{a}{b_1 b_2}\right) = \left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right)$.
(4) $a_1 \equiv a_2 \pmod{b}$ *implies* $\left(\frac{a_1}{b}\right) = \left(\frac{a_2}{b}\right)$.
(5) $\left(\frac{-1}{b}\right) = (-1)^{\frac{b-1}{2}}$.
(6) $\left(\frac{2}{b}\right) = (-1)^{\frac{b^2-1}{8}}$.
(7) If $\gcd(b_1, b_2) = 1$, *then*

$$\left(\frac{b_1}{b_2}\right)\left(\frac{b_2}{b_1}\right) = (-1)^{\frac{b_1-1}{2}\cdot\frac{b_2-1}{2}}.$$

Proof. The first four items are consequences of the definition of the Jacobi symbol and the corresponding properties for the Legendre symbol, so they are left as an exercise for the reader.

Write $b = q_1 \cdots q_s$ for the prime factorisation of b, where repeats are allowed, i.e. the q_i may not be distinct.

Item (5) If x and y are odd, then

$$\frac{xy-1}{2} - \left(\frac{x-1}{2} + \frac{y-1}{2}\right) = \frac{(x-1)(y-1)}{2} \equiv 0 \pmod{2}.$$

Hence

$$\frac{x-1}{2} + \frac{y-1}{2} \equiv \frac{xy-1}{2} \pmod{2}$$

Therefore,

$$\sum_{i=1}^{s} \frac{q_i - 1}{2} \equiv \frac{b-1}{2} \pmod{2}.$$

But therefore,

$$\left(\frac{-1}{b}\right) = \prod_{i=1}^{s}\left(\frac{-1}{q_i}\right) = (-1)^{\sum \frac{q_i-1}{2}} = (-1)^{\frac{b-1}{2}}.$$

Item (6) If x and y are odd, then

$$\frac{x^2y^2-1}{8} - \left(\frac{x^2-1}{2} + \frac{y^2-1}{8}\right) = \frac{(x^2-1)(y^2-1)}{8} \equiv 0 \pmod{8}.$$

Hence

$$\frac{x^2-1}{8} + \frac{y^2-1}{8} \equiv \frac{x^2y^2-1}{8} \pmod{2}$$

Therefore,

$$\sum_{i=1}^{s} \frac{q_i^2 - 1}{8} \equiv \frac{b^2-1}{8} \pmod{2}.$$

But therefore,

$$\left(\frac{2}{b}\right) = \prod_{i=1}^{s}\left(\frac{2}{q_i}\right) = (-1)^{\sum \frac{q_i^2-1}{8}} = (-1)^{\frac{b^2-1}{8}}.$$

Item (7) Write prime factorisations for both b_1 and b_2, repeats allowed:

$$b_1 = p_1 \cdots p_s, \quad b_2 = q_1 \cdots q_t.$$

Then

$$\begin{aligned}\left(\frac{b_1}{b_2}\right) &= \prod_j \prod_i \left(\frac{p_i}{q_j}\right) \\ &= \prod_j \prod_i \left(\frac{q_j}{p_i}\right) (-1)^{\frac{p_i-1}{2}\cdot\frac{q_j-1}{2}}. \\ &= \left(\frac{b_2}{b_1}\right) (-1)^{\sum_i \sum_j \frac{p_i-1}{2}\cdot\frac{q_j-1}{2}}.\end{aligned}$$

However, we have

$$\sum_i \sum_j \frac{p_i-1}{2}\cdot\frac{q_j-1}{2} = \left(\sum_i \frac{p_i-1}{2}\right)\left(\sum_j \frac{q_j-1}{2}\right) \equiv \frac{b_1-1}{2}\cdot\frac{b_2-1}{2} \pmod 2.$$

□

The following theorem is, in fact, equivalent to Quadratic Reciprocity in the sense that the proof of each follows from the other fairly quickly.

Theorem 34.4. *Let p and q be distinct odd primes, and let $a \geq 1$. Then if $p \equiv \pm q \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.*

Proof. First, I claim that it suffices to do the case of odd prime a. For, suppose I already know this. If $a = 2^{e_0}p_1^{e_1}\cdots p_s^{e_s}$ is the prime factorisation of a, then

$$p \equiv \pm q \pmod{4a} \implies p \equiv \pm q \pmod{4p_i} \implies \left(\frac{p_i}{p}\right) = \left(\frac{p_i}{q}\right).$$

But since $p \equiv \pm q \pmod 8$, then $\left(\frac{2}{p}\right) = \left(\frac{2}{q}\right)$, and therefore we are done, by the definition of the Jacobi symbol.

Hence, we will assume that a is odd and prime. We will now do the case that $p \equiv q \pmod{4a}$. The other case is similar and is left for the

reader. We have $\left(\frac{p}{a}\right) = \left(\frac{q}{a}\right)$, and therefore,

$$\begin{aligned}\left(\frac{a}{p}\right) &= (-1)^{\frac{p-1}{2}\cdot\frac{a-1}{2}}\left(\frac{p}{a}\right)\\ &= (-1)^{\frac{p-1}{2}\cdot\frac{a-1}{2}}\left(\frac{q}{a}\right)\\ &= (-1)^{\frac{p-1}{2}\cdot\frac{a-1}{2}+\frac{q-1}{2}\cdot\frac{a-1}{2}}\left(\frac{a}{q}\right)\\ &= (-1)^{(a-1)\cdot\frac{p+q-2}{4}}\left(\frac{a}{q}\right)\end{aligned}$$

Write $p = q + 4at$ for some t. Then since q is odd,

$$p + q - 2 = q + 4at + q - 2 = 2(q-1) + 4at \equiv 0 \pmod 4.$$

Therefore, $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$. □

Theorem 34.5. *An integer a is a square if and only if a is a quadratic residue or 0 modulo all primes p.*

Proof. Exercise. □

This is an example of the Hasse principle.

35. Some questions about QRs and QNRs

Let P_r be the set of QRs modulo p, and let P_n be the set of QNRs modulo p. One of the main questions a number theorist may ask is to describe the distribution of P_r in the interval $1, 2, \ldots, p-1$.

We already know a few things:

(1) The sets P_r and P_n are both of size $\frac{p-1}{2}$.
(2) If $p \equiv 1 \pmod 4$, then:
$$a \in P_r \iff p - a \in P_r$$
(3) If $p \equiv 3 \pmod 4$, then:
$$a \in P_r \iff p - a \in P_n$$

Here's something else we can show:

Proposition 35.1. *The maximum number of residues between successive elements of P_n in the list*

$$1, 2, \ldots, p-1.$$

is $2\sqrt{p} + 1$.

Proof. The squares

$$1, 4, 9, 16, \ldots, (\lfloor\sqrt{p}\rfloor)^2$$

all lie in this interval and are QRs. Furthermore, the next square, $(\lfloor\sqrt{p}\rfloor + 1)^2 > p - 1$ so it is not in the interval. The difference of consecutive squares is odd numbers:

$$(n+1)^2 - n^2 = 2n + 1.$$

Therefore, the largest gap we see is at most $2\sqrt{p} + 1$. □

Vinogradov has conjectured the following:

Conjecture 35.2. *Let $\epsilon > 0$. Then*

(1) The number of elements between successive elements of P_n in the list

$$1, 2, \ldots, p - 1$$

is $O_\epsilon(p^\epsilon)$, i.e. it is bounded by $C_\epsilon p^\epsilon$ for some constant C_ϵ depending on ϵ.

(2) Let $N(p)$ be defined as the smallest element of P_n in the list

$$1, 2, \ldots, p - 1.$$

Then,

$$\lim_{p\to\infty} \frac{N(p)}{p^\epsilon} = 0.$$

This conjecture is still outstanding, although Burgess in 1957 showed that for large p, $N(p) < p^{1/4+\epsilon}$. The Riemann Hypothesis would imply that $N(p) \leq c_1(\ln p)^2$. On the other hand, this must be close to best possible, since Salié (1949) showed that $N(p) > c_2(\ln p)$ for infinitely many p.

36. Binary Quadratic Forms

A monomial $ax_1^{k_1} \cdots x_n^{k_n}$ has degree $k_1 + \cdots + k_n$. A polynomial in n variables has degree equal to the maximum of the degrees of its monomials. Such a polynomial is *homogeneous* (or a *form*) if all monomials have the same degree.

A *quadratic form* is often defined as a homogeneous polynomial of degree two. A *binary quadratic form* is a quadratic form in two variables. An *integral binary quadratic form* is a quadratic form with integer coefficients.

In other words, we are talking about things of the form

$$f(x, y) = ax^2 + bxy + cy^2$$

for some $a, b, c \in \mathbb{Z}$. An integral binary quadratic form is *primitive* if a, b, c do not have a common factor.

Here's the more abstract approach to quadratic forms.

Let V be an n-dimensional K-vector space. A *symmetric bilinear form* is a map $B : V \times V \to K$ such that

(1) $B(v, w) = B(w, v)$
(2) $B(\lambda u + v, w) = \lambda B(u, w) + b(v, w)$

Note that this implies that $B(0, v) = B(v, 0) = 0$ for all v.

Definition 36.1. *A function $f : V \to K$ is a quadratic form in terms of any basis of V if*

(1) $f(av) = a^2 f(v)$
(2) The function $B(v, w) := \frac{1}{2}(f(v + w) - f(v) - f(w))$ is a symmetric bilinear form.

Theorem 36.2. *These two definitions of a quadratic form agree! To be precise, a quadratic form as in the previous definition, when expressed as a function of coefficients with respect to a basis, is a binary degree two homogeneous polynomial. Conversely, such a polynomial, considered as a function on a vector space of dimension two, is a quadratic form as in the previous definition.*

We will only prove this for V of dimension 2, to keep notation manageable. A corresponding result for higher dimension applies.

Note that one of the definitions is basis dependendent and the other is not. In fact, a choice of basis specifies one of many polynomial quadratic forms that agrees with one functionally-defined quadratic form.

Proof. In the forward direction, we must simply check the properties:

(1) $f(a(x, y)) = f(ax, ay) = a^2 f(x, y)$.
(2) If we write $f(x, y) = ax^2 + bxy + cy^2$ and expand

$$B((x_1, y_1), (x_2, y_2)) = \frac{1}{2}\left(f(x_1 + x_2, y_1 + y_2) - f(x_1, y_1) - f(x_2, y_2)\right),$$

then the resulting polynomial is symmetric in (x_1, y_1) versus (x_2, y_2) and is linear in (x_1, y_1).

Let e_1, e_2 be a basis for V over K. Then

$$B(xe_1 + ye_2, xe_1 + ye_2) = x^2 B(e_1, e_2) + xy(2B(e_1, e_2)) + y^2 B(e_2, e_2)$$

By the definition of B, we have $f(v) = B(v, v)$. Then since

$$B(v + w, v + w) - B(v, v) - B(w, w) = 2B(v, w),$$

we have

$$f(xe_1 + ye_2) = x^2 f(e_1) + xy(f(e_1 + e_2) - f(e_1) - f(e_2)) + y^2 f(e_2).$$

□

Once we choose a basis, any symmetric bilinear form has a matrix representation as

$$\begin{pmatrix} x_2 & y_2 \end{pmatrix} \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = ax_1x_2 + b/2x_1y_2 + b/2x_2y_1 + cy_1y_2.$$

And corresponding quadratic form

$$ax^2 + bxy + cy^2.$$

It is crucial here that a quadratic form in the sense of the functional equation definition corresponds to many polynomial forms – in fact, one for each basis! If we change basis, this affects the matrix equation above; instead of matrix M, we obtain B^TMB for some invertible B. This corresponds to a change of variables $\mathbf{x} \mapsto B\mathbf{x}$. Since, as number theorists, we are interested in integral forms, we should require $B \in \mathrm{GL}_2(\mathbb{Z})$, so that it is a change of basis for the lattice $\mathbb{Z}^2$, which is where we consider our form to take values. This is called $\mathrm{GL}_2(\mathbb{Z})$-*equivalence* of forms.

Proposition 36.3. *Any quadratic form satisfies*

$$f(v+w) + f(v-w) = 2(f(v) + f(w)).$$

Proof. This is a consequence of the relation

$$B(v,w) + B(v,-w) = 0$$

where we plug in $B(v,w) = \frac{1}{2}(f(v+w) - f(v) - f(w))$ and use the fact that $f(w) = f(-w)$. □

37. The big questions for a quadratic form

The study of quadratic forms had its origin in the question of which numbers or primes are the sums of two squares. The big questions are:

(1) What integers does an integral binary quadratic form represent?
(2) What is the smallest value it represents?
(3) How many representations does a given integer have?

To access these questions, we'll consider quadratic forms that correspond to different bases (i.e. the same functionally-defined form) to be equivalent (we'll call this $\mathrm{GL}_2(\mathbb{Z})$-equivalent), and study these equivalence classes. Equivalent forms represent the same set of integers.

38. Conway's Sensual Quadratic Form

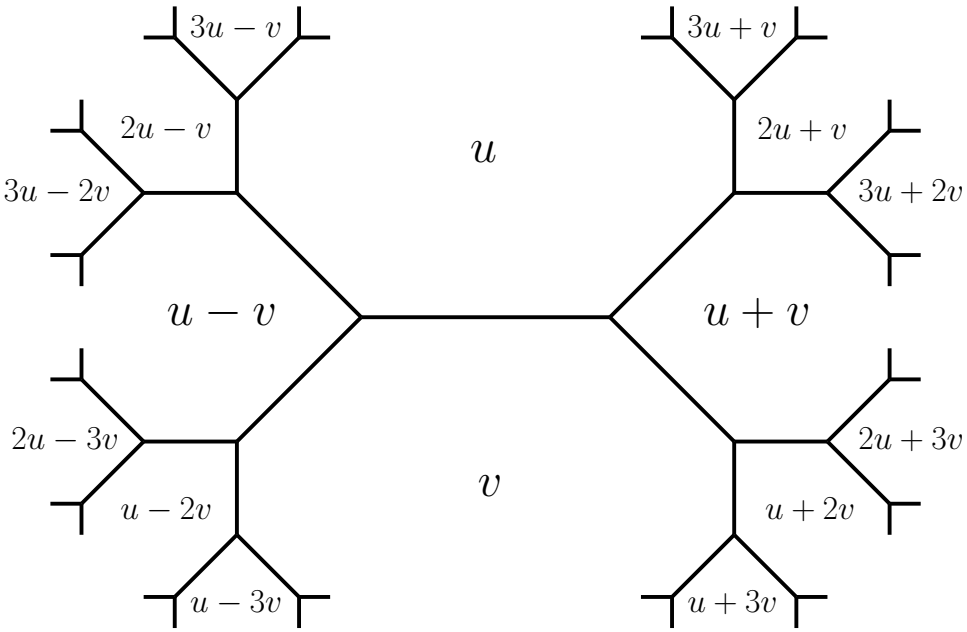
This section follows Conway and Fung, "The Sensual Quadratic Form", first chapter. An absolutely delightful book!

A *lax vector* in $\mathbb{Z}^2$ is an equivalence class of vectors up to sign, i.e. $\pm u$. A *basis* can also refer to a set of two lax vectors which form a basis for $\mathbb{Z}^2$ (a change of signs has no bearing on whether two vectors form a basis).

A *superbasis* is a set of three lax vectors wuch that any two of them form a basis for $\mathbb{Z}^2$. If u and v form a basis, then the only superbases containing u and v are

$$\{\pm u, \pm v, \pm(u+v)\} \quad \text{and} \quad \{\pm u, \pm v, \pm(u-v)\}.$$

Each superbasis contains three bases. Since each basis is contained in two superbases, we can draw a valence three graph representing the bases and superbases: edges are bases and vertices are superbases. We can place this graph into the plane in such a way that we can label the regions of the plane with lax vectors so that the following property holds: the boundary of each region consists of all bases and superbases containing that vector.

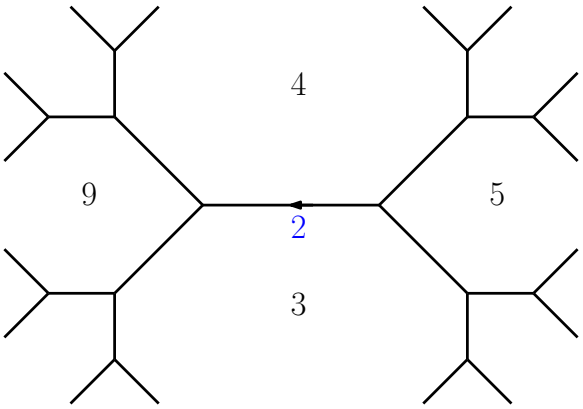


The parallelogram law then has a nice interpretation in terms of the picture: the three terms

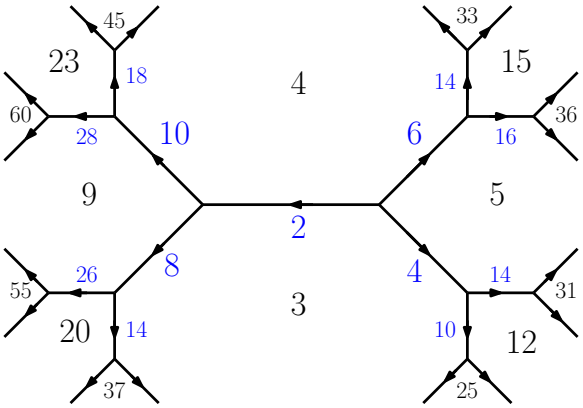
$$f(u-v), \quad f(u)+f(v), \quad f(u+v)$$

form an arithmetic progression. The vectors $u, v, u+v$ and $u-v$ surround one edge of the topograph: we can label these regions with the values of f at the respective vectors, and then we can label that edge with the common difference of the progression. The arrow will indicate the direction in which the progression increases.

For example, if $f(u+v) = 5$, $f(u) = 4$, $f(v) = 3$ and $f(u-v) = 9$, we have the following picture:

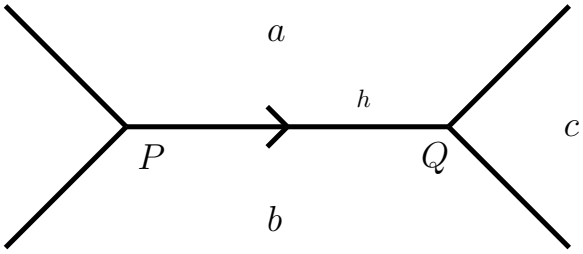


Starting with values surrounding any one vertex, this process, using the parallelogram law, will fill out all the values in the topograph. For example,



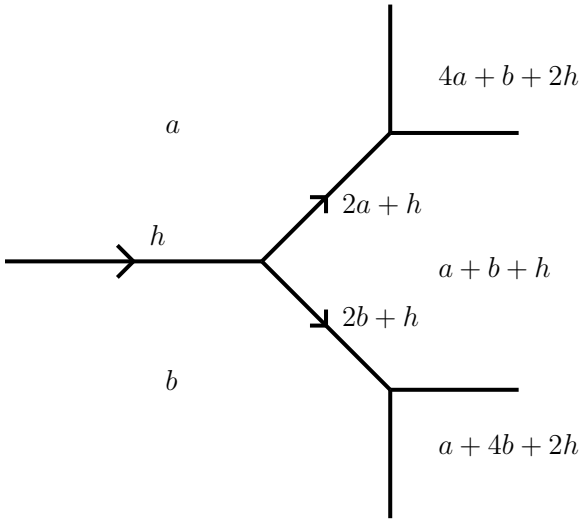
Of course, I haven't shown that this topograph is a tree, or is one connected piece; maybe it has many components, or maybe it has loops. We will do a little work to verify this:

Lemma 38.1 (The Climbing Lemma)**.** *Consider one location in the tree, given by this picture:*



Suppose $a, b, h > 0$. Then $c > 0$ and the edges emerging from the vertex labelled Q point away from Q, and furthermore, the edge-labels are larger than h.

Proof. By the arithmetic progression rule, $c = a + b + h > 0$. Therefore we can fill out some more of the topograph:



□

We will set some terminology concerning the values of a quadratic form.

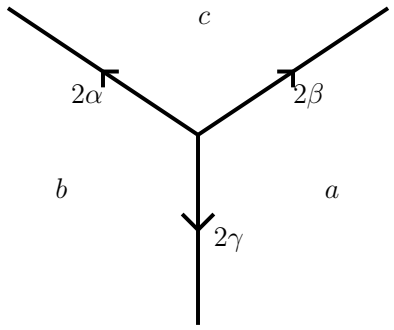
Definition 38.2. *A form is* indefinite *if it takes on both positive and negative values. A form is* positive semidefinite (negative semidefinite) *if $f(x, y) \geq 0$ ($f(x, y) \leq 0$) and the form represents zero. A form is* positive definite (negative definite) *if $f(x, y) > 0$ ($f(x, y) < 0$).*

A *well* is a vertex in the topograph such that no arrows point into it. Note that if $h = 0$, we may label an edge with no arrow.

Now, consider a positive definite form. By the climbing lemma, if we start anywhere and walk downhill (i.e. against arrows), we'll eventually have to stop, since we will see decreasing numbers as we walk, by the Climbing Lemma. Thus we'll have to have reach a well.

Lemma 38.3 (The Well Lemma)**.** *Let f be a positive definite quadratic form. The values around a well are the smallest values of the form.*

Proof. A well looks like this:



We have

$$2\alpha = b + c - a$$
$$2\beta = c + a - b$$
$$2\gamma = a + b - c$$

$$\alpha, \beta, \gamma \geq 0$$

Therefore

$$a = \beta + \gamma, \quad b = \alpha + \gamma, \quad c = \beta + \alpha.$$

So

$$a + b \geq c, \quad a + c \geq b, \quad b + c \geq a.$$

Let e_1, e_2, e_3 be the superbasis surrounding the well; suppose $e_1 + e_2 + e_3 = 0$, and suppose that

$$f(e_1) = a, f(e_2) = b, f(e_3) = c.$$

Let us write a general vector v as

$$v = m_1 e_1 + m_2 e_2 + m_3 e_3$$

Claim: the following formula (Selling's Formula) holds:

$$f(v) = \alpha(m_2 - m_3)^2 + \beta(m_1 - m_3)^2 + \gamma(m_1 - m_2)^2.$$

The proof is that both sides are quadratic forms which agree on the superbasis e_1, e_2, e_3. But the values on a superbasis determine a form: for example, recall the useful formula

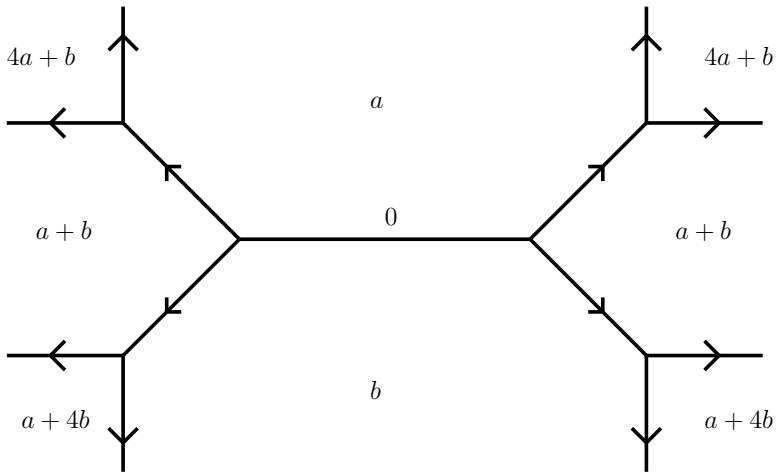
$$f(xe_1 + ye_2) = x^2 f(e_1) + xy(f(e_1 + e_2) - f(e_1) - f(e_2)) + y^2 f(e_2).$$

For all v besides the e_i, the corresponding m_i are all distinct so, by Selling's Formula,

$$f(v) \geq \alpha + \beta + \gamma \geq a, b, c.$$

□

As a consequence, if $\alpha, \beta, \gamma > 0$, then the well is unique. There's also the possibility of a double well, i.e. where one of α, β or γ is zero. It looks like this:

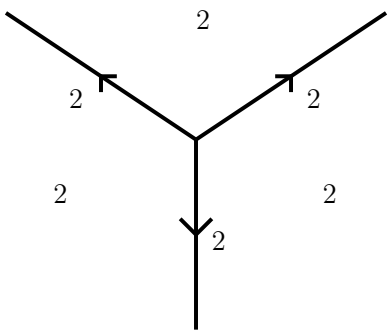


In this case, the well is not unique, but there are two identical wells adjacent to each other.

A 'triple well', i.e. where two of α, β and γ are zero, is not possible. For, suppose without loss of generality that $b + c = a$ and $c + a = b$. Then $c = 0$ which is not possible for a positive definite quadratic form.

Proposition 38.4. *The topograph is a connected tree.*

Proof. Consider a form with a simple well (i.e. not a double well). We might as well take the form with $\alpha = \beta = \gamma = 1$. Then, starting anywhere in the topograph, climb down against the arrows. One must eventually stop, and if one stops, one has reached a well. But this well is the unique well, i.e. the spot in the topograph with the three smallest values. Here it is:

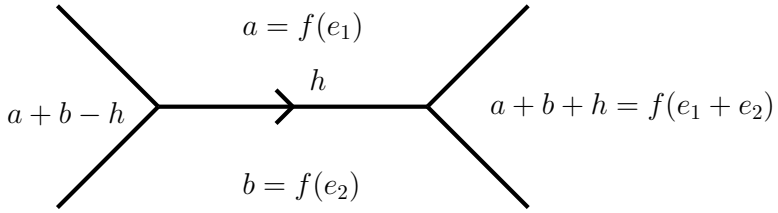


So all components contain the well: in other words, the topograph is connected.

Furthermore, the climbing lemma also rules out loops in the topograph; since on a path, the numbers in your neighbourhood can only increase, you cannot return to your starting point.[4] □

39. The discriminant of a quadratic form

Let's look at a piece of the topograph surrounding any one edge, and the corresponding superbasis:



Knowing the values of the form around this edge tells us the whole form. The form is given by

$$f(xe_1+ye_2) = x^2f(e_1)+xy(f(e_1+e_2)-f(e_1)-f(e_2))+y^2f(e_2) = ax^2+hxy+by^2.$$

In other words, this is the form when expressed in terms of basis e_1, e_2. We can see all the different $\mathrm{GL}_2(\mathbb{Z})$-equivalent polynomial forms by looking around each edge! Actually, we need to be slightly careful here; if we change the order of the basis, we get $bx^2+hxy+ay^2$. If we imagine pointing the other way on the same edge, we get $ax^2-hxy+by^2$. And if we do both, we get $bx^2-hxy+ay^2$. All of these are equivalent, but they are all different ways to "read" the one edge.

The discriminant of a quadratic form is defined in terms of its coefficients:

$$\Delta_f = h^2 - 4ab.$$

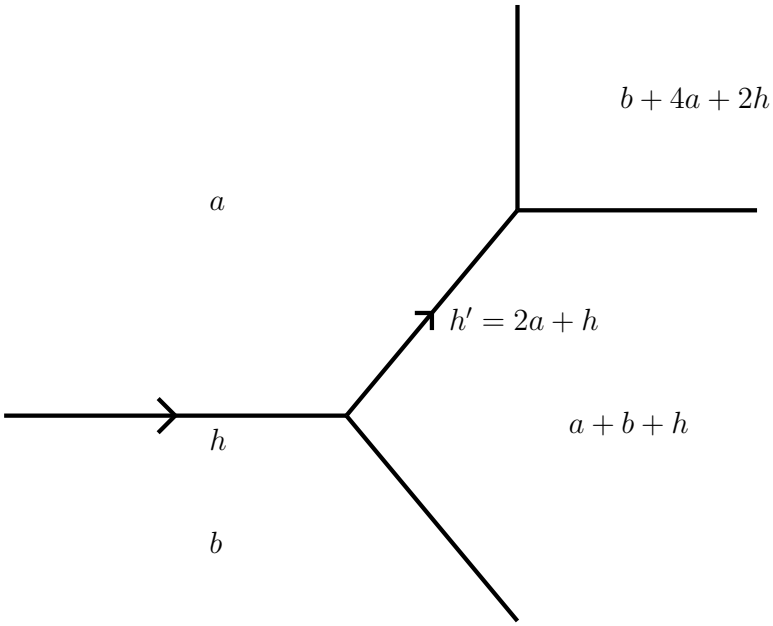
[4] I forgot to mention this in class!

(This is just the same as the discriminant of the quadratic equation we obtain if $y = 1$.) We can view the discriminant as given by the values surrounding an edge in the topograph.

The discriminant is a square in $\mathbb{Z}$ if and only if f factors into linear terms in $\mathbb{Z}$. For example, $x^2 - y^2 = (x - y)(x + y)$ has $\Delta = 4$, while $10x^2 - 27xy + 18y^2 = (2x - 3y)(5x - 6y)$ has $\Delta = 9$. I leave the formal proof of this as an exercise.

Proposition 39.1. *The discriminant doesn't depend on the choice of edge used to obtain it; therefore it is a property of the form itself.*

Proof. Here are two adjacent edges:



It is a brief calculation to check that the discriminants are the same at both edges h and h':

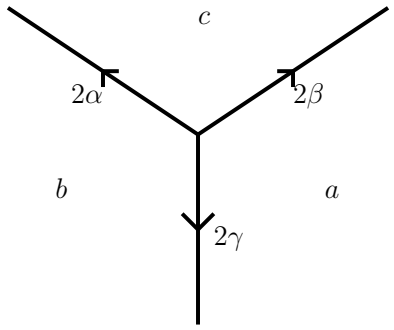
$$(2a + h)^2 - 4a(a + b + h) - (h^2 - 4ab) = \cdots = 0.$$

Since the topograph is connected, this is enough to guarantee that the discriminant is an invariant of the form, even though we defined it by looking at one edge. □

This tells us that the discriminant is an invariant under $\mathrm{GL}_2(\mathbb{Z})$-equivalence of the polynomial forms.

Proposition 39.2. *A positive definite or negative definite form has a negative discriminant.*

Proof. We will do the proof for a positive definite form; a negative definite one is similar. We can define the discriminant using one edge of the well.



We obtain

$$\Delta_f = (c+a-b)^2 - 4ac = a^2 + b^2 + c^2 - 2ac - 2ab - 2cb \leq 0.$$

Recall that

$$\begin{aligned} a + b &\geq c > 0 \\ a + c &\geq b > 0 \\ b + c &\geq a > 0 \end{aligned}$$

and furthermore, at least two of the '$\geq$' are strict, by the same computation that showed there are no 'triple wells'. This implies that

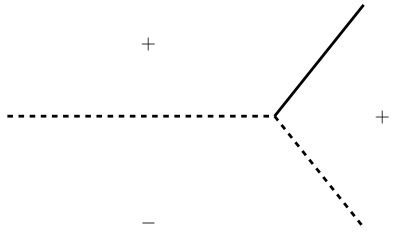
$$\begin{aligned} ac + bc &\geq c^2 > 0 \\ ab + bc &\geq b^2 > 0 \\ ab + ac &\geq a^a > 0 \end{aligned}$$

where again at least two of the '$\geq$' are strict. From this we conclude that

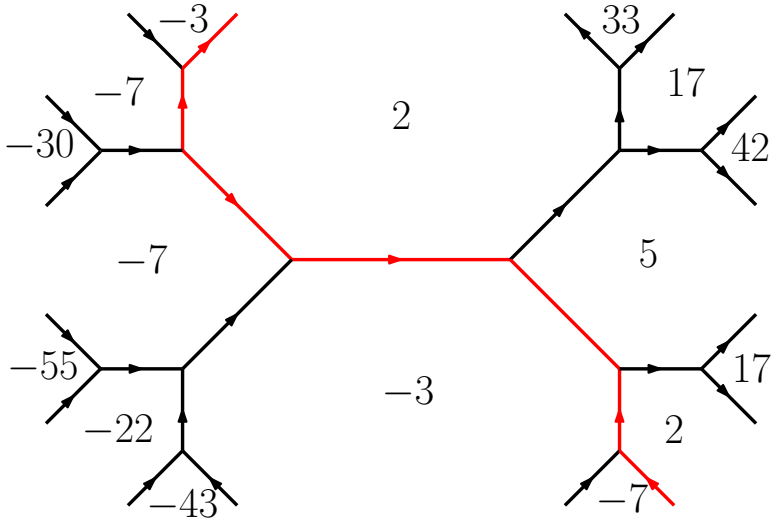
$$2(ac + ab + cb) > a^2 + b^2 + c^2.$$

This tells us that the discriminant is negative. □

An indefinite form which does not represent zero must have a positive region adjacent to a negative one. An edge between regions of different sign will be called a *river edge* (shown in the diagrams as a dotted edge). If we have a river edge entering a vertex, another river edge must come out:



In this way, we obtain a path with no endpoints. This is the *river*. Here's an example of an indefinite form with its river:



By the Climbing Lemma, as we move away from the river, we larger numbers in absolute value (either negative or positive, depending which 'bank' we are climbing). This implies that the river is unique – we can't run into another river by climbing away from the first.

Proposition 39.3. *An indefinite form not representing zero has a positive discriminant.*

Proof. Take a river edge to compute the discriminant, so the two adjacent regions have values $a > 0$ and $c < 0$. Then

$$\Delta_f = h^2 - 4ac = (+) - 4(+)(-) > 0.$$

□

Proposition 39.4. *The river is periodic.*

Proof. Consider the discriminant as defined by a river edge. Then

$$|\Delta_f| = h^2 + 4|ac|$$

so that

$$|h| \leq \sqrt{\Delta_f}, \quad |ac| = \frac{1}{4}(|\Delta_f| - h^2).$$

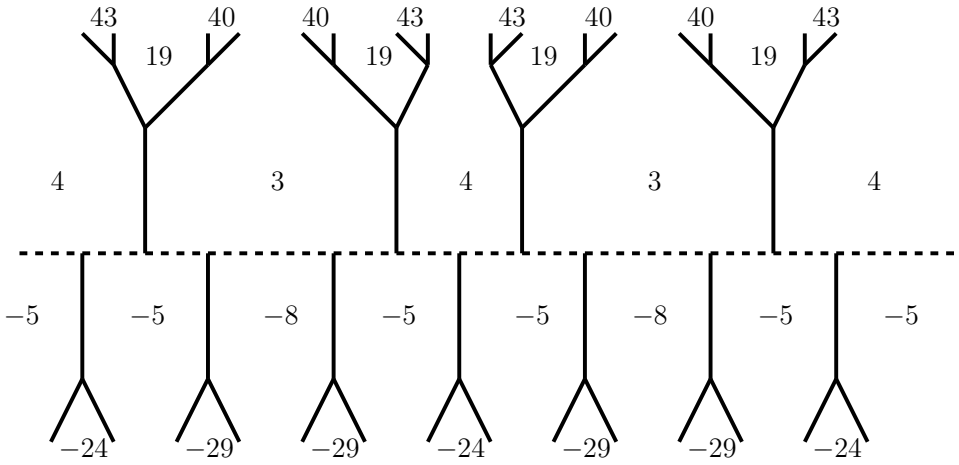
From this we see that there are only finitely many possible values for h and only finitely many possible value for a and c along a river edge. Therefore, since the river is infinite, it must eventually see the same a, c, h again.

But once we see the same values surrounding an edge, we must begin repeating everything, as the values around and on an edge determine the values everywhere else. □

Here's an example river, given by $3x^2 + 6xy - 5y^2$. We have

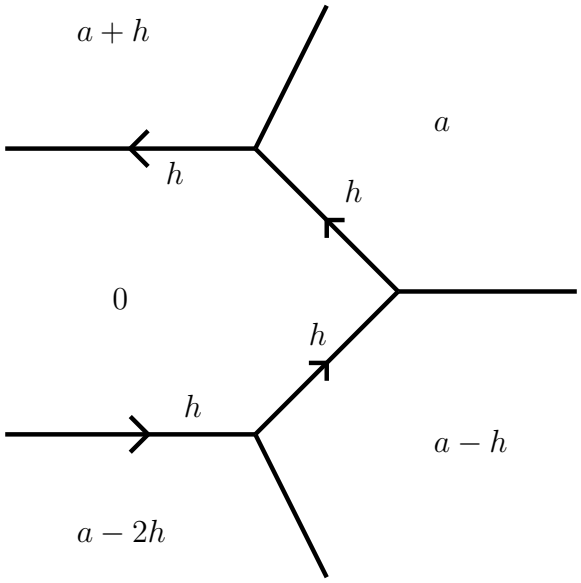
$$f(1,0) = 3, \quad f(0,1) = -5, \quad f(1,1) = 4$$

and therefore these values surround a piece of the river. Following it, we see



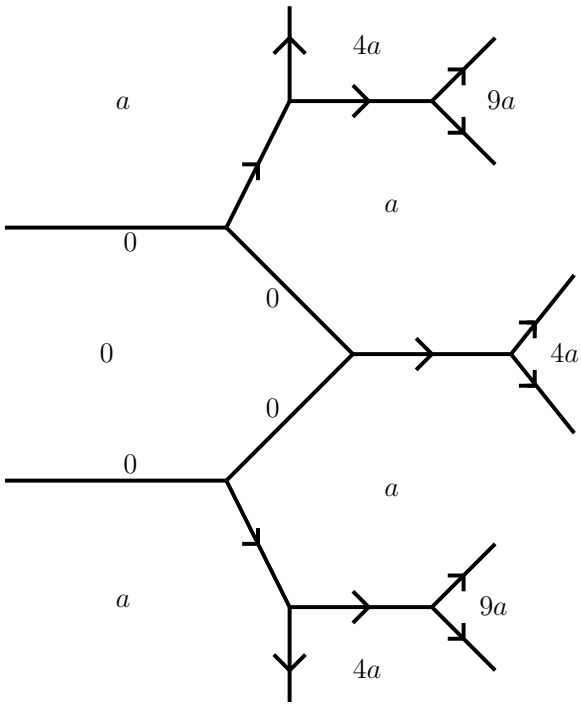
From this picture (and the Climbing Lemma), we may immediately conclude that $3x^2 + 6xy - 5y^2 = 7$ has no solution in the integers.

If a form represents zero, the corresponding region is called a *lake*. In general, by the arithmetic progression rule (aka parallelogram law), the regions surrounding the lake have values in arithmetic progression:



Proposition 39.5. *A positive or negative semidefinite form has zero discriminant.*

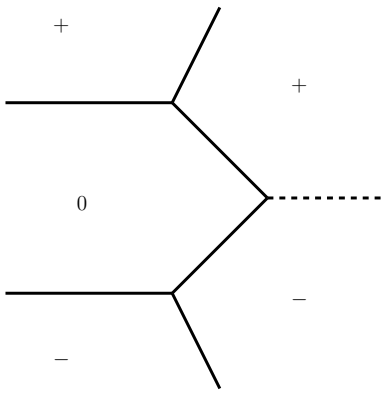
Proof. If f is a positive semidefinite form, then $h = 0$ (because otherwise this arithmetic progression along the shore contains negative values). Then we are in a very special situation that looks like this:



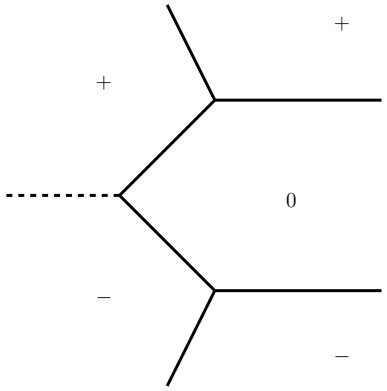
Taking the discriminant from a lake edge, say, we see that $\Delta_f = 0$. $\square$

Proposition 39.6. *An indefinite form which represents zero has non-zero square discriminant.*

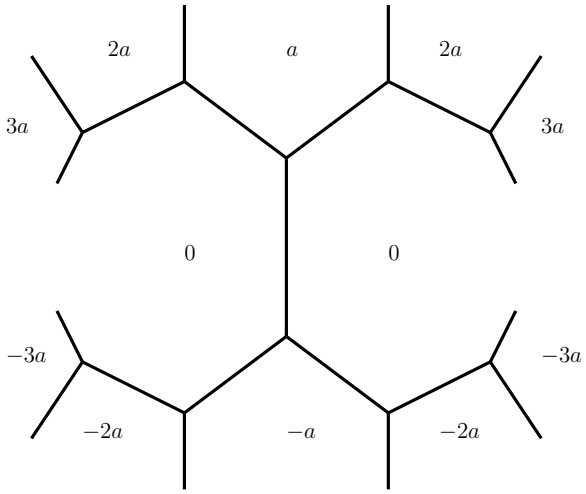
Proof. We have a lake, but we are not in the special case of the special semidefinite form lake. So along the shore, we see positive and negative terms. Therefore we cross the river somewhere (or another lake). If we cross a river, the river flows into (out of?) the lake. We have something like this:



By periodicity, the river must eventually hit another lake:



Therefore the river is of finite length. It is possible, of course, that it has zero length:



In any of these cases, take an edge which is a shore of a lake, and consider the discriminant one obtains: it is h^2. □

If the discriminant of a form is a square, that means the form factors over $\mathbb{Z}$: if so, then it must take the value 0 (for example, one factor $ax + by$ vanishes when $x = -b, y = a$). With this remark, we have now collected information about the discriminant of all the different types of forms. We can form a table summarizing what we've learned. To tell the difference between positive or negative (semi)definite forms, it suffices to check whether the coefficients a or c are positive or negative.

In summary, then for a quadratic form $ax^2 + bxy + cy^2$,

condition	values of form
$\Delta_f > 0$ square	$+, -, 0$
$\Delta_f > 0$ nonsquare	$+, -$
$\Delta_f = 0$, and $a > 0$ or $c > 0$	$+, 0$
$\Delta_f = 0$, and $a < 0$ or $c < 0$	$-, 0$
$\Delta_f < 0$, $a > 0$	$+$
$\Delta_f < 0$, $a < 0$	$-$

40. Projective Linear Groups

Here are some extremely useful matrix groups:

$$\begin{aligned}
\mathrm{GL}_2(\mathbb{Z}) &= \{\text{ordered bases of } \mathbb{Z}^2\} \\
&= \{\text{changes of bases of } \mathbb{Z}^2\} \\
&= \{2 \times 2 \text{ matrices with determinant } \pm 1\}. \\
\mathrm{PGL}_2(\mathbb{Z}) &= \{\text{ordered bases of } \mathbb{Z}^2 \text{ up to scaling by } \pm 1\} = \mathrm{GL}_2(\mathbb{Z})/\{\pm I\}. \\
\mathrm{SL}_2(\mathbb{Z}) &= \{\text{positively oriented ordered bases of } \mathbb{Z}^2\} \\
&= \{\text{changes of bases of } \mathbb{Z}^2 \text{ which preserve orientation}\} \\
&= \{M \in \mathrm{GL}_2(\mathbb{Z}) : \det(M) = 1\}. \\
\mathrm{PSL}_2(\mathbb{Z}) &= \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}.
\end{aligned}$$

The group $\mathrm{PSL}_2(\mathbb{Z})$ contains the element $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \sim \begin{pmatrix} -1 & -1 \\ 0 & -1 \end{pmatrix}$ (this notation denotes that these two matrices are equivalent in $\mathrm{PSL}_2(\mathbb{Z})$), and the element $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ which correspond to a shear and a ninety-degree rotation. (Note: we will say 'contains the matrix', meaning, of course, that it contains an equivalence class represented by that matrix.)

The group $\mathrm{PGL}_2(\mathbb{Z})$ contains the matrix $U = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \sim \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$, which reverses the direction of one basis vector.

We can think of $\mathrm{PSL}_2(\mathbb{Z})$ as a subgroup of $\mathrm{PGL}_2(\mathbb{Z})$ because $\mathrm{SL}_2(\mathbb{Z})$ is a subgroup of $\mathrm{GL}_2(\mathbb{Z})$. In particular, $\mathrm{PSL}_2(\mathbb{Z})$ is the subgroup of exactly those equivalence classes in $\mathrm{PGL}_2(\mathbb{Z})$ which contain a matrix representative of determinant 1. (However, note that if any one representative has determinant 1, then so does the other representative.)

Theorem 40.1. *$\mathrm{PSL}_2(\mathbb{Z})$ is generated by S and T. $\mathrm{PGL}_2(\mathbb{Z})$ is generated by S, T and U. Furthermore,*

$$[\mathrm{PGL}_2(\mathbb{Z}) : \mathrm{PSL}_2(\mathbb{Z})] = 2.$$

Proof. The fact that

$$[\mathrm{PGL}_2(\mathbb{Z}) : \mathrm{PSL}_2(\mathbb{Z})] = 2$$

is actually easier than the statements about generation, for, $\mathrm{PSL}_2(\mathbb{Z})$ is the kernel of the determinant map.

$$\det : \mathrm{PGL}_2(\mathbb{Z}) \to \{\pm 1\}, \quad M \mapsto \det(M).$$

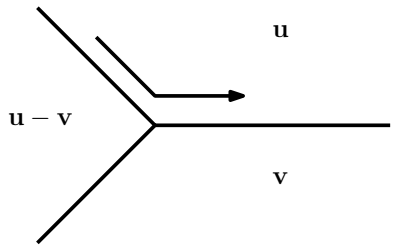
which is a well-defined group homomorphism (this requires checking that the $\det(M)$ is the same on both elements of an equivalence class, but this check is easy). Once we know S, T and U generate $\mathrm{PGL}_2(\mathbb{Z})$, then one can check that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by T and S easily. For, every element of $\mathrm{PGL}_2(\mathbb{Z})$ can be expressed as a word in S, T and U with exactly one instance of U at the beginning (this follows from the identities $SU = US$ and $TU = UT^{-1}$). So PGL_2 is a union of two cosets: PSL_2 and $U \cdot \mathrm{PSL}_2$.

On to the topograph. It will suffice to show how each of S, T and U act as changes of bases on the topograph. Consider the following action of $\mathrm{PGL}_2(\mathbb{Z})$ on ordered bases up to ± 1:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot (\mathbf{e}_1, \mathbf{e}_2) = (a\mathbf{e}_1 + c\mathbf{e}_2, b\mathbf{e}_1 + d\mathbf{e}_2).$$

First, I claim that this action can take any ordered basis to any other. But this is clear, since every pair of ordered bases is a linear combination in terms of any other, with coefficients forming a matrix of determinant one.[5]

We will associate the data of an ordered basis up to ± 1 to a *path through a superbasis*. That is, if we specify a path by an arrow thus:

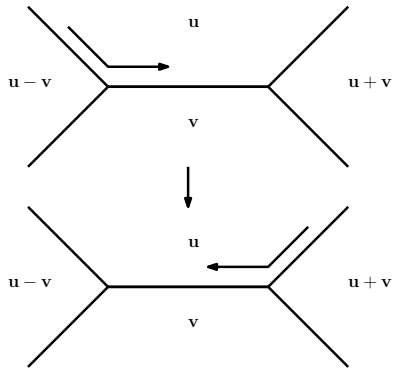


then the associated ordered basis is given by first the basis element common to both edges in the path ($\mathbf{u}$ in the picture), then the other basis element on the output edge ($\mathbf{v}$ in the picture). We must choose signs on these lax vectors so that the third superbasis element consists of $\mathbf{u} - \mathbf{v}$: if we change the sign of $\mathbf{u}$, we must also change the sign of $\mathbf{v}$. Therefore it specifies an ordered basis up to ± 1.

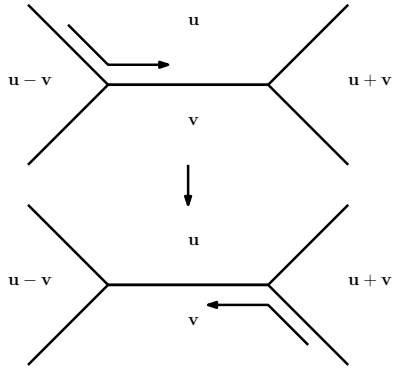
Then the actions of U, S and T are as follows:

U is a 'reflection':

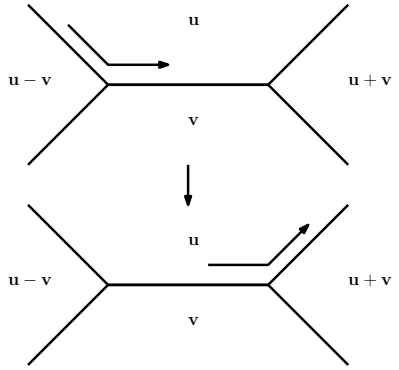
[5]Note that my action is a little odd, perhaps a transpose of what you'd expect?



S is a 'rotation':



T is a 'translation':



With these moves available to us, we see that we can traverse the topograph to arrive at any other basis in the same component. For example, T allows us to move along the shore of any region, and S allows us to cross an edge to a new basin. U lets us change the direction we're facing along a shore. Because the topograph is connected, this implies that S, T and U must be sufficient to arrive at any other ordered basis. So they generate all of $\mathrm{PGL}_2(\mathbb{Z})$.

□

Remark. Note that there are actually at least two different ways to imagine $\mathrm{PGL}_2(\mathbb{Z})$ acting on the collection of ordered bases up to ± 1. You could write your basis in terms of the standard basis and use the usual interpretation of a matrix as a transformation in terms of the standard basis. That's actually not what we're doing here. If we use that action, then it acts on the whole topograph, preserving adjacencies. In our action in the proof above, adjacencies are not preserved; this is an action that describes 'walking around on the topograph.' Note that if 'walk forward and take the left fork' on two edges adjacent at their 'tail,' then the new edges we arrive at are no longer adjacent! In other words, applying an element of $\mathrm{PGL}_2(\mathbb{Z})$ (with the action of the proof) to all three bases of a superbasis could result in three new bases whose union is *not* a superbasis.

41. Equivalence of Quadratic Forms

Recapping what we've seen already in one convenient location:

If $f : \mathbb{Z}^2 \to \mathbb{Z}$ is a quadratic form, then for each choice of ordered basis e_1, e_2 for $\mathbb{Z}^2$, we can write down f as a polynomial in two variables:

$$f(xe_1+ye_2) = x^2 f(e_1)+xy(f(e_1+e_2)-f(e_1)-f(e_2))+y^2 f(e_2) = ax^2+hxy+by^2.$$

Any two forms obtained this way from the same function on $\mathbb{Z}^2$ are called $\mathrm{GL}_2(\mathbb{Z})$-equivalent. Therefore, $AX^2+BXY+CY^2$ is equivalent to $ax^2+bxy+cy^2$ if and only if there's a change of basis $Y = \alpha x + \beta y$, $Y = \gamma x + \delta y$ taking one to the other. That is, plugging in the change of variable for X and Y to the form $AX^2 + BXY + CY^2$ gives us $ax^2 + bxy + cy^2$.

Let us write

$$M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}.$$

Then, another way to say what we've just said is that

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

It is easy to see that equivalent forms represent the same integers. One question we would like to answer is whether the converse is true.

It is also a consequence of the last section that *equivalent quadratic forms have the same discriminant*. It is also easy to see the invariance of the discriminant directly: Δ is the determinant of the matrix of the form. In the last equation, we are multiplying by M and M^T, so the determinant changes by $\det(M)\det(M^T) = \det(M)^2 = 1$.

42. Reduction of Quadratic Forms

To compare whether two numbers are in the same equivalence class modulo N, we reduce them both to residues in the window $0 \leq x \leq N$ and see if we obtain the same residue. In other words, each equivalence class modulo N has a 'least positive' representative. By reducing any given number to that choice within its equivalence class, we can compare equivalence classes.

To do the same thing with quadratic forms is called the 'reduction of quadratic forms'. We designate one representative of each equivalence class of quadratic forms as 'reduced' and we develop an algorithm for 'reducing' to that choice.

To do this, we'll use the group $\mathrm{PGL}_2(\mathbb{Z})$, since these represent changes of bases. Note that the change of basis given by

$$X \mapsto -X, \quad Y \mapsto -Y$$

does not change a quadratic form $aX^2 + bXY + cY^2$. Therefore, the two matrices $M = I, -I$ fix the form. That is why is makes sense to use $\mathrm{PGL}_2(\mathbb{Z})$ instead of $\mathrm{GL}_2(\mathbb{Z})$.

We obtain an action of $\mathrm{PGL}_2(\mathbb{Z}) = \langle S, T, U \rangle$ on the members of one equivalence class of quadratic forms. By acting on a given form with appropriate choices from $\mathrm{PGL}_2(\mathbb{Z})$, we can 'move' around in the equivalence class until we reach the reduced form.

We will restrict for the moment to positive definite quadratic forms. Each element of $\mathrm{PGL}_2(\mathbb{Z})$ acts on the coefficients of the form. We summarise the actions of the generating elements in a table:

transformation	element of $\mathrm{PGL}_2(\mathbb{Z})$	where $ax^2 + bxy + cy^2$ goes, i.e. $(a, b, c) \mapsto (a', b', c')$	usefulness or possible effect
$X = y, Y = -x$	S	$(c, -b, a)$	make $a \leq c$
$X = x + y, Y = y$	T	$(a, b + 2a, a + b + c)$	make b smaller
$X = x - y, Y = y$	T^{-1}	$(a, b - 2a, a - b + c)$	until $\lvert b \rvert \leq a$
$X = x, Y = -y$	U	$(a, -b, c)$	change sign of b

By using these, we can alter our quadratic form bit by bit until $0 \leq b \leq a \leq c$. We call a quadratic form $ax^2 + bxy + cy^2$ satisfying these inequalities on its coefficients *reduced*.

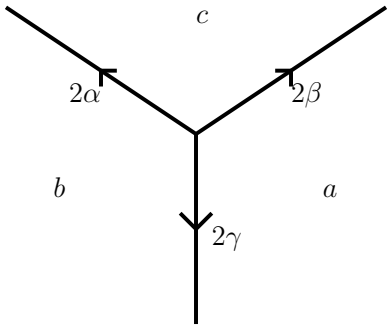
Warning: some authors use PSL_2 instead of PGL_2 and obtain a different, slightly weaker meaning of 'reduced.'

Proposition 42.1. *There is exactly one reduced form in every equivalence class of positive definite quadratic forms.*

Proof. I claim that the reduced form is the form obtained at the well in the topograph. To be precise, we take the edge of the well surrounded by the two smallest values of the quadratic form. Recall that an edge surrounded by values a, b, $a+b-h$ and $a+b+h$ corresponds to the four forms

$$ax^2 \pm hxy + by^2, \quad bx^2 \pm hxy + ay^2.$$

Only one of these can be reduced. Recall that the well looks like this for some a, b, c:



where

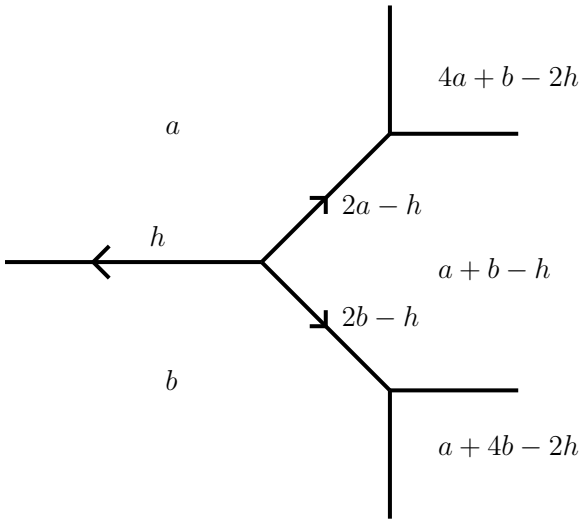
$$\begin{aligned} 2\alpha &= b + c - a \\ 2\beta &= c + a - b \\ 2\gamma &= a + b - c \end{aligned}$$

Suppose that $a \leq b \leq c$ (these are different uses of a, b, c than in the previous paragraphs). Then the edge 2γ corresponds to form

$$ax^2 + 2\gamma xy + by^2.$$

Since $0 \leq 2\gamma = a + b - c \leq a \leq b$, this form is reduced.

Now suppose we have found an edge which is a reduced form. Then $0 \leq h \leq a \leq b$ in the picture



so that by the climbing lemma, the well is to the right of this edge. But in fact, $2a-h, 2b-h > 0$, so the well is the vertex immediately to the right. Furthermore, $a+b-h > a, b$, so the edge is between the two smallest values of the form. □

If a form is reduced, then $b^2 \leq ac$, so

$$-\Delta = 4ac - b^2 \geq 3ac$$

which implies that $a, b, c \leq \frac{1}{3}|\Delta|$. Therefore, there are only finitely many reduced forms of a given discriminant. This implies there are only finitely many equivalence classes of reduced forms for a given discriminant. Question: just how many are there?

Example 42.2. *Here's an example of reducing a quadratic form. Consider the form* $5x^2 - 16xy + 14y^2$. *This has* $\Delta = 16^2 - 4 \cdot 5 \cdot 14 = -24$.

$$(5, -16, 14) \xrightarrow{T} (5, -6, 3) \xrightarrow{T} (5, 4, 2) \xrightarrow{S} (2, -4, 5) \xrightarrow{T} (2, 0, 3).$$

The result is reduced because $0 \leq b \leq a \leq c$. *The reduced form is* $2x^2 + 3y^2$ *with* $\Delta = 0^2 - 4 \cdot 2 \cdot 3 = -24$. *(It's always good to check that the discriminant is invariant under your supposed reduction.)*

A useful pattern for reducing efficiently is as follows: apply S *when* $a > c$ *and then as many* T*'s as needed to get* $|b| \leq a$, *then repeat as necessary. Finally, at the end apply* U *if you have a negative* b.

We can view reduction as 'moving around the topograph' to find the 'reduced' form at the well.

43. Algebraic Number Theory

For this I'm following Samuel, Chapter 2. So please refer to that. I'll just make a few extra comments here. He does things in somewhat more generality, which will be useful to you later, and is no more work, but our concern here is with *number fields*, i.e. finite extensions of $\mathbb{Q}$. Each such field K has a ring of integers, $\mathcal{O}_K$, i.e. the integral closure of $\mathbb{Z}$ inside K. It is necessarily an integral domain (for example, it can be embedded in $\mathbb{C}$).

Samuel defines elements integral over a ring. But the two basic cases to keep in mind are *algebraic numbers*, i.e. roots of polynomials in $\mathbb{Q}[x]$, equivalently of monic polynomials in $\mathbb{Q}[x]$, equivalently roots of polynomials in $\mathbb{Z}[x]$; and *algebraic integers*, i.e. roots of monic polynomials in $\mathbb{Z}[x]$.

Here's a useful lemma that Samuel doesn't explicitly do:

Proposition 43.1. *If α is an algebraic number, then there is a rational integer b such that $b\alpha$ is an algebraic integer.*

Proof. Let $f \in \mathbb{Q}[x]$ be the minimal polynomial of α. Write

$$f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

Let b be the smallest positive rational integer such that all $ba_i \in \mathbb{Z}$. Then $b\alpha$ is a root of

$$x^n + ba_{n-1}x^{n-1} + b^2 a_{n-2}x^{n-2} + \cdots + b^{n-1}a_1 x + b^n a_0 \in \mathbb{Z}[x].$$

Therefore $b\alpha$ is an algebraic integer. □

Example 43.2. *If $K = \mathbb{Q}(i)$, we saw last time that the algebraic integers in K are exactly the Gaussian integers, $\mathbb{Z}[i]$.*

44. Units in rings of integers

Write $\mathcal{O}_K^*$ to denote the units in $\mathcal{O}_K$.

Proposition 44.1. *$\alpha \in \mathcal{O}_K$ is a unit if and only if $N(\alpha) = \pm 1$.*

Proof. If α is a unit, then $\alpha\beta = 1$ for some $\beta \in \mathcal{O}_K$. So $N(\alpha)N(\beta) = 1$. Since α and its conjugates are integers and $N(\alpha) \in \mathbb{Q}$, we must have $N(\alpha) \in \mathbb{Z}$, and similarly $N(\beta) \in \mathbb{Z}$. Therefore, $N(\alpha) = N(\beta) \in \{\pm 1\}$. Conversely, if $N(\alpha) = \pm 1$, then $\prod_{i=1}^n \alpha_i = \pm 1$, so that α has as its inverse the product of all other conjugates (possibly negated). But if $\alpha \in \mathcal{O}_K$, then so are all its conjugates and hence this product. So α is a unit in $\mathcal{O}_K$. □

45. Some Linear Algebra

The main adjustment I make to Samuel is to spend a bit more time on the linear algebra. Let $A \subset B$ be rings, with B a free A-module of finite rank n. Then B has a basis b_i over A, and any element $\alpha \in B$ gives a map

$$x \mapsto \alpha x$$

which is a linear transformation on B. Therefore it has a matrix with respect to the basis of B over A.

For example, if α is integral over A, and $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in A[x]$ is an irreducible polynomial of which α is a root, then $B = A[\alpha]$ has basis $1, \alpha, \alpha^2, \ldots, \alpha^{n-1}$ and the matrix of α is

$$\begin{pmatrix} 0 & 0 & \cdots & -a_0 \\ 1 & 0 & \cdots & -a_1 \\ 0 & 1 & \cdots & -a_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -a_{n-1} \end{pmatrix}.$$

This matrix can be diagonalized if we extend scalars to some larger ring C containing A.

Formally, we can do this with a tensor product. Replace $B = \sum Ab_i$ with $C \otimes_A B = \sum C(1 \otimes b_i)$. The elements of this ring are tensors, i.e. pairs $c \otimes a$ where you think of c as the coefficient and a as the vector; the basis is $1 \otimes b_i$. The tensors are linear in each factor, and have properties

$$c(d \otimes a) = cd \otimes a,$$
$$ac \otimes b = c \otimes ab \text{ for } a \in A$$

Extending scalars to B itself, we find that α is an eigenvalue of the matrix, since it satisfies its own characteristic polynomial.

In the case of a field extension, the characteristic polynomial is a power of the minimal polynomial and the eigenvalues are the conjugates of α, so the diagonalized matrix has $\sigma_1(\alpha), \sigma_2(\alpha), \ldots, \sigma_n(\alpha)$ along the diagonal.

46. Quadratic Fields

A number field K is called a *quadratic field* if it has $[K : \mathbb{Q}] = 2$.

Let $K = \mathbb{Q}(\alpha)$, where α is an algebraic number of degree 2. As a root of a quadratic equation, α must be of the form $a + b\sqrt{D}$ where $a, b \in \mathbb{Q}$, $D \in \mathbb{Z}$. So $K = \mathbb{Q}(\sqrt{D})$. In fact, we may replace D with its

squarefree part, so that D becomes squarefree. All elements α of K are of the form

$$\alpha = a + b\sqrt{D}$$

for some rational a and b. If $D < 0$, we call this field *imaginary* and if $D > 0$, it is called *real*.

Since both roots of $x^2 - D$ are in K, it is a *Galois extension* of $\mathbb{Q}$, i.e. the images of its two embeddings in an algebraically closed field are the same, so it has two automorphisms. The automorphism group of K consists of the identity and the map $a + b\sqrt{D} \mapsto a - b\sqrt{D}$. Therefore,

$$N(\alpha) = (a+b\sqrt{D})(a-b\sqrt{D}) = a^2 - Db^2, \quad tr(\alpha) = a+b\sqrt{D}+a-b\sqrt{D} = 2a.$$

The goal of this section is to describe the ring of integers $\mathcal{O}_K$ of K.

Proposition 46.1.

$$\alpha = a + b\sqrt{D} \in \mathcal{O}_K \iff N(\alpha), tr(\alpha) \in \mathbb{Z}.$$

Proof. Suppose that $\alpha = a + b\sqrt{D} \in K$. Then the minimal polynomial of α is

$$x^2 - tr(\alpha)x + N(\alpha).$$

But $\alpha \in \mathcal{O}_K$ if and only if this polynomial is in $\mathbb{Z}[x]$. □

Proposition 46.2. *If $D \equiv 2, 3 \pmod 4$, then*

$$\mathcal{O}_K = \mathbb{Z}[\sqrt{D}] = \mathbb{Z} + \sqrt{D}\mathbb{Z}.$$

If $D \equiv 1 \pmod 4$, then

$$\mathcal{O}_K = \mathbb{Z}\left[\frac{1+\sqrt{D}}{2}\right] = \mathbb{Z} + \frac{1+\sqrt{D}}{2}\mathbb{Z}.$$

Proof. Let $\alpha = a+b\sqrt{D}$. We know $\alpha \in \mathcal{O}_K$ if and only if $2a, a^2 - Db^2 \in \mathbb{Z}$.

If $\alpha \in \mathcal{O}_K$, then $a \in \frac{1}{2}\mathbb{Z}$ and so $Db^2 \in \frac{1}{4}\mathbb{Z}$, which implies $b \in \frac{1}{2}\mathbb{Z}$ also. Therefore, all elements of $\mathcal{O}_K$ are of the form

$$\alpha = \frac{A + B\sqrt{D}}{2}, \quad A, B \in \mathbb{Z}$$

so we may restrict consideration only to α of this form. Then we have

$$\begin{aligned} \alpha \in \mathcal{O}_K &\iff A^2 - DB^2 \equiv 0 \pmod 4 \\ &\iff A^2 \equiv DB^2 \pmod 4 \end{aligned}$$

The solutions of this congruence depend on D modulo 4. Since the only squares modulo 4 are 0 and 1, we have

$$D \equiv 2, 3 \pmod 4 \implies A \equiv B \equiv 0 \pmod 2$$

and

$$D \equiv 1 \pmod 4 \implies A \equiv B \pmod 2.$$

In the former case, we conclude that

$$\mathcal{O}_K = \{a + b\sqrt{D} : a, b \in \mathbb{Z}\} = \mathbb{Z}[\sqrt{D}].$$

In the latter case, we can express α in the form

$$\alpha = \frac{A + B\sqrt{D}}{2} = \frac{A - B}{2} + B\frac{1 + \sqrt{D}}{2}.$$

Therefore,

$$\mathcal{O}_K = \{a + b\frac{1 + \sqrt{D}}{2} : a, b \in \mathbb{Z}\} = \mathbb{Z}\left[\frac{1 + \sqrt{D}}{2}\right].$$

□

What are the units of a quadratic field (i.e. units of its ring of integers)?

Proposition 46.3. *If $D < 0$ is squarefree, then*

(1) If $D = -1$, $\mathcal{O}_K^* = \{1, -1, i, -i\}$.
(2) If $D = 3$, $\mathcal{O}_K^* = \{1, -1, \omega, -\omega, \omega^2, -\omega^2\}$ for $\omega = \frac{1+\sqrt{-3}}{2}$.
(3) Otherwise, $\mathcal{O}_K^* = \{1, -1\}$.

Proof. Let $\alpha \in \mathcal{O}_K$. If $D \equiv 2, 3 \pmod 4$, then write $\alpha = A + B\sqrt{D}$, while if $D \equiv 1 \pmod 4$, write $\alpha = \frac{A+B\sqrt{D}}{2}$. Then the equation $N(\alpha) = \pm 1$ becomes, respectively for the two cases,

$$A^2 - DB^2 = \pm 1, \quad A^2 - DB^2 = \pm 4.$$

For $-D > 1$, the former has no solutions except $A = \pm 1, B = 0$. When $-D = 2$ or $-D > 3$, the latter has no solutions except $A = \pm 2$.

If $-D = 1$, then the former also has solutions $A = 0$, $B = \pm 1$. If $-D = 3$, then there are solutions

$$(A, B) = (\pm 2, 0), (\pm 1, \pm 1)$$

These give the units of the statement. □

To study the units of a real quadratic field, we will use the theory of continued fractions, later in the course.

47. Different and Discriminant

A bilinear pairing

$$\langle , \rangle : B \times B \to A$$

is called *non-singular* or *non-degenerate* if there is no element $b \in B$ such that $\langle b, B\rangle = 0$ (i.e. nothing that pairs to zero with everything).

Whenever you see a pairing, you should think of a dual space. Given an A-module B, the dual space B^* consists of all linear functionals on B, i.e. linear maps $B \to A$. Every element of b gives a linear functional via the pairing:

$$\langle b, \cdot\rangle : B \to A, \quad x \mapsto \langle b, x\rangle.$$

It would be nice if every linear functional arose in this manner. Unfortunately, that doesn't always happen.

We will be concerned in this section with a field extension L/K where K is the field of fractions of some integral domain A. Then we can define a pairing

$$\langle x, y\rangle = Tr_{L/K}(xy).$$

which makes L into a *metric K-vector space*. Since L is a field, for any x there is a y so that $xy = 1$; hence the pairing is non-degenerate. In this case, the Riesz Representation Theorem guarantees that every linear functional in the dual space is realized as 'pairing with an element,' as we hoped for above. Given a basis e_i for L, there is a natural basis for the dual space L^*, i.e. the linear functionals

$$f_i : \sum a_k e_k \mapsto a_i$$

which takes the value 1 on e_i and 0 on the other basis vectors. Therefore, by the Riesz Representation Theorem, there's an element $e_i^\vee$ of L such that

$$f_i(\ell) = \langle e_i^\vee, \ell\rangle.$$

Definition 47.1. *A* lattice Λ *in L is a free A-module of rank n. The* dual lattice $\Lambda^\vee$ *is defined by*

$$\Lambda^\vee = \{\alpha \in L : \langle \alpha, \Lambda\rangle \subset A\}.$$

In particular, $\alpha \in \Lambda^\vee$ if and only if $\langle \alpha, e_i\rangle \in A$ for the basis vectors e_i of Λ.

Example 47.2. *Let $\Lambda = \mathbb{Z}[i]$. Then $a + bi \in \mathbb{Q}(i)$ is in the dual lattice $\Lambda^\vee$ if and only if $Tr_{\mathbb{Q}(i)/\mathbb{Q}}((a+bi)\cdot 1) \in \mathbb{Z}$ and $Tr_{\mathbb{Q}(i)/\mathbb{Q}}((a+bi)\cdot i) \in \mathbb{Z}$. This happens if and only if $2a, -2b \in \mathbb{Z}$. Therefore,*

$$\Lambda^\vee = \frac{1}{2}\mathbb{Z} + \frac{i}{2}\mathbb{Z} = \frac{1}{2}\mathbb{Z}[i].$$

Proposition 47.3. *$\Lambda^\vee$ is a free rank n A-module with basis $e_i^\vee$.*

Proof. Claim: The $e_i^\vee$ are independent. Proof of claim: Suppose $\sum k_i e_i^\vee = 0$. Then $k_i = \langle 0, e_i \rangle = 0$ for all i.

Claim: $\Lambda^\vee = \sum Ae_i^\vee$. Proof of claim: From the first claim, we know the $e_i^\vee$ form a basis for L. Write $s = \sum k_i e_i^\vee$, $k_i \in K$. Then $s \in \Lambda^\vee$ if and only if $k_i = \langle s, e_i \rangle \in A$ for all i. □

Proposition 47.4. *Some properties of the dual:*

(1) $\Lambda^{\vee\vee} = \Lambda$
(2) $\Lambda_1 \subset \Lambda_2 \iff \Lambda_2^\vee \subset \Lambda_1^\vee$
(3) $(\Lambda_1 + \Lambda_2)^\vee = \Lambda_1^\vee \cap \Lambda_2^\vee$
(4) $(\Lambda_1 \cap \Lambda_2)^\vee = \Lambda_1^\vee + \Lambda_2^\vee$
(5) $(\alpha\Lambda)^\vee = \frac{1}{\alpha}\Lambda^\vee$ *for* $\alpha \neq 0 \in K$

Proof. Exercise. □

48. The Discriminant and Dual of $\mathcal{O}_K$ for Quadratic Fields

Let $K = \mathbb{Q}(\sqrt{d})$, d squarefree.

Theorem 48.1. *(1) If $d \equiv 2, 3 \pmod 4$, then $\Delta_K = 4d$ and $\mathcal{O}_K^\vee = \frac{1}{2\sqrt{d}}\mathcal{O}_K$.*
(2) If $d \equiv 1 \pmod 4$, then $\Delta_K = d$ and $\mathcal{O}_K^\vee = \frac{1}{\sqrt{d}}\mathcal{O}_K$.

Proof. Suppose $d \equiv 2, 3 \pmod 4$. Then, as we have seen previously, an integral basis is $\omega_1 = 1, \omega_2 = \sqrt{d}$, i.e. $\mathcal{O}_K = \mathbb{Z} + \sqrt{d}\mathbb{Z}$. Then

$$\Delta_K = \det(Tr(\omega_i\omega_j)) = \begin{vmatrix} 2 & 0 \\ 0 & 2d \end{vmatrix} = 4d.$$

To compute the dual, we find all elements $a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ having $Tr(a + b\sqrt{d}) = 2a \in \mathbb{Z}$ and $Tr(db + a\sqrt{d}) = 2bd \in \mathbb{Z}$, i.e. $\frac{1}{2}\mathbb{Z} + \frac{1}{2\sqrt{d}}\mathbb{Z} = \frac{1}{2\sqrt{d}}\mathcal{O}_K$. On the other hand, suppose $d \equiv 1 \pmod 4$. Then an integral basis is $\omega_1 = 1, \omega_2 = \frac{1+\sqrt{d}}{2}$. Then

$$\Delta_K = \det(Tr(\omega_i\omega_j)) = \begin{vmatrix} Tr(1) & Tr\left(\frac{1+\sqrt{d}}{2}\right) \\ Tr\left(\frac{1+\sqrt{d}}{2}\right) & Tr\left(\frac{1+d+2\sqrt{d}}{4}\right) \end{vmatrix} = \begin{vmatrix} 2 & -1 \\ -1 & \frac{1+d}{2} \end{vmatrix} = d.$$

To compute the dual, we find all elements $a + b\sqrt{d} \in \mathbb{Q}(\sqrt{d})$ having $Tr(a + b\sqrt{d}) = 2a \in \mathbb{Z}$ and $Tr((a + b\sqrt{d})(\frac{1}{2} + \frac{1}{2}\sqrt{d})) = Tr(\frac{a+db}{2} + \frac{a+b}{2}\sqrt{d}) = a + db \in \mathbb{Z}$, i.e. $\frac{1}{\sqrt{d}}\mathbb{Z} + (\frac{1}{2} + \frac{1}{2\sqrt{d}})\mathbb{Z} = \frac{1}{\sqrt{d}}\mathcal{O}_K$. □

49. Relation between dual and discriminant

We expect a relation between the dual and discriminant, which is easy to describe if we choose a basis e_i for L over K and write the pairing in terms of its Gram matrix,

$$G = (\langle e_i, e_j \rangle).$$

Then to pair any elements x and y in L, we write

$$\langle x, y \rangle = x^T G y.$$

By definition, if M is the matrix whose columsn are some basis u_i for a lattice Λ, and $M^\vee$ is the matrix whose columns are the dual basis $u_i^\vee$, then

$$M^T G M^\vee = I.$$

If we write this in terms of the basis $e_i = u_i$, this simplifies, and we discover that

$$M^\vee = G^{-1}.$$

To change the basis of L (respectively Λ) we are using to write the matrices involves a matrix B of determinant a unit in L (respectively A) and the following transformation:

$$M' = BM, \quad M'^\vee = BM^\vee, \quad G' = B^T G B.$$

So if we restrict to bases of Λ, then $\det G$ is determined free of basis up to the square of a unit in A. Further $(\det M^\vee)^{-1}$ is determined up to a unit, and they are equal up to a unit.

Conversely, if $\det G = \det G'$ for two bases of L, then it must be the case that $\det B = \pm 1$, and therefore the two bases generate the same A-module in L. From this we discover that

Proposition 49.1. *Any two bases of L over K have the same discriminant up to the square of a unit if and only if they generate the same lattice in L.*

Thus it makes sense to talk about the *discriminant of the ring of integers $\mathcal{O}_K$ of a number field K* by taking $\det G$ for a basis of $\mathcal{O}_K$ as a $\mathbb{Z}$-module. We often call this the *discriminant of K* or the *absolute discriminant of K*.

50. Computing Discriminants

The Gram matrix of the trace pairing can be diagonalized, provided we extend scalars. In particular, if $\sigma_1, \ldots, \sigma_n$ are the embeddings of

L fixing K, and if we extend scalars to some algebraically closed field, then we can write

$$(\sigma_i(x_j))\,(\sigma_j(x_k))^T = \left(\sum_j \sigma_j(x_i x_k)\right) = (Tr(x_i x_k))\,.$$

In other words, we diagonalize by changing basis from the standard basis in terms of $x_1, \ldots, x_n$ to

$$\begin{pmatrix}\sigma_1(x_1)\\ \sigma_2(x_1)\\ \vdots \\ \sigma_n(x_1)\end{pmatrix}, \begin{pmatrix}\sigma_1(x_2)\\ \sigma_2(x_2)\\ \vdots \\ \sigma_n(x_2)\end{pmatrix}, \ldots, \begin{pmatrix}\sigma_1(x_n)\\ \sigma_2(x_n)\\ \vdots \\ \sigma_n(x_n)\end{pmatrix}.$$

The pairing's Gram matrix becomes I_n.

For example, the Gram matrix for $\mathbb{Z}[i]$ was

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

Our change of basis matrix is

$$\begin{pmatrix}1 & 1\\ i & -i\end{pmatrix}$$

And the relation above becomes

$$\begin{pmatrix}1 & 1\\ i & -i\end{pmatrix}^T \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \begin{pmatrix}1 & 1\\ i & -i\end{pmatrix} = \begin{pmatrix}2 & 0\\ 0 & -2\end{pmatrix}$$

Or, in other words, the Gram matrix in terms of the new 'σ-basis' is:

$$\begin{pmatrix}1 & 1\\ i & -i\end{pmatrix}^{-T} \begin{pmatrix}2 & 0\\ 0 & -2\end{pmatrix} \begin{pmatrix}1 & 1\\ i & -i\end{pmatrix}^{-1} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Remark. Samuel uses a lemma of Dedekind on independence of embeddings to show that the Trace pairing is non-degenerate and so the Discriminant is non-zero. In our case (a field extension of the field of fractions of an integral domain), we have already seen this. He's working in more generality, where the trace pairing may not always be non-degenerate (can you come up with an example?).

At this point go back to Samuel for the remainder of Chapter 2.

51. But wait, so what's this 'Different', anyway?

For this we need to define invertible fractional ideals; it's the inverse of the dual of the ring of integers. More on this soon.

52. Factorisation of Ideals in Rings of Integers

Samuel; we did Chapters 2,3, and 1.4

53. Prime Ideals in Quadratic Fields

We'd like to essentially list the prime ideals of a quadratic field. First, it is not hard to show that the prime ideals of $\mathbb{Z}$ are exactly (p) where $p \in \mathbb{Z}$ is prime. (We never count the whole ring, i.e. unit ideal, as a prime ideal.) Now, instead consider $K = \mathbb{Q}(\sqrt{d})$. If $P \in \mathcal{O}_K$ is a prime ideal, then we have seen that there is some $a \in P$, $0 \neq a \in \mathbb{Z}$, so that $(a) \subset P$. Since 'to contain is to divide,' we just need to factor all ideals of the form (a) for non-zero integers a in order to find all primes P. Of course, we can do a little better: a itself has a prime factorisation,

$$a = \prod_i p_i$$

in $\mathbb{Z}$ which gives a corresponding equation for ideals:

$$(a) = \prod_i (p_i).$$

Of course, this may not be a prime factorisation in $\mathcal{O}_K$, as perhaps the (p_i) factor further. But we have now reduced the question to finding all the prime ideal factors of ideals of the form (p) where $p \in \mathbb{Z}$.

As an example, recall that we've already essentially done this for the Gaussian integers. For primes in $\mathbb{Z}$, we know how they factor in the UFD $\mathbb{Z}[i]$:

$p \equiv 1$	(mod 4)	$p = (a+bi)(a-bi)$
$p \equiv 3$	(mod 4)	p is prime
$p = 2$		$p = (1+i)(1-i)$

where $1+i$ and $1-i$ are associates, while $a+bi$ and $a-bi$ are not. These give rise to prime factorisations:

$p \equiv 1$	(mod 4)	$(p) = (a+bi)(a-bi)$
$p \equiv 3$	(mod 4)	(p) is prime
$p = 2$		$(p) = (1+i)^2$

It turns out that these three possible types of factorisation already demonstrate the possibilities.

Proposition 53.1. *Let K be a quadratic number field. Let p be a rational prime (i.e. prime in $\mathbb{Z}$). Then the ideal (p) in $\mathcal{O}_K$ factors in one of the following three ways:*

(1) (p) is prime;

(2) $(p) = P_1P_2$ *where* $P_1 \neq P_2$ *are distinct prime ideals;*
(3) $(p) = P^2$ *where* P *is a prime ideal.*

Proof. We know the norm of (p):

$$N((p)) = N(p) = p^2$$

Note that $N(I) = 1$ if and only if $\mathcal{O}_K/I$ is trivial if and only if $I = \mathcal{O}_K$. Since the norm is multiplicative on ideals, and prime ideals have norm $N(P) > 1$, it must be that (p) has at most two factors. □

We are ready to classify the factorisation of all ideals (p), where p is odd.

Theorem 53.2. *Let* $K = \mathbb{Q}(\sqrt{d})$ *where* d *is squarefree. Then the following hold:*

(1) If $p \nmid \Delta_K$, *and* $\left(\frac{d}{p}\right) = 1$, *then* $(p) = P_1P_2$ *where* $P_1 \neq P_2$ *are prime.*
(2) If $p \nmid \Delta_K$, *and* $\left(\frac{d}{p}\right) = -1$, *then* (p) *is prime.*
(3) If $p \mid \Delta_k$, *then* $(p) = P^2$ *where* P *is prime.*

Proof. First, suppose that $p \nmid \Delta_K$, and $\left(\frac{d}{p}\right) = 1$. Then let $a \in \mathbb{Z}$ be such that $a^2 \equiv d \pmod{p}$. Then

$$(p, a+\sqrt{d})(p, a-\sqrt{d}) = (p^2, p(a+\sqrt{d}), p(a-\sqrt{d}), a^2-d) = (p)(p, a+\sqrt{d}, a-\sqrt{d}, \frac{a^2-d}{p})$$

Note that $\frac{a^2-d}{p}$ is an integer because of our choice of a. Now, the right hand ideal contains p and $2a$, which are necessarily coprime, since $p \nmid d$ (this is a consequence of $p \nmid \Delta_K$). Hence the right hand ideal is the unit ideal, $(1) = \mathcal{O}_K$. Therefore,

$$(p, a + \sqrt{d})(p, a - \sqrt{d}) = (p)$$

We now need to check that these two ideals are proper and are not equal; by the previous proposition this suffices. But note that they are conjugates, meaning, conjugating all elements in one ideal gives all elements in the other ideal. Hence if one is the unit ideal, the other is; then their product would not be (p). Hence they are proper ideals. If they were equal, then they would both contain p and $2a$, and would again be the unit ideal. So they are different.

Second, suppose $p \nmid \Delta_K$. There is an element $\alpha \in \mathcal{O}_K$ such that $\alpha^2 = d$. (For, $\sqrt{d} \in \mathcal{O}_K$.) Let's suppose P is a prime ideal dividing (p). Suppose $N(P) = p$. Then we have maps

$$\mathcal{O}_K \to \mathcal{O}_K/(p) \to \mathcal{O}_K/P \to \mathbb{Z}/p\mathbb{Z}$$

The last of these maps is an isomorphism, since $N(P) = p$, and there's only one right of size p. Integers $a \in \mathbb{Z} \subset \mathcal{O}_K$ must map under this sequence of maps to $a \pmod p$. The element α must map to something, call it $\phi(\alpha)$. But since $\alpha^2 = d$ in $\mathcal{O}_K$, it must be that

$$\phi(\alpha)^2 \equiv d \pmod p$$

in $\mathbb{Z}/p\mathbb{Z}$. This implies $\left(\frac{d}{p}\right) = 1$. So whenever (p) is not prime, $\left(\frac{d}{p}\right) = 1$; the contrapositive gives case two.

Finally, suppose $p \mid \Delta_k$. Then $p \mid d$ (since p is odd). So we have

$$(p, \sqrt{d})^2 = (p^2, p\sqrt{d}, d) = (p)(p, \sqrt{d}, d/p) = (p)\mathcal{O}_K = (p)$$

because d/p and p are coprime integers (a consequence of the fact that d is squarefree). □

Hence, we have classified the primes for $\mathbb{Q}(\sqrt{d})$ in the case that d is odd, except for those primes that are factors of (2). We have:

(1) (p) whenever $p \nmid \Delta_K$ and $\left(\frac{d}{p}\right) = -1$
(2) $(p, a \pm \sqrt{d})$ whenever $p \nmid \Delta_K$ and $d \equiv a^2 \pmod p$
(3) $(p, \sqrt{d})$ when $p \mid \Delta_K$

We also have a result for the prime $p = 2$:

Theorem 53.3.
If $2 \nmid \Delta_K$, and $d \equiv 1 \pmod 8$, then $(2) = P_1P_2$ where $P_1 \neq P_2$ are prime.
If $2 \nmid \Delta_K$, and $d \equiv 5 \pmod 8$, then (2) is prime.
If $2 \mid \Delta_k$, then $(2) = P^2$ where P is a prime.

54. The Correspondence between Ideals and Quadratic Forms

Let $K = \mathbb{Q}(\sqrt{-d})$ be a quadratic field of discriminant Δ and suppose $d > 0$ (so it is an imaginary quadratic field). Define two sets:

$$\mathcal{I} = \{\text{equivalence classes of ideals of } \mathcal{O}_K\},$$

$$\mathcal{Q} = \left\{\begin{matrix}\text{proper equivalence classes of primitive positive} \\ \text{definite quadratic forms of discriminant } \Delta\end{matrix}\right\}.$$

First we need a definition.

Definition 54.1. *For any ideal $I \subset \mathcal{O}_K \subset \mathbb{C}$, an integral basis α, β for I is called* admissible *if $\mathrm{Im}(\frac{\beta}{\alpha}) > 0$. (Note that this depends on a choice of embedding $\mathcal{O}_K \subset \mathbb{C}$.)*

Our goal is to show that the two sets $\mathcal{I}$ and $\mathcal{Q}$ are in bijection. We'll define a function

$$\mathbf{Q} : \mathcal{I} \to \mathcal{Q}$$

as follows. Let I be a representative of an equivalence class $[I] \in \mathcal{I}$, and let α, β be an admissible integral basis for I. Then

$$\mathbf{Q}([I]) = \left[\frac{N(\alpha x + \beta y)}{N(I)} \right].$$

where the right hand side is a quadratic form in variables x, y. We can also define

$$\mathbf{I} : \mathcal{Q} \to \mathcal{I}$$

as follows. Let $f = ax^2 + bxy + cy^2$ be a representative of a proper equivalence class $[f] \in \mathcal{Q}$. Then

$$\mathbf{I}([f]) = \left[\left(a, \frac{b + \sqrt{\Delta}}{2} \right) \right].$$

We'll need to show that these two functions are well-defined and supply a bijection between $\mathcal{I}$ and $\mathcal{Q}$. But first let's look at our continuing example from the midterm:

Example 54.2. *Let $K = \mathbb{Q}(\sqrt{-5})$ with ring of integers $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ and discriminant $\Delta = -20$. (Note that $-5 \equiv 3 \pmod 4$).*

Then the ideal $(2, 1 + \sqrt{-5})$ has integral basis $2, 1 + \sqrt{-5}$ and so maps to

$$\frac{N(2x + y(1 + \sqrt{-5}))}{N(I)} = \frac{1}{2}((2x + y)^2 + 5y^2) = 2x^2 + 2xy + 3y^2,$$

and this quadratic form yields the ideal $(2, 1 + \sqrt{-5})$. (Of course, we can't always expect to get back the same ideal, but in general one which is equivalent.)

A principal ideal (γ) can be taken to have integral basis γ, $\gamma\sqrt{-5}$, so it maps to

$$\frac{N(\gamma(x + y\sqrt{-5}))}{N(\gamma)} = N(x + y\sqrt{-5}) = x^2 + 5y^2.$$

This quadratic form maps to the ideal $(1, \sqrt{-5}) = (1)$, which is equivalent to any principal ideal.

Lemma 54.3. *If α, β and α', β' are both admissible integral bases for I, then there exists M with $\det(M) = 1$ such that*

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = M \begin{pmatrix} \alpha' \\ \beta' \end{pmatrix}.$$

Proof. First note that

$$\frac{\beta}{\alpha} = \frac{\beta\overline{\alpha}}{\alpha\overline{\alpha}} = \frac{\beta\overline{\alpha}}{N(\alpha)}.$$

The denominator of the right-hand quantity is positive, hence a basis α, β is admissible if and only if $\mathrm{Im}(\beta\overline{\alpha}) > 0$.

We can write $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, so that

$$\alpha = a\alpha' + b\beta', \qquad \beta = c\alpha' + d\beta'.$$

And we have $\mathrm{Im}(\beta\overline{\alpha}) > 0$ and $\mathrm{Im}(\beta'\overline{\alpha'}) > 0$.

$$\begin{aligned} \beta\overline{\alpha} &= (c\alpha' + d\beta')(a\overline{\alpha'} + b\overline{\beta'}) \\ &= acN(\alpha') + ad\beta'\overline{\alpha'} + cb\alpha'\overline{\beta'} + bdN(\beta') \end{aligned}$$

Since it is only the imaginary part we are concerned with, we can subtract various real parts (norms and traces) and we are left to consider

$$(ad - bc)(\beta'\overline{\alpha'})$$

By the assumption that both bases are admissible, we find that $ad - bc > 0$. Since they are both bases, $ad - bc = 1$. □

Proposition 54.4. Q *is well-defined on ideal classes in* $\mathcal{I}$.

Proof. Let α', β' and γ', δ' be two admissible integral bases of I and I', where $I \sim I'$.

Then there are some θ, $\phi \in \mathcal{O}_K$ such that

$$(\phi)I = (\theta)I'.$$

So we can write

$$\alpha = \phi\alpha', \beta = \phi\beta',$$

which is an integral basis of $(\phi)I$; and

$$\gamma = \theta\gamma', \delta = \theta\delta',$$

which is an integral basis of $(\theta)I$.

These two bases are related by a change-of-basis matrix $M \in SL_2(\mathbb{Z})$ (by the lemma, since they are both admissible). In fact, this implies that if we write

$$x\alpha + y\beta = X\gamma + Y\delta$$

then

$$\begin{pmatrix} \gamma & \delta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \alpha & \beta \end{pmatrix} M \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

so that the two quadratic forms

$$\frac{N(\alpha x+\beta y)}{N(\phi)N(I)} \quad \text{and} \quad \frac{N(\gamma X+\delta Y)}{N(\theta)N(I')}$$

are related by a linear transformation of determinant 1, taking (x,y) to (X,Y). Therefore the choice of ideal within an equivalence class and the choice of admissible basis does not alter the proper equivalence class of the resulting quadratic form.

□

Proposition 54.5. **I** *is well-defined on proper equivalence classes in* $\mathcal{Q}$.

Proof. Let $ax^2+bxy+cy^2$ be a representative of a proper equivalence class of forms. We will denote forms by a triple of their coefficients. Then

$$\mathbf{I}([(a,b,c)]) = \left(a, \frac{b+\sqrt{\Delta}}{2}\right).$$

We will apply $S,T \in \mathrm{SL}_2(\mathbb{Z})$ to f and verify that the ideal class obtained is the same. First T:

$$\mathbf{I}([T(a,b,c)]) = \mathbf{I}([(a,b+2a,a+b+c)]) = \left(a, a+\frac{b+\sqrt{\Delta}}{2}\right) = \left(a, \frac{b+\sqrt{\Delta}}{2}\right).$$

Now S, which is more difficult. We have

$$\mathbf{I}([S(a,b,c)]) = \mathbf{I}([(c,-b,a)]) = \left(c, \frac{-b+\sqrt{\Delta}}{2}\right).$$

Note that since $\Delta = b^2-4ac$,

$$\left(\frac{b+\sqrt{\Delta}}{2}\right)\left(\frac{-b+\sqrt{\Delta}}{2}\right) = ac$$

Therefore,

$$(c)\left(a, \frac{b+\sqrt{\Delta}}{2}\right) = \left(ac, c\frac{b+\sqrt{\Delta}}{2}\right)$$

and

$$\left(\frac{b+\sqrt{\Delta}}{2}\right)\left(c, \frac{-b+\sqrt{\Delta}}{2}\right) = \left(c\frac{b+\sqrt{\Delta}}{2}, ac\right)$$

So the resulting ideals are equivalent. □

Proposition 54.6. *Let $I \subset \mathcal{O}_K$ be an ideal. Then*

$$\Delta_I = N(I)^2 \Delta_K.$$

Proof. Let α, β be an integral basis for the ideal I, and let θ, ϕ be an integral basis for $\mathcal{O}_K$. Then these bases are related by a change-of-basis matrix M. By the same proof we saw in homework for principal ideals, $\det(M) = N(I)$. Then, by a property of determinants which we've seen earlier,

$$\Delta_I = N(I)^2 \Delta_K.$$

□

Proposition 54.7. *If $[I] \in \mathcal{I}$, and $[f] \in \mathcal{Q}$, then*

(1) $\mathbf{Q}([I]) \in \mathcal{Q}$,
(2) $\mathbf{I}([f]) \in \mathcal{I}$.

Proof. For the first part, let $\overline{\alpha}$ and $\overline{\beta}$ represent the conjugates of α and β. Then

$$\begin{aligned} N(\alpha x + \beta y) &= (\alpha x + \beta y)(\overline{\alpha} x + \overline{\beta} y) \\ &= N(\alpha)x^2 + Tr(\alpha\overline{\beta})xy + N(\beta)y^2 \end{aligned}$$

This has discriminant

$$Tr(\alpha\overline{\beta}) - 4N(\alpha)N(\beta) = (\alpha\overline{\beta} + \overline{\alpha}\beta)^2 - 4\alpha\overline{\alpha}\beta\overline{\beta} = (\alpha\overline{\beta} - \overline{\alpha}\beta)^2.$$

On the other hand,

$$\begin{aligned} \Delta_I &= Tr(\alpha^2)Tr(\beta^2) - Tr(\alpha\beta)^2 \\ &= (\alpha^2 + \overline{\alpha}^2)(\beta^2 + \overline{\beta}^2) - (\alpha\beta + \overline{\alpha}\overline{\beta})^2 \\ &= (\overline{\alpha}\beta - \alpha\overline{\beta})^2 \end{aligned}$$

Therefore,

$$\frac{N(\alpha x + \beta y)}{N(I)}$$

has discriminant

$$\Delta_I / N(I)^2 = \Delta.$$

For the second part, begin with a form $ax^2 + bxy + cy^2$. Then $a \in \mathcal{O}_K$ since $\mathbb{Z} \subset \mathcal{O}_K$, and $\frac{b-\Delta}{2} \in \mathcal{O}_K$ because

(1) if $4 \mid \Delta$, then $\mathcal{O}_K = \mathbb{Z}[\sqrt{\Delta/4}]$; whereas
(2) if $4 \nmid \Delta$, then $\mathcal{O}_K = \mathbb{Z}[\frac{1+\sqrt{\Delta}}{2}]$.

So

$$\left(a, \frac{b-\Delta}{2}\right)$$

is an ideal of $\mathcal{O}_K$. □

Theorem 54.8. *The two sets $\mathcal{I}$ and $\mathcal{Q}$ are in bijection under* $\mathbf{Q}$ *and* $\mathbf{I}$.

Proof. We will show that $\mathbf{Q} \circ \mathbf{I}$ and $\mathbf{I} \circ \mathbf{Q}$ are the identity maps on $\mathcal{Q}$ and $\mathcal{I}$, respectively. Let α, β be an admissible integral basis for I. Then

$$\mathbf{Q}([I]) = \frac{1}{N(I)} \left(N(\alpha)x^2 + Tr(\alpha\overline{\beta})xy + N(\beta)y^2 \right).$$

Applying $\mathbf{I}$, we obtain

$$\left(\frac{N(\alpha)}{N(I)}, \frac{\frac{Tr(\alpha\overline{\beta})}{N(I)} + \sqrt{\Delta}}{2} \right).$$

Our task is to show that

$$\left(\frac{N(\alpha)}{N(I)}, \frac{\frac{Tr(\alpha\overline{\beta})}{N(I)} + \sqrt{\Delta}}{2} \right) \sim (\alpha, \beta).$$

Let $1, \theta$ be an integral basis for $\mathcal{O}_K$, where $\theta = \sqrt{d}$ or $\frac{1+\sqrt{d}}{2}$ as appropriate (depending on d modulo 4). Then write $X_2 : \mathcal{O}_K \to \mathbb{Z}$ for the map which takes α to its second coordinate in terms of the basis $1, \theta$.

Claim: $N(I) = X_2(\overline{\alpha}\beta)$ **Proof of Claim:** Write

$$\alpha = \alpha_1 + \alpha_2\theta, \quad \beta = \beta_1 + \beta_2\theta, \quad \alpha_i, \beta_i \in \mathbb{Z}.$$

Then

$$N(I) = |\alpha_1\beta_2 - \alpha_2\beta_1| = |X_2(\overline{\alpha}\beta)|$$

But by admissibility, $\mathrm{Im}(\overline{\alpha}\beta) > 0$, so that $X_2(\overline{\alpha}\beta) > 0$. End proof of claim.

So we have

$$(X_2(\overline{\alpha}\beta)) \left(\frac{N(\alpha)}{N(I)}, \frac{\frac{Tr(\alpha\overline{\beta})}{N(I)} + \sqrt{\Delta}}{2} \right) = \left(N(\alpha), \frac{Tr(\overline{\alpha}\beta) + X_2(\overline{\alpha}\beta)\sqrt{\Delta}}{2} \right).$$

Now, if $\Delta = -4d$, then $X_2(\cdot) = \mathrm{Im}(\cdot)/\sqrt{d}$. If $\Delta = -d$, then $X_2(\cdot) = 2\,\mathrm{Im}(\cdot)/\sqrt{d}$. In either case, the second generator of the ideal above becomes

$$\frac{Tr(\overline{\alpha}\beta) + 2\,\mathrm{Im}(\overline{\alpha}\beta)}{2} = \overline{\alpha}\beta.$$

So the ideal above becomes $(\overline{\alpha}\alpha, \overline{\alpha}\beta) \sim (\alpha, \beta)$.

Fortunately, the other direction is easier. Begin with a form $ax^2 + bxy + cy^2$. This gives an ideal

$$I = \left(a, \frac{b + \sqrt{\Delta}}{2} \right).$$

In turn, this gives a form

$$\frac{1}{N(I)}\left(a^2x^2 + Tr\left(a\frac{b-\sqrt{\Delta}}{2}\right)xy + acy^2\right),$$

because

$$\frac{b+\sqrt{\Delta}}{2}\cdot\frac{b-\sqrt{\Delta}}{2} = \frac{b^2-\Delta}{4} = ac.$$

We also have

$$N(I) = X_2(a(b+\sqrt{\Delta})/2) = a$$

so that this form is actually

$$ax^2 + bxy + cy^2$$

as required. □

55. Diophantine Approximation

Theorem 55.1 (Dirichlet's Theorem)**.** *For all $\alpha \in \mathbb{R}$, and $1 < Q \in \mathbb{R}$, there exist $p, q \in \mathbb{Z}$ such that $0 < q < Q$ and*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{qQ}.$$

Proof. First, suppose that Q is an integer. Then $Q \geq 2$. Consider the $Q+1$ real numbers

$$0, 1, \{\alpha\}, \{2\alpha\}, \ldots, \{(Q-1)\alpha\}.$$

Divide the line segment $[0,1]$ into Q disjoint segments of length $1/Q$ in the obvious way. Then at least one of these segments must contain two numbers from the list above. Since $Q \geq 2$, 0 and 1 are not in the same segment. So there are two cases:

First, we may have $\{r_1\alpha\}$ and $\{r_2\alpha\}$ in the same segment, with $r_1 < r_2$. Then there exist $s_1, s_2 \in \mathbb{Z}$ with

$$|(r_1\alpha - s_1) - (r_2\alpha - s_2)| \leq \frac{1}{Q}.$$

On the other hand, we may have 0 or 1 together with $\{r\alpha\}$ in a segment. In this case, there exists an $s \in \mathbb{Z}$ with

$$|r\alpha - s| < \frac{1}{Q}.$$

In either case, we are done.

Now suppose that Q is not an integer. Then just take the smallest integer $Q_0 \geq Q$. Since $q < Q_0$, then $q < Q$ (since $q \in \mathbb{Z}$). □

Note that we can replace $\leq$ with $<$ if α is irrational. In fact, the only time we will get $\leq$ is if two extreme endpoints of the interval are used, i.e. α is rational with denominator Q.

Corollary 55.2. *Let α be a real number. Then α is irrational if and only if there exist infinitely many rational numbers $p/q \in \mathbb{Q}$ such that*

$$0 < \left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$$

Proof. Choose any Q_1 and apply Dirichlet to obtain a $p_1/q_1 \in \mathbb{Q}$ such that

$$\left|\alpha - \frac{p_1}{q_1}\right| \leq \frac{1}{q_1 Q_1} \leq \frac{1}{q_1^2}$$

Then choose Q_2 such that

$$\left|\alpha - \frac{p_1}{q_1}\right| > \frac{1}{q_1 Q_2}.$$

Apply Dirichlet's theorem again with this Q_2. Etc. etc.

However, if α is rational, say $\alpha = \frac{r}{s}$, then for any $p/q \in \mathbb{Q}$ with $q > s$,

$$\left|\frac{p}{q} - \frac{r}{s}\right| \geq \frac{1}{qs} > \frac{1}{q^2}$$

So there are only finitely many rational approximations to a rational number.

□

This characterises irrationality! So it is possible to prove a number is irrational in this way. Do you remember Liouville's number, which we showed was transcendental? It was

$$\sum_k \frac{1}{10^{k!}}$$

You could show this is irrational using this theorem. Similar methods can be used to show that $\zeta(3)$ is irrational (Apery, 1978).

Theorem 55.3 (Liouville's Theorem)**.** *Let $\alpha \in \mathbb{R}$ be an algebraic number of degree $d > 1$. Then for all $\epsilon > 0$, there are only finitely many $p, q \in \mathbb{Z}$ such that*

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^{d+\epsilon}}$$

Proof. Let $f(x) \in \mathbb{Q}[x]$ be the minimal polynomial of α. Multiplying up to remove denominators, α is a root of some

$$g(x) = a_d x^d + \cdots + a_0 \in \mathbb{Z}[x]$$

Then

$$\left|g\left(\frac{p}{q}\right)\right| = \frac{1}{q^d}\left|a_d p^d + \cdots + a_0 q^d\right| \geq \frac{1}{q^d}.$$

since $g(x)$ has no rational roots, and hence $a_d p^d + \cdots + a_0 q^d$ is a non-zero rational integer.

Now apply the mean value theorem:

$$g(\alpha) - g\left(\frac{p}{q}\right) = \left(\alpha - \frac{p}{q}\right) g'(\beta)$$

for some β in the interval $[\alpha, \frac{p}{q}]$.

Since $g(\alpha) = 0$ but $g'(\beta) \neq 0$ (since left side of the last equation doesn't vanish), we get

$$\left|\alpha - \frac{p}{q}\right| = \frac{\left|g\left(\frac{p}{q}\right)\right|}{|g'(\beta)|} \geq \frac{1}{|g'(\beta)|q^d} > \frac{1}{Mq^d}$$

for some constant M independent of p and q (the fact that solutions p, q to the inequality of the theorem satisfy $\left|\alpha - \frac{p}{q}\right| < 1$ means $\beta \in [\alpha - 1, \alpha + 1]$, wherein g' has some maximum). Any potential solution to the inequality of the theorem having large enough q so that $q^\epsilon > M$ is ruled out by this last inequality; hence there are only finitely many solutions. □

Note that the proof gives a bound to the size of solutions!

Since Liouville's theorem (1844), there has been gradual progress in improving the result, i.e. giving smaller exponents on q under the same hypotheses.

Liouville (1844): exponent d
Thue (1909): exponent $\frac{d}{2} + 1$
Siegel (1921): exponent $2\sqrt{d}$
Gelfand/Dyson (1947): exponent $\sqrt{2d}$
Roth (1955): exponent 2

Roth's result, that for every algebraic number α and for every $\epsilon > 0$, we have only finitely many solutions $p, q \in \mathbb{Z}$ to

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^{2+\epsilon}}$$

earned him the Field's Medal, and has wide-ranging number theoretical consequences. Compare it to Dirichlet's Corollary that α is rational

(i.e. algebraic of degree 1) if and only if

$$\left|\alpha - \frac{p}{q}\right| \leq \frac{1}{q^2}$$

has only finitely many solutions.

In other words, algebraic numbers are poorly approximable, but rationals are even more poorly approximable.

Note: both Dirichlet's Corollary and Liouville's proof rely fundamentally on the fact that if some polynomial vanishes at α, but not at p/q, then its value at p/q is bounded below somehow (using the fact that integers are discretely separated). In the case of Dirichlet the polynomial is just linear. This fundamental idea is at root of Roth's proof, too.

Liouville's Theorem implies that Liouville's number is transcendental. We saw this before, but now it's a quick consequence:

Theorem 55.4. *The number*

$$z = \sum_{n=1}^{\infty} \frac{1}{10^{i!}}$$

is transcendental.

Proof. Let d, t be integers. Let p/q be the partial sum

$$p/q = \sum_{i=1}^{d+t} \frac{1}{10^{i!}}$$

Then the denominator $q = 10^{(d+t)!}$. We have

$$\left|\alpha - \frac{p}{q}\right| = \frac{1}{10^{(d+t+1)!}}\left(1 + \frac{1}{10^{d+t+2}} + \cdots\right) < \frac{2}{10^{(d+t+1)!}} < \frac{2}{q^{d+t+1}} < \frac{1}{q^{d+t}}.$$

This works for all t, hence we find infinitely many 'good approximations' for any d. □

56. Continued Fractions

Given a real number $\alpha > 0$, if it is > 1, then subtract 1 and if it is < 1, then invert it. This process expresses any α as a *continued fraction*:

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \cdots}}}$$

In other words, given α we obtain a well-defined sequence of integers a_n: a_0 is defined as the integer part of α, and then a_1 is the integer part of $(\alpha - a_0)^{-1}$ etc.

Every real number has a continued fraction expansion. A few famous continued fraction expansions are:

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{4+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{6+\cdots}}}}}}}}$$

which has the pattern $2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, \ldots$, and

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15+\cfrac{1}{1+\cfrac{1}{292+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{2+\cdots}}}}}}}}$$

which has no discernable pattern. Chopping off the fraction at any finite point, we obtain a rational approximations to α. Are these good in the sense of Dirichlet's Theorem? That's our question for now.

The processes of subtracting 1 and inverting can be codified with matrices. Think of a fraction as a vector with numerator and denominator as its entries. Two vectors which are scalar multiples of each other represent the same fraction. Thus, we think of such vectors as 'equivalent' and write

$$\begin{pmatrix} r \\ s \end{pmatrix} \sim a \begin{pmatrix} r \\ s \end{pmatrix}$$

The space of all vectors in $\mathbb{R}^2$ modulo this equivalence is called *real projective space*, $\mathbb{P}^1(\mathbb{R})$. The matrices in $\mathrm{GL}_2(\mathbb{R})$, i.e. those with determinant ± 1, act on $\mathbb{P}^1(\mathbb{R})$.

For example, the action of subtracting one from a number, i.e.

$$r/s \mapsto r/s - 1,$$

can be codified in this language as multiplication by a matrix:

$$\begin{pmatrix} r \\ s \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} r+s \\ s \end{pmatrix}.$$

Notice that if we use a different, but equivalent, vector to represent r/s, we get an equivalent answer:

$$\begin{pmatrix} r \\ s \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r/s \\ 1 \end{pmatrix} = \begin{pmatrix} r/s+1 \\ 1 \end{pmatrix}.$$

Inverting a number, i.e.

$$r/s \mapsto 1/(r/s) = s/r,$$

can also be codified as multiplication by a matrix:

$$\begin{pmatrix} r \\ s \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} s \\ r \end{pmatrix}.$$

Since matrices which are multiples of one another act the same way on equivalence classes of vectors, we can think of this as an action of $\mathrm{PGL}_2(\mathbb{R})$.

Given $\alpha > 0$, write p_n/q_n for the continued fraction of α truncated after n steps, i.e. if α has continued fraction expansion

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

which we will henceforth write in a more compact notation as $[a_0; a_1, a_2, a_3, a_4, \ldots]$, then

$$p_n/q_n = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

In particular, $p_0/q_0 = a_0$ and $p_1/q_1 = a_0 + 1/a_1$.

Definition 56.1. *The quantity p_n/q_n is called the n-th* approximant *or* convergent *of α and a_n is called the n-th* partial quotient *of α.*

If we start with $\infty = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, we can invert and add a_n to get a_n, then invert and add a_{n-1} to get $a_{n-1} + \frac{1}{a_n}$ and so on, building up the continued fraction from below to obtain p_n/q_n once we finally reach a_0.

We could even start with $0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, which, when inverted and adding a_n gives us ∞. Then, continuing, we build up p_{n-1}/q_{n-1} (since a_n went missing).

In other words, in the matrix notation above,

$$\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_3 & 1 \\ 1 & 0 \end{pmatrix} \cdots \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}$$

We obtain recurrence relations

$$p_{n+1} = a_{n+1}p_n + p_{n-1}, \quad q_{n+1} = a_{n+1}q_n + q_{n-1},$$

meaning the ps and qs are very easy to compute from the a_n.

Considering determinants, we find that for $n > 1$,

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

This implies p_n and p_{n+1} are relatively prime, as are q_n and q_{n+1}.

As a consequence,

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

So we obtain an alternating series whose partial sums are p_n/q_n:

$$\frac{p_n}{q_n} = a_0 + \sum_{i=1}^{n} \frac{(-1)^{i+1}}{q_i q_{i-1}}.$$

Do we have a limit

$$\lim_{n\to\infty} \frac{p_n}{q_n}?$$

If $\alpha > 0$, then the $a_n \geq 0$. If α is such that the a_n are strictly positive, then the q_n are growing as $n \to \infty$, by the recurrence relations. By the alternating series test, then, the limit exists.

Now, given a_0, a_1 etc.,

$$\alpha = [a_0, a_1, \ldots, a_{n-1}, \alpha_n]$$

where α_n is not necessarily rational, but is chosen so that this *finite* continued fraction is equal to α. In other words,

$$\alpha_n = [a_n, a_{n+1}, \ldots].$$

Note that a_n is the integer part of α_n. In particular, $a_n \leq \alpha_n < a_n + 1$. Then, from the recurrence relations above,

$$\alpha = [a_0, a_1, \ldots, a_{n-1}, \alpha_n] = \frac{\alpha_n p_{n-1} + p_{n-2}}{\alpha_n q_{n-1} + q_{n-2}}$$

Lemma 56.2. *Suppose $a, b, c, d > 0$. If $\frac{a}{b} < \frac{c}{d}$, then $\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$.*

Proof. The hypothesis guarantees that $ad < bc$. Multiply up denominators to compare three integers, and use this inequality. □

From the lemma, then, α lies between $\frac{p_{n-1}}{q_{n-1}}$ and $\frac{p_{n-2}}{q_{n-2}}$. This may be shown for all n. Hence, by the convergence of the alternating series, $\alpha = \lim_{n\to\infty} \frac{p_n}{q_n}$.

Assuming that the a_n are positive, then from a property of the partial sums of an alternating series, we have the following arrangement of the convergents:

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \alpha < \cdots < \frac{p_{2m+1}}{q_{2m+1}} < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}.$$

Furthermore, also a property of alternating series,

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

So, the $\frac{p_n}{q_n}$ form a sequence of infinitely many good approximations of α, and by Dirichlet's Corollary, α is irrational.

On the other hand, if $a_n = 0$ at some point, our continued fraction expansion terminates and clearly α is rational.

We have shown

Theorem 56.3. *A real number $\alpha > 0$ is irrational if and only if its continued fraction expansion is infinite. Furthermore, the partial quotients $\frac{p_n}{q_n}$ are all 'good approximations' in the sense of Dirichlet.*

Of course, this leaves us with the question – are these all the good approximations? Or the best ones in some sense?

57. Pell's Equation

Pell's Equation (erroneously attributed to Pell by Euler; it should be Brouncker's equation), is the following, for d not a square:

$$x^2 - dy^2 = \pm 1$$

From the theory of quadratic fields, we see that this is the equation $N(x + y\sqrt{d}) = \pm 1$, i.e. the solutions give all units $x + y\sqrt{d}$ in $\mathbb{Q}(\sqrt{d})$ for $d \equiv 2, 3 \pmod 4$. The equation has only finitely many solutions for negative d. But for positive d, it may have many more. In this section we'll use continued fractions to describe the solutions. We will assume throughout that d is positive and not a square.

A solution to the Pell equation must satisfy

$$1 = \left|x^2 - dy^2\right| = \left|x - \sqrt{d}y\right| \left|x + \sqrt{d}y\right|$$

so that

$$\left|\frac{x}{y} - \sqrt{d}\right| < \frac{1}{y^2 \left|\frac{x}{y} + \sqrt{d}\right|} < \frac{1}{2y^2}$$

The final inequality follows since

$$\left|\frac{x}{y} + \sqrt{d}\right| = \sqrt{d \pm \frac{1}{y^2}} + \sqrt{d} > 2.$$

This shows that $\frac{x}{y}$ is a convergent to $\sqrt{d}$, in fact it is a p_n/q_n for some n.

This gives the basic relationship; we need to ask now, is every convergent to $\sqrt{d}$ a solution? Not necessarily, but we can use continued fractions to provide a general solution anyway. There are close ties to the units in the ring of integers of a real quadratic field.

58. Elliptic Curves

In what follows, I'll largely be simply expanding in these notes on *Lectures on Elliptic Curves* by J.W.S Cassels (LMS Student Texts).

An elliptic curve E over a field K is a non-singular plane curve defined by an equation of the form

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

where $a_i \in K$. Henceforth we will assume that most of these coefficients are zero and write instead

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

where $a, b \in K$. (Unless we are in characteristic $p = 2, 3$, one can always perform a change of variables to obtain this simpler form.) The equation is called a *Weierstrass equation*. To check that it is non-singular is to check that the quantity

$$\Delta = -16(4a^3 + 27b^2)$$

is non-zero. This is the *discriminant* of the curve. The quantity $4a^3 + 27b^2$ is the discriminant of the cubic polynomial on the right; we've adjusted it by a factor of -16.

The K-rational points of E are the points $(x, y) \in K$ satisfying the equation of E. The points of an elliptic curve form a group under the following group law:

To add P_1 and P_2, draw a straight line through P_1 and P_2. This will intersect the curve at three points. Call the third P'. Then draw a vertical line through P'. This will intersect the curve at two points. Call the second $P_1 + P_2$; that is the sum.

This sounds a little strange, but actually the group law can be characterised by the following fact: any line will intersect the curve at three points (over the algebraic closure). These three points sum to the identity.

To see that one always obtains *three* points, we need to count correctly, and that means considering the curve as a curve in the projective plane $\mathbb{P}^2$. Let K be a field.

$$\mathbb{P}^2(K) = \{[X, Y, Z] : X, Y, Z \in K \text{ not all zero}\}/\sim$$

where $\sim$ denotes the following equivalence relation:

$$[X, Y, Z] \sim [X', Y', Z'] \iff \exists a \neq 0 s.t. X = aX', Y = aY', Z = aZ'.$$

We can identify $[X, Y, Z]$ with the line along the vector (X, Y, Z) in K^3. In this way, $\mathbb{P}^2(K)$ becomes the collection of lines through the origin in K^3.

We now rewrite the Weierstrass equation by homogenizing:

$$ZY^2 = X^3 + aZ^2X + bZ^3$$

This can now be considered an equation in $\mathbb{P}^2$, since any two points under the equivalence $\sim$ either both satisfy or neither satisfy the equation. The K-points of the usual equation (called *affine*) are points of the new one under the embedding

$$(x, y) \mapsto [x, y, 1].$$

The homogeneous equation has a single additional point, which is

$$[0, 1, 0]$$

and we call this *the point at infinity*. There are no other additional points, since if the third coordinate is non-zero, it is of the form $[x, y, 1]$ and if the third coordinate is zero, then the equation dictates that the first coordinate is zero.

Now suppose we intersect the equation with a line, say $\alpha X + \beta Y + \gamma Z = 0$. If γ is non-zero, then solving for Z and substituting, we obtain a degree 3 homogeneous equation in two variables X and Y, with coefficient 1 on X^3. Since X and Y are in proportion, we can divide through by Y^3, so that $X^3 + \cdots + zY^3$ has solutions $X = x_0Y$ where x_0 is a root of $X^3 + \cdots + z$. Thus, as a projective equation, this has exactly three solutions over an algebraically closed field, counted with multiplicity. Note that a vertical line (one having $\beta = 0$) intersects the line at infinity (the line $Z = 0$) at the one point at infinity.

If two of these points are defined over a field K, then the third is also, since in this case the polynomial factors over that field.

If $\gamma = 0$, then solve for Y instead for the same conclusion. If both β and γ are zero, then the line is $X = 0$, which intersects the curve at the solutions to $ZY^2 = bZ^3$, i.e. $[0, 1, 0]$, $[0, \pm\sqrt{b}, 1]$.

Finally, it's possible we want to add the same point to itself. But that's just using multiplicities; the line goes through one point with multiplicity two, i.e. it is tangent.

Thus we have shown that whenever two K-rational points are chosen, their sum is well-defined and is a K-rational point.

Theorem 58.1. *The K-rational points of E, denoted $E(K)$, form an abelian group.*

Proof. The product is well-defined and again in $E(K)$.

The group law is clearly abelian.

It has a zero, which is the $[0, 1, 0]$, since the line through any point P and $[0, 1, 0]$ is vertical. To see this, consider $\alpha X + \beta Y + \gamma Z = 0$; if $[0, 1, 0]$ is on this line, then $\beta = 0$. Hence the line is of the form

$\alpha X + \gamma Z = 0$. If $\alpha \neq 0$, this represents a vertical line (verify by setting $Z = 1$). If $\alpha = 0$, then $\gamma \neq 0$ and this is the line at infinity we discussed before. Thus the group law produces $P + [0, 1, 0] = P$.

It has inverses: any two points on a vertical line are inverses, using the group law and the fact that every vertical line passes through the identity. (Draw a picture.)

What remains is associativity. This is rather tedious to do with the equations, but it is possible. I'll give a sketch of a geometric proof which should give some intuition, but it lacks details. First, we draw a diagram of 9 points and the horizontal and vertical lines passing through them:

It is actually constructed piece by piece as follows: locate P, Q and R, and draw a line through P and Q and through Q and R. Then these lines pass through $P * Q$ and $Q * R$ reppectively. Here $*$ denotes simply the third point of the cubic on that line, not the addition law. The addition law is actually: $P + Q = (P * Q) * \infty$. Locate the point ∞ and draw the lines through $P * Q$ and ∞ and $Q * R$ and ∞ respectively; then $P + Q$ and $Q + R$ reside on these lines, so locate those. Finally, draw the lines through P and $Q + R$ and through R and $P + Q$. If their intersection point is actually on E, then we will have proved

$$P * (Q + R) = (P + Q) * R$$

which suffices to show associativity.

Notice that the three vertical lines form a cubic C_1 and the three horizontal lines form a cubic C_2 (it is not an irreducible cubic, but three lines do form a cubic). There are 9 points of intersection, 8 of which we know are on E. So we only need to show that if E goes through 8 points on two different cubics, then it goes through their 9th point of intersection.

To see this, notice first that 10 coefficients determine a cubic:

$$x^3, y^3, xy^2, x^2y, x^2, y^2, xy, x, y, 1.$$

Of course scaling doesn't change the cubic, but we're parametrizing equations. Those that pass through 8 distinct points satisfy 8 conditions, so they form a 2-dimensional family (you have to be convinced the 8 conditions are independent; this is true if the points are in 'general position', i.e. no more collinear than we have constructed them to be, which is the case so long as the lines are all distinct; details omitted as actually you need to do some special cases separately). So if we have two independent elements of this family (C_1 and C_2 are not the same cubic so their equations are not scalar multiples), then any other in the

family (in our case E) is a linear combination of these two. So it also passes through their 9th point of intersection. □

So we have a group!

The arithemtic study of elliptic curves is largely centred on and motivated by the question: which groups can you get?

Here's a famous starting point:

Theorem 58.2 (Mordell)**.** *The group $E(K)$ is a finitely generated abelian group.*

Therefore the group has the form

$$\mathbb{Z}^r \times \mathbb{Z}/N_1\mathbb{Z} \times \cdots \mathbb{Z}/N_\ell\mathbb{Z}.$$

That is, it consists of a free abelian part of *rank* r, and a torsion part. The torsion part is fairly well tamed over $\mathbb{Q}$, since we know the following theorem.

Theorem 58.3 (Mazur)**.** *Let E be an elliptic curve defined over $\mathbb{Q}$. Then the torsion part of $E(\mathbb{Q})$ is one of the following groups:*

$$\mathbb{Z}/n\mathbb{Z}, n = 1, 2, \ldots, 10, 12, \quad \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2n\mathbb{Z}, n = 1, 2, 3, 4.$$

Similar precise results are known for some small number fields, and it is known that there is a finite list for each number field.

However, we have no idea what ranks are possible. Ranks up to size 28 have been discovered, but it is a major open question to determine if they are bounded.

The Singular Case

A point on the Weierstrass equation $F(X, Y, Z) = 0$ is a singular point if and only if the partials $\frac{\partial F}{\partial X}$, $\frac{\partial F}{\partial Y}$, $\frac{\partial F}{\partial Z}$ all vanish at that point.

Write the Weierstrass equation as

$$Y^2Z = G(X, Z)$$

If there is a singular point $P = [X_0, Y_0, 1]$, then

$$2Y_0 = 0, \quad \frac{dG}{dx}(X_0, 1) = 0$$

which means that $Y_0 = 0$ and X_0 is a repeated root of $G(X, 1)$. This gives the discriminant condition,

$$4A^3 - 27B^2 = 0.$$

There are two possibilities:

First, $A = B = 0$. The cuspidal cubic

$$ZY^2 = X^3.$$

A line not passing through the origin can be written

$$\alpha X + \beta Y + \gamma Z = 0$$

where $\gamma \neq 0$, so by scaling we may assume $\gamma = 1$. Then, intersecting with the cubic, we obtain

$$(\alpha X + \beta Y)Y^2 + X^3 = 0$$

This has three roots. Write the points of intersection as

$$[X_1, Y_1, Z_1], [X_2, Y_2, Z_2], [X_3, Y_3, Z_3].$$

Note that Y_i are not zero (line not through the origin), so if we prefer, we can use $Y_i = 1$. Then we have that the X_i are roots of the cubic

$$(\alpha X + \beta) = -X^3.$$

Since this has zero X^2 coefficient, we obtain

$$X_1 + X_2 + X_3 = 0.$$

Therefore we can identify the non-singular points with the additive group of the field, under the same geometric group law as before.

Second case. This case we leave as an exercise, but the end result is as follows: the non-singular points form a group isomorphic to the multiplicative group of the field.

Reduction of Elliptic Curves

If you recall, we learned that for degree 2 homogeneous equations, there are solutions over $\mathbb{Q}$ if and only if there are solutions over $\mathbb{Q}_p$ for all primes p and over $\mathbb{R}$ (the Hasse-Minkowski Theorem). This is part of the *local-global principal*. So, not surprisingly, one of the main tools in the study of elliptic curves is to look 'locally.'

Consider E, an elliptic curve over $\mathbb{Q}$. We can consider the same curve over $\mathbb{Q}_p$ since $\mathbb{Q}$ embeds in $\mathbb{Q}_p$. Then we can reduce modulo p, using the reduction map

$$\mathbb{Z}_p \to \mathbb{F}_p.$$

To be precise, any point in $\mathbb{P}^2(\mathbb{Q}_p)$ with rational coordinates can be written with integer coordinates where the valuations of at least one coordinate is zero. (Restricting to $\mathbb{Q}$ coordinates, we could use rational integer coordinates with gcd 1). Applying the reduction map to these coordinates, we obtain a map

$$\mathbb{P}^2(\mathbb{Q}_p) \to \mathbb{P}^2(\mathbb{F}_p).$$

Any curve (including lines) in $\mathbb{P}^2$ which has coefficients which are integers with at least one of zero valuation, can be reduced this way also.

For that reason, we will assume our elliptic curve has integer coefficients in Weierstrass form. (If it does not, we can change coordinates, but this adds a layer of complication we'll just ignore for now.) If we are given a line, we can multiply by a constant to obtain an equation in the required form.

For example, the elliptic curve

$$y^2 = x^3 + 8$$

becomes the curve

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

modulo 3, and the point $(1, 3)$ on the curve becomes $(1, 0)$. If P lies on a curve C, then the reduction $\overline{P}$ lies on the reduced curve $\overline{C}$. So the reduction map gives rise to a reduction of the curves

$$C(\mathbb{Q}_p) \to \overline{C}(\mathbb{F}_p).$$

Proposition 58.4. *The non-singular points of $\overline{C}$ are all in the image of this map.*

Proof. Let $P_0 \in \overline{C}$ be non-singular. Then

$$\frac{\partial \overline{F}}{\partial X}(P_0) \neq 0$$

where $\overline{F}(X, Y, Z) = 0$ is the equation of $\overline{C}$. Write $P_0 = [X_0, Y_0, Z_0]$, where $X_0, Y_0, Z_0 \in \mathbb{F}_p$. Lift these coordinates to any $X, Y, Z \in \mathbb{Z}_p$. Note that $F(X, Y, Z)$ does not necessarily vanish. Let

$$G(T) = F(T, Y, Z).$$

Then $\frac{dG}{dT}(X) \not\equiv 0 \pmod{p}$, while $G(X) \equiv 0 \pmod{p}$. So we may apply Hensel's lemma to find a lift X' of X_0 for which $F(X', Y, Z) = G(X') = 0$. □

Lemma 58.5. *Let L be a line and C be a cubic. Suppose that L meets C in the three points P, Q, R, counted with multiplicities. Then either*

(1) $\overline{L}$ is entirely contained in $\overline{C}$; or
(2) $\overline{L}$ meets $\overline{C}$ at $\overline{P}$, $\overline{Q}$, $\overline{R}$, counted with multiplicities.

Proof. Write L as

$$\alpha X + \beta Y + \gamma Z = 0$$

Write C as

$$F(X, Y, Z) = 0$$

where not all coefficients vanish modulo p in either equation. Assume without loss of generality that γ does not vanish mod p. Solve for Z in the line and substitute into F, obtaining some

$$G(X,Y) = F(X,Y,-\alpha/\gamma X - \beta/\gamma Y) = 0$$

If this equation is $0 = 0$ modulo p, then $\overline{L}$ is contained in $\overline{C}$. So let us assume not. In any case, $G(X,Y)$ has the form

$$\lambda(p_y X - p_x Y)(q_y X - q_x Y)(r_y X - r_x Y) = 0$$

where $P = [p_x, p_y, p_z]$, $Q = [q_x, q_y, q_z]$, $R = [r_x, r_y, r_z]$. However, these points do not have both of their first two coordinates vanishing modulo p, and λ does not vanish mod p (as otherwise G would vanish). Then we may reduce this equation and we obtain another of degree three: $\overline{P}$, $\overline{Q}$ and $\overline{R}$ are the three points on the intersection of $\overline{L}$ and $\overline{C}$. □

Definition 58.6.

$$E_0(\mathbb{Q}_p) = \{P \in E(\mathbb{Q}_p) : \overline{P} \text{ is non-singular}\}.$$

Theorem 58.7. *$E_0(\mathbb{Q}_p)$ is a subgroup of $E(\mathbb{Q}_p)$. The reduction induces a homomorphism $E_0(\mathbb{Q}_p) \to \overline{E}(\mathbb{F}_p)$ of groups.*

Proof. Reduction takes lines to lines. E_0 fails to be a group if it is not closed under addition or inverses. But this would entail a line two of whose points are in E_0 and one which is not. This would then give a line in $\mathbb{P}_2(\mathbb{F}_p)$ which intersects $\overline{E}$ at two non-singular points and at the singular point. Such lines don't exist (the node or cusp is always a multiplicity two intersection). So E_0 is a subgroup. From the previous lemma and the definition of the group law, $\overline{P+Q} = \overline{P} + \overline{Q}$. So the map is a homomorphism. □

Thus we obtain homomorphisms:

$$E_0(\mathbb{Q}) \to E_0(\mathbb{Q}_p) \to \overline{E}(\mathbb{F}_p)$$

This motivates our study of elliptic curves over the p-adics and over finite fields.

Elliptic curves over the p-adics

Let us suppose that we are studying an elliptic curve

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

where $a, b \in \mathbb{Z}_p$. If $a, b \in \mathbb{Q}_p$ only, then there is a change of coordinates $x \mapsto u^2 x, y \mapsto u^3 y$ which can fix the problem. We can transfer the information obtained here through the change of coordinates later. So

for now let's stick with the simple case. But keep in mind that reduction modulo p depends on the particular equation.

The reduced curve over $\mathbb{F}_p$ doesn't contain a line, even though it may be singular (i.e. the homogenized equation can't have a linear factor).

We write

$$E_0(\mathbb{F}_p) = \{P \in E(\mathbb{F}_p) : P \text{ is non-singular}\}.$$

This lets us also write

$$E_0(\mathbb{Q}_p) = \{P \in E(\mathbb{Q}_p) : \overline{P} \in E_0(\mathbb{F}_p)\}.$$

The map

$$E_0(\mathbb{Q}_p) \to E_0(\mathbb{F}_p)$$

is a surjective group homomorphism, as we saw earlier. We will now study its kernel, which is the collection of points of $E(\mathbb{Q}_p)$ whose reduction modulo p is the point at infinity.

Proposition 58.8. *The kernel is*

$$\mathcal{K} := \{[0,1,0]\} \cup \{[X,Y,1] : X, Y \notin \mathbb{Z}_p\}.$$

Proof. Suppose $[X,Y,Z]$ is in the kernel, but $Z \neq 0$. Then, multiplying up the coordinates of $[X,Y,Z]$ so that they are in $\mathbb{Z}_p$, we obtain X and Z of positive valuation, Y of zero valuation, and $Z \neq 0$. The claim is that:

$$v_p(X) < v_p(Z)$$

The Weierstrass equation gives $ZY^2 = X^3 + aXZ^2 + bZ^3$, where we are assuming $v_p(a), v_p(b) \geq 0$. Since $v_P(Y) = 0$, we find that

$$\begin{aligned} v_p(Z) &\geq \min\{3v_p(X), v_p(a) + v_p(X) + 2v_p(Z), v_p(b) + 3v_p(Z)\} \\ &\geq \min\{v_p(X) + 1, 3v_p(Z)\}. \end{aligned}$$

The claim follows from this and the fact that $v_p(Z) > 0$. Therefore, dividing by the $p^{v_p(Z)}$, we obtain the form in the statement. □

Now suppose that $[X,Y,1]$ is in the kernel $\mathcal{K}$, so that $v_p(X), v_p(Y) < 0$. Then

$$2v_p(Y) = 3v_p(X)$$

since of the three terms on the right in the Weierstrass equation, X^3 must have smallest valuation. So, we can write

$$v_p(Y) = -3n, \quad v_p(X) = -2n$$

where $n \in \mathbb{Z}^{\geq 0}$. Call n the *level* of the point. Something not in the kernel is defined to have level 0. The level of $[0,1,0]$ is defined to be ∞.

The important fact is that the kernel of this reduction is all points of level $n \geq 1$.

Let $N \geq 1$. Let's change coordinates:

$$X_N = p^{2N}X, \quad Y_N = p^{3N}Y, \quad Z_N = Z.$$

The Weierstrass equation becomes a new curve:

$$E^{(N)} : Y_N^2 Z_N = X_N^3 + p^4 a X_N Z_N^2 + p^6 b Z_N^3.$$

Now, if we reduce modulo p in this *new* situation, we get

$$\overline{E^{(N)}} : Y_N^2 Z_N = X_N^3.$$

(Note that in class I fixed my conflict of notation between E_N and $E^{(N)}$ differently. Upon reflection, I like this fix better.)

A point (X, Y) changes to a point $(p^{2N}X, p^{3N}Y)$ which

(1) reduces to the singular point of $\overline{E^{(N)}}$ if $\text{level}(X, Y) < N$.
(2) maps to the identity of $\overline{E^{(N)}}$ (is in the kernel) if $\text{level}(X, Y) > N$.

But $\overline{E^{(N)}}$ is that cuspidal cubic we saw earlier; the non-singular points form an additive group.

Define for $N \geq 1$,

$$E_N(\mathbb{Q}_p) := \{P \in E(\mathbb{Q}_p) : \text{level}(P) \geq N\}.$$

Lemma 58.9. *We have containments*

$$E(\mathbb{Q}_p) \supset E_0(\mathbb{Q}_p) \supset E_1(\mathbb{Q}_p) \supset E_2(\mathbb{Q}_p) \supset \cdots .$$

This is called the p-adic filtration. *We have for* $N \geq 1$,

$$E_N(\mathbb{Q}_p)/E_{N+1}(\mathbb{Q}_p) \cong \mathbb{Z}/p\mathbb{Z}.$$

And

$$E_0(\mathbb{Q}_p)/E_1(\mathbb{Q}_p) \cong \overline{E}_0(\mathbb{F}_p).$$

i.e. the group of non-singular points on $\overline{E}$.

Proof. Containment is clear.

The group

$$E_N(\mathbb{Q}_p)/E_{N+1}(\mathbb{Q}_p)$$

is the non-singular points of $E^{(N)}$ (the same as those which reduce to non-singular points on $\overline{E^{(N)}}$), modulo the kernel of reduction. Since reduction is a homomorphism of groups, and the image is surjective, we find that this group is exactly $\overline{E^{(N)}}_0(\mathbb{F}_p)$, which we saw in the cuspidal reduction case, is isomorphic to the additive group of the field, i.e. $\mathbb{Z}/p\mathbb{Z}$.

By the same argument, the last displayed equation in the statement holds, but here we don't know that $\overline{E}$ is cuspidal; in fact, it could be various different things. So we can't say more.

□

Corollary 58.10. *Let $P \in E(\mathbb{Q}_p)$ be of finite order $m > 1$, where m and p are coprime. Then $P = [x, y, 1]$ where $x, y \in \mathbb{Z}_p$.*

Proof. If not, then P is of some level $N \geq 1$. Then $P \in E_N(\mathbb{Q}_p)$ but $P \notin E_{N+1}(\mathbb{Q}_p)$. So it maps to a non-zero element of $E_N(\mathbb{Q}_p)/E_{N+1}(\mathbb{Q}_p)$, which must have order p, so $p \mid m$. □

Plan: improve this so doesn't depend on N and p being coprime. Note that what this theorem says, in another language, is this:

If P has level N, then so does mP. But if mP is the identity, this can't be so (since the identity has infinite level). Contradiction.

But we have to improve the proof for the case that $p \mid m$. In this case, the level of P will rise. We have to determine how much it rises, and show that it does not rise to ∞ to reach the same contradiction.

So look closer at the homomorphism above

$$\phi : E_N(\mathbb{Q}_p)/E_{N+1}(\mathbb{Q}_p) \to \mathbb{Z}/p\mathbb{Z}.$$

This is done by first changing coordinates, and then using our analysis of the cuspidal cubic. The map ϕ looks in practice like the following:

$$[x, y, 1] \mapsto [p^{2N}x, p^{3N}y, 1] = [p^{-N}x/y, 1, p^{-3N}/y] \mapsto p^{-N}x/y.$$

If $[x, y, 1]$ is of level N, then x/y is coprime to p.

So, if we define $u([x, y, 1]) = x/y$ for $[x, y, 1]$ of all levels, and $u([0, 1, 0]) = 0$, then x/y is coprime to p and $v_p(u(P)) = \text{level}(P)$.

Lemma 58.11. *Let $P_1, P_2 \in E_1(\mathbb{Q}_p)$.*

$$v_p(u(P_1 + P_2) - u(P_1) - u(P_2)) \geq 5 \min\{v_p(u(P_1)), v_p(u(P_2))\}.$$

Proof. First, suppose one of $P_1 = [x, y, 1]$, P_2, $P_1 + P_2$ is the identity. In the first two cases, the relation reduces to a trivially true statement. (Note, the valuation of 0 is infinity.) In the third case, note that $u(P) = -u(-P)$ since the inverse of $[x, y, 1]$ is $[x, -y, 1]$.

Without loss of generality, let's assume

$$v_p(u(P_1)) \leq v_p(u(P_2)).$$

Let $N \geq 1$ be the level of P_1 (which is the smaller of the two levels). Then

$$v_p(x) = -3N, v_p(y) = -2N.$$

Let's consider E_N as above. Since P_1 and P_2 do not map to the singularity, neither does $-P_1 - P_2$ (the third point on the line joining them),

so the line doesn't pass through $[0,0,1]$, and hence has non-zero coefficient on Z_N:

$$Z_N = \alpha Z_N + \beta Y_N.$$

Plugging $[x,y,1]$ into the line, we find that, looking at valuations,

$$v_p(\alpha) \geq 0, v_p(\beta) \geq 0.$$

Now we'll intersect with E:

$$0 = -Y_N^2(\alpha X_N + \beta Y_N) + X_N^3 + p^{4N} a X_N (\alpha X_N + \beta Y_N)^2 + p^{6N} b (\alpha X_N + \beta Y_N)^3$$

Simplifying

$$0 = c_3 X_N^3 + c_2 X_N^2 Y_N + c_1 X_N Y_N^2 + c_0 Y_N^3$$

where

$$c_3 = 1 + p^{4N} a\alpha^2 + p^{6N} b\beta^3,$$
$$c_2 = 2p^{4N}\alpha\beta a + 3p^{6N}\alpha^2\beta b.$$

Therefore,

$$v_p(c_3) = 0, \quad v_p(c_2) \geq 4N.$$

The roots of this equation are, by construction,

$$X_N/Y_N = -p^{-N}u(P_1+P_2), p^{-N}u(P_1), p^{-N}u(P_2).$$

The sum of these roots must be $-c_2/c_3$ from the equation, and we are done. □

As a consequence, u has a sort of linearity:

Lemma 58.12. *Let $P \in E_1(\mathbb{Q}_p)$, $s \in \mathbb{Z}^{>0}$.*

$$v_p(u(sP)) = v_p(s) + v_p(u(P))$$

Proof. One can prove from the previous lemma by induction that for integers $s > 0$,

$$v_p(u(sP) - su(P)) \geq 5 v_p(u(P))$$

The three quantities

$$u(sP) - su(P), u(sP), su(P)$$

are three sides of a triangle, i.e. their valuations are two smaller and equal, one potentially larger. If $p \nmid s$, so that $v_p(su(P)) = v_p(u(P))$, then it must be (because of the inequality above) that $u(sP) - su(P)$ is the largest, so

$$v_p(u(sP)) = v_p(u(P)).$$

The same holds if $s = p$, since then $v_p(su(P)) = v_p(u(P)) + 1$.

Now induct on the power of p dividing s using the statement of the lemma, pulling out one power at a time. □

Theorem 58.13. *The group $E_1(\mathbb{Q}_p)$ is torsion free. In other words, torsion is injective under reduction.*

Proof. Suppose $P \in E_1(\mathbb{Q}_p)$ has order m and level N. Then $mP = 0$, so

$$\infty = v_p(u(0)) = v_p(m) + v_p(u(P)) = v_p(m) + N.$$

which is a contradiction. □

Corollary 58.14. *Suppose that $p \neq 2$, and $v_p(4a^3 + 27b^2) = 0$. Then the torsion subgroup of $E(\mathbb{Q}_p)$ is isomorphic to a subgroup of $\overline{E}(\mathbb{F}_p)$.*

Proof. By the discriminant condition, $E(\mathbb{Q}_p) = E_0(\mathbb{Q}_p)$, so

$$\overline{E}(\mathbb{F}_p) = E(\mathbb{Q}_p)/E_1(\mathbb{Q}_p).$$

But the quotient includes none of the torsion. □

Now we return to E being a curve over $\mathbb{Q}$, with coefficients in $\mathbb{Z}$. Our work is going to pay off to tell us about the torsion points on the elliptic curve.

Theorem 58.15 (Nagell-Lutz)**.** *The torsion part of $E(\mathbb{Q})$ is finite. A non-identity point $P = [x, y, 1] \in E(\mathbb{Q})$ of finite order satisfies $x, y \in \mathbb{Z}$ and either $y = 0$ or $y^2 \mid (4a^3 + 27b^2)$.*

Proof. By the injection $E(\mathbb{Q}) \mapsto E(\mathbb{Q}_p)$, we find $x, y \in \mathbb{Z}_p$ for all p, so that

$$x, y \in \mathbb{Z}.$$

Now suppose p is prime, but $p \neq 2$ and $p \nmid 4a^3 + 27b^2$. Then the torsion group of $E(\mathbb{Q})$ is isomorphic to a subgroup of $\overline{E}(\mathbb{F}_p)$. So it is finite. (In a given situation, one could look modulo p for various p to restrict the torsion group.) If $2P = 0$, then $P = -P$, so $y = 0$.

Otherwise, write $2P = [x_2, y_2, 1]$. We have established that $x_2, y_2 \in \mathbb{Z}$. Now one needs to work out the doubling formula and then some manipulations then show $y^2 \mid 4a^3 + 27b^2$. Here is the sketch:

The doubling formula tells us x_2 in terms of x and y:

$$x_2 + 2x = \frac{(3x^2 + a)^2}{4(x^3 + ax + b)}$$

Since the left side is an integer, looking modulo $x^3 + ax + b$, we have

$$(3x^2 + a)^2 \equiv 0$$

But at the same time, it is a simple calculation that

$$(3x^2 + 4a)(3x^2 + a)^2 \equiv 4a^3 + 27b^2$$

Since $y^2 = x^3 + ax + b$, we discover that

$$y^2 \mid 4a^3 - 27b^2.$$

□

Now we have a tool to compute the torsion on an elliptic curve! One way to bound it: look modulo p for various p. To exhaustively enumerate it: use the theorem.

A bit about Finite Fields

Recall these statements from earlier in the course:

Lemma 58.16. *(previously Lemma 27.1) Let $f(x) \in k[x]$, where k is a field, and suppose that $f(x)$ is not identically zero. Let $n = \deg f(x)$. Then f has at most n distinct roots in k.*

Corollary 58.17. *Let $f(x), g(x) \in k[x]$, where k is a field. Suppose that $n = \deg f(x) = \deg g(x)$. If $f(\alpha_i) = g(\alpha_i)$ for $n+1$ distinct values*

$$\alpha_1, \alpha_2, \ldots, \alpha_{n+1},$$

then $f(x) = g(x)$.

Proposition 58.18.

$$x^{p-1} - 1 \equiv (x-1)(x-2)\cdots(x-(p-1)) \pmod{p}$$

We can update this. Consider a finite field $\mathbb{F}_q$ of q elements (we don't assume anything about q yet, although we'll discover shortly that it is a power of a prime). Since it is a field, $\mathbb{F}_q^*$ has $q-1$ elements. By group theory, then $x^{q-1} = 1$ for all such elements. Also $0^q = 0$, so every element of the field satisfies $x^q = x$.

Proposition 58.19. *Let $\mathbb{F}_q$ be a finite field of q elements. Then*

$$x^q - x = \prod_{\alpha \in \mathbb{F}_q} (x - \alpha).$$

Proof. Let $f(x) = x^q - x - \prod_{\alpha \in \mathbb{F}_q}(x-\alpha)$. Then $\deg f(x) < q$ since the leading terms cancel. But it has q distinct roots in $\mathbb{F}_q$ (all elements). So $f(x)$ must be identically zero, by Lemma 27.1. □

Corollary 58.20. *Suppose that L is a field extension of $\mathbb{F}_q$. Then $\alpha \in L$ is an element of $\mathbb{F}_q$ if and only if $\alpha^q = \alpha$.*

Here's another proposition from before:

Proposition 58.21. *If $d \mid p-1$, then $x^d \equiv 1 \pmod{p}$ has exactly d solutions.*

Here's the updated version we'll prove:

Proposition 58.22. *Let k be an y field. Then $x^d - 1 \mid x^n - 1$ in $k[x]$ if and only if $d \mid n$. Similarly, for a nonzero integer a, $a^d - 1 \mid a^n - 1$ if and only if $d \mid n$.*

Proof. Let $n = dd' + r$ (division algorithm!), where $0 \leq r < d$. If $r = 0$, then

$$g(x) = \frac{x^n - 1}{x^d - 1} = \frac{(x^d)^{d'} - 1}{x^d - 1} = (x^d)^{d'-1} + (x^d)^{d'-2} + \cdots + x^d + 1$$

and

$$x^n - 1 = (x^d - 1)g(x).$$

In general,

$$\frac{x^n - 1}{x^d - 1} = \frac{(x^d)^{d'} - 1}{x^d - 1} + x^{dd'} \frac{x^r - 1}{x^d - 1} = g(x) + x^{dd'} \frac{x^r - 1}{x^d - 1}$$

is a polynomial if and only if $\frac{x^r-1}{x^d-1}$ is; i.e. if and only if $r = 0$.

The second statement has the exact same proof. □

Proposition 58.23. *Let $\mathbb{F}_q$ be a finite field of q elements. If $d \mid q - 1$, then $x^d = 1$ has exactly d solutions in $\mathbb{F}_q$.*

Proof. If $x^d - 1$ had fewer than d roots, then by the divisibility just witnessed, and Lemma 58.16, $x^{q-1} - 1$ would have fewer than $q - 1$ roots. But it has as roots all elements of $\mathbb{F}_q^*$, i.e. it has $q - 1$ roots. By this contradiction, we have proven the proposition. □

Finally, we showed before that:

Theorem 58.24. *$(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic.*

We can update the proof to see that

Theorem 58.25. *Let $\mathbb{F}_q$ be a finite field of q elements. Then $\mathbb{F}_q^*$ is cyclic.*

Proof. In any finite commutative group G, there exists $y \in G$ whose order is the least common multiple of the orders of all the elements of G. In particular, if n is the order of y, then $x^n = 1$ for all $x \in G$.

So all elements of $\mathbb{F}_q^*$ are roots of $x^n - 1$, but 0 is not, so it has exactly $q - 1$ roots in $\mathbb{F}_q$.

But $x^n - 1$ has at most n roots in $\mathbb{F}_q$ (by Lemma 27.1).

So

$$q - 1 \leq n.$$

On the other hand, $1, y, y^2, \ldots, y^{n-1}$ are all distinct, by the fact that y has order n. Therefore, $x^n - 1$ has at least n roots in $\mathbb{F}_q$, so

$$q - 1 \geq n.$$

Hence $n = q - 1$ and

$$\mathbb{F}_q^* = \{1, y, y^2, \ldots, y^{n-1}\} = \langle y \rangle.$$

□

The rest of this section is new, not a repeat of $\mathbb{Z}/p\mathbb{Z}$.

Lemma 58.26. *Let $\mathbb{F}_q$ be a finite field of q elements. Then there is exactly one ring homomorphism[6] $\mathbb{Z} \to \mathbb{F}_q$, and its image is isomorphic to $\mathbb{Z}/p\mathbb{Z}$ for some prime p.*

Proof. Any such ring homomorphism is forced to take the identity to the identity, so the map must be $n \mapsto ne$ where e is the identity of $\mathbb{F}_q$. The image is a finite subring of $\mathbb{F}_q$, which must be an integral domain. Hence the kernel is a prime ideal. □

In this case, the field must have characteristic p, i.e. $px = p(ex) = (pe)x = 0x = 0$ for all $x \in \mathbb{F}_q$.

Corollary 58.27. *Any finite field of size p is isomorphic to $\mathbb{Z}/p\mathbb{Z}$.*

We write $\mathbb{F}_p$ for this field.

Theorem 58.28. *The number of elements in a finite field is a power of its characteristic.*

Proof. The field $\mathbb{F}_q$ has subfield $\mathbb{F}_p$. Then in particular it is a finite dimensional vector space over $\mathbb{F}_p$. It must therefore have size p^n where n is the dimension. □

Proposition 58.29. *Let k be a field of characteristic p. Then*

$$(x + y)^{p^d} = x^{p^d} + y^{p^d}, \quad x, y \in \mathbb{F}_q$$

for all $d \in \mathbb{Z}^{>0}$. This implies that the map

$$x \mapsto x^{p^d}$$

is a ring homomorphism on any field of characteristic p.

Proof. For $d = 1$, this is an application of binomial theorem; all the terms besides the first and last have coefficient divisible by p.

For higher $d = 2$, just repeat, etc. etc.

The ring homomorphism property that wasn't immediate is the one we checked (that is respects addition). □

[6]of commutative rings with identity, don't forget we're number theorists

Theorem 58.30. *The subfields of a finite field $\mathbb{F}_q$ of size $q = p^n$, are in one-to-one correspondence with the divisors of n, where for each divisor d of n, we have a subfield of size p^d.*

Proof. Each subfield is a vector space over $\mathbb{F}_p$ and so has size p^d for some d. We will show that $d \mid n$. Call this subfield E. Then E^* is of size $p^d - 1$ and $x^{p^d-1} - 1$ is the polynomial over $\mathbb{F}_p$ having roots exactly all elements of E^*. Since this is true for $\mathbb{F}_q$ also, we have

$$x^{p^d-1} - 1 \mid x^{p^n-1} - 1.$$

Therefore $d \mid n$.

Conversely, suppose that $d \mid n$. Define E to be the subset of $\mathbb{F}_q$ consisting of elements fixed by the p^d Frobenius:

$$E = \{x \in F : x^{p^d} = x\}.$$

The Frobenius is a ring homomorphism by a recent lemma, so the elements fixed by it form a field (just check; being fixed is closed under the various properties, including inverses). We're not done, though, because we need to discover that E actually has p^d elements (if $d \nmid n$, this fails).

Since $d \mid n$, we have $p^d - 1 \mid p^n - 1$, so $x^{p^d-1} - 1 \mid x^{p^n-1} - 1$. So $x^{p^d} - x$ has exactly p^d roots (and not fewer), but these roots are exactly E, by definition.

Now, if E' is another subfield having the same size, then its elements also satisfy the same polynomial $x^{p^d} - x$ which has as roots a unique subset of the elements of F, so it is exactly E. □

Corollary 58.31. *There is at most one finite field of size $q = p^d$, up to isomorphism.*

Proof. We already did the case $d = 1$.

Now suppose $d > 1$. Then the p^d elements all satisfy $x^{p^d} - x$ and in particular all roots of this polynomial are in the field. We call a minimal field over $\mathbb{F}_p$ in which a polynomial has all its roots a *splitting field*. In fact, *the* splitting field is unique, up to isomorphism, which shows finite fields of each size are unique up to isomorphism.

Here's a proof that splitting fields are unique up to isomorphism. First, assume $g(x)$ is an irreducible factor, of degree greater than 1 (if the polynomial is linear, the base field is its splitting field). Adjoining one root of $g(x)$ results in a field isomorphic to

$$F[x]/(g(x)).$$

Now over this field, by induction, the splitting field of the polynomial is a unique extension, up to isomorphism. Does starting with a different

irreducible factor matter? No: any other splitting field of our original polynomial must contain a field isomorphic to $F[x]/(g(x))$ and hence is isomorphic to the splitting field thus obtained. □

Proposition 58.32. *Write $F_d(x)$ to be the product of the finitely many monic irreducible polynomials in $\mathbb{F}_p[x]$ of degree d. Then*

$$x^{p^n} - x = \prod_{d|n} F_d(x).$$

Proof. First, $x^{p^n} - x$ is squarefree, since if it were of the form $f(x)^2 g(x)$, then

$$-1 = 2f(x)f'(x)g(x) + f(x)^2 g'(x)$$

which would imply $f(x) \mid -1$.

Now we show that all monic irreducible factors of $x^{p^n} - x$ are exactly those of degree dividing n.

Claim 1: Let $f(x)$ be a monic irreducible polynomial dividing $x^{p^n} - x$. Then $d = \deg f$ divides n. Let α be a root of f. Then $\mathbb{F}_p(\alpha)$ is a vector space of dimension d over $\mathbb{F}_p$ and hence has p^d elements, the roots of $x^{p^d} - x$. So α is a root of a factor of $x^{p^n} - x$, and hence α is fixed by the p^n-Frobenius. As Frobenius is a ring homomorphism fixing $\mathbb{F}_p$, every element of $\mathbb{F}_p(\alpha)$ must be fixed by it. So all elements of $\mathbb{F}_p(\alpha)$ satisfy $x^{p^n} - x$, and so

$$x^{p^d - 1} - 1 \mid x^{p^n - 1} - 1$$

and therefore $p^d - 1 \mid p^n - 1$ so $d \mid n$.

Claim 2: Let $f(x)$ be a monic irreducible polynomial such that $d = \deg f \mid n$. Then $f(x) \mid x^{p^n} - x$. Any root α is fixed by p^d-Frobenius, so α is a root of $x^{p^d} - x$ so

$$f(x) \mid x^{p^d} - x \mid x^{p^n} - x.$$

□

Corollary 58.33. *The field $\mathbb{F}_q(\alpha)$, where α is a root of a monic irreducible $f(x) \in \mathbb{F}_q[x]$, is of size q^n where n is the degree of f, and it is the splitting field of $f(x)$.*

Proof. We found that $f(x) \mid x^{p^d} - x$, but the latter factors completely into linear factors in $\mathbb{F}_q(\alpha)$. □

Let

$$N_d = \#\{\text{monic irreducible polynomials of degree } d\}.$$

Corollary 58.34.

$$p^n = \sum_{d|n} dN_d$$

and by Mobius inversion

$$N_d = \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) p^d.$$

In particular, the number is at least one in each case, since it is a sum of distinct powers of p with coefficients ± 1. This shows that

Theorem 58.35. *There exists a finite field of $q = p^n$ elements for all primes p and positive integers n.*

Elliptic Curves over Finite Fields

There's a simple bound for the size of $E(\mathbb{F}_p)$: there are p possible x coordinates, each of which may have 2 possible y coordinates, plus the point at infinity: $2p + 1$.

In fact, what is known is that

Theorem 58.36 (Hasse Theorem)**.**

$$|\#E(\mathbb{F}_p) - p - 1| \leq 2\sqrt{p}.$$

The quantity in brackets is called a_p, the *trace of Frobenius*. Knowing the traces of Frobenius determine the curve and its arithmetic information in several different ways, and collecting information about their behaviour (as p varies) is one of the main avenues of research.

59. Rough Notes of Cryptography Stuff

We'll get to that, but it's Friday before break, so first a quick detour to elliptic curve factoring and elliptic curve cryptography.

Factoring:

Let $n \geq 2$ be composite we want to factor.

First, check that $\gcd(n, 6) = 1$, and n isn't a perfect power (try taking roots and finding small factors first).

Choose random $1 < b, x_1, y_1 < n$.

Let $c = y_1^2 - x_1^3 - bx_1$ modulo n. Then $P = (x_1, y_1)$ must be on the curve

$$y^2 = x^3 + bx + c$$

considered modulo n.

Check that $\gcd(\Delta, n) = 1$ (It might give us a factor! But if it is n, go back and choose a different b.)

Choose k a product of small primes to smallish powers, e.g. k is the lcm of the first K integers.

Compute kP performing computations modulo n. If this fails because something isn't invertible, we found a factor of n. As a rational, it will always have the form $\frac{a_k}{d_K^2}, \frac{b_k}{d_k^3}$), but when we're working mod n, the addition law breaking is the same as $\gcd(d_k, n) > 1$.

Compute $\gcd(d_k, n)$. If we find a factor of n, yaya. If the gcd is 1, go back and pick a larger k, or choose a new curve. If it is n, decrease k.

Double and add to compute kP efficiently.

Why does it work? Suppose $p \mid n$. If $E(\mathbb{F}_p)$ (which has point P on it) has size dividing k, then the order of P divides k. So computing kP will give the point at infinity, and d_k vanishes modulo p.

Diffie Hellman Key Exchange

Agree on $\mathbb{F}_q$, $E/\mathbb{F}_q$, $P \in E$. Alice: aP. Bob bP.

ElGamal

aree on suff. Alice: aP as usual.

Bob's message $M \in E$, random integer k, compute

$$B_1 = kP, \quad B_2 = M + kA$$

These B's are ciphertext. Alice computes $B_2 - aB_1 = M$.

$$B_2 - aB_1 = M + kA - akP = M + kaP - akP = M.$$

ECDSA (digital signature)

Agree as usual, N is order of p, Alice aP.

document: d mod N, and random integer k mod N. Alic computes kP, signature is

$$(s_1, s_2) = (x(kP) \pmod N, (d + as_1)k^{-1} \pmod N)$$

Note: choose integer representative of x(kP) that is between 0 and p-1.

Bob verifies using alice's public A. Compute

$$v_1 = ds_1^{-1} \pmod N, v_2 = s_1 s_2^{-1} \pmod N.$$

Then compute

$$v_1 P + v_2 A$$

this ought to have x coordinate s_1 modulo N.

CHeck:

$$v_1 P + v_2 A = ds_2^{-1} P + s_1 s_2^{-1} aP = (s_2^{-1}(d + as_1))P = kP$$

General group methods for ECDLP or any DLP.

Baby-Step-Giant-Step (Shanks)

Q = kP, find k

Let m = ceiling of $\sqrt{N}$.
compute P, 2P, ..., mP
compute R = -mP
compute Q+R, Q+2R, Q+3R, ..., Q + mR
match? iP = Q + jR; Q = iP + jmP.
Why must there be a match? Write $k = jm + i$ Then $0 \leq i < m$ and $j = \frac{m-i}{N}$ is also i nthat range.

60. Elliptic Curves over Finite Fields

We haven't got enough background to properly do this section justice, so we will provide only a sketch of further ideas.

Theorem 60.1 (Hasse)**.** *Let E be an elliptic curve defined over $\mathbb{F}_q$, a finite field of q elements. Then*

$$|\#E(\mathbb{F}_q) - q - 1| \leq 2\sqrt{q}.$$

Proof. The proof relies on the automorphisms of finite fields. In particular, the automorphisms of extensions of $\mathbb{F}_q$ are generated by the *Frobenius automorphism*,

$$\phi(x) = x^q$$

This automorphism has the property of 'picking out' elements of $\mathbb{F}_q$ inside larger extensions, i.e. $x \in \overline{\mathbb{F}_q}$ is actually in $\mathbb{F}_q$ if and only if $\phi(x) = x$. This is a little like the way complex conjugation picks out if a complex number is actually a real number.

This is an automorphism of the curve also, and it does the same work for us: a point in $\overline{\mathbb{F}_q}$ is actually in $\mathbb{F}_q$ if and only if it is fixed by the Frobenius map on E:

$$\phi([x, y, 1]) = [x^q, y^q, 1].$$

Therefore, the $\mathbb{F}_q$ points of E are exactly the kernel of $1 - \phi$. It remains to show that these are all multiplicity one, and to compute the degree of the map, so as to determine the size of the kernel and hence the size of $E(\mathbb{F}_q)$.

□

We write

$$a_q = q + 1 - \#E(\mathbb{F}_q).$$

and we call this the *trace of Frobenius*. The reason is as follows. The endomorphisms of E are, as a ring (one can add and compose endomorphisms), isomorphic to an order in a quadratic imaginary field, or

an order in a quaternion algebra. In particular, ϕ always satisfies a quadratic equation in this ring, i.e.

$$\phi^2 - a_q\phi + q = 0.$$

One can also look at the action of Frobenius on $E[\ell]$, the ℓ-torsion, which is isomorphic to $\mathbb{F}_\ell^2$ for ℓ a prime coprime to p. Then the equation above is its characteristic equation.

Hasse's Theorem is called the *Riemann hypothesis for elliptic curves*, because of an analogy with the Riemann zeta function. We'll turn to this next.

For a more elementary proof, see "The Riemann Hypothesis for Elliptic Curves" by Chahal and Osserman. The next section will roughly follow their exposition of the zeta function of an elliptic curve.

61. The Zeta function for elliptic curves

We defined a zeta function for $\mathbb{Q}$, and it can more generally be done for a number field. Then, we'll turn to function fields, i.e. fields which are fraction fields of

$$\mathbb{F}_q[x,y]/(f(x,y))$$

where $f(x,y)$ is an irreducible polynomial with coefficients in $\mathbb{F}_q$. There are a great many parallels between number fields and function fields.

We use the Euler product version of ζ to rewrite it a few different ways.

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Our first task is to rewrite this in terms of valuations. Recall that there's a valuation $v = v_p$ for each prime p (the p-adic valuation). Define

$$\mathcal{O}_v := \{x \in \mathbb{Q} : v(x) \geq 0\}, \quad \mathfrak{p}_v := \{x \in \mathbb{Q} : v(x) > 0\}.$$

Then $\mathcal{O}_v$ consists of rationals with no p in the denominator, and $\mathfrak{p}_v$ consists of those which do have a p in the numerator. The quotient

$$\mathcal{O}_v/\mathfrak{p}_v$$

then consists of rationals (with no p in the denominator) modulo p, so it is isomorphic to $\mathbb{Z}/p\mathbb{Z}$. We define the norm of a valuation as

$$n_v = \#\mathcal{O}_v/\mathfrak{p}_v.$$

So we can rewrite the Euler product as

$$\zeta(s) = \prod_v \left(1 - \frac{1}{n_v^s}\right)^{-1}.$$

where v ranges over the discrete valuations of $\mathbb{Q}$. (Note: the archimedean absolute valuation doesn't give a discrete valuation and is excluded.)

More generally, we can define everything analogously for a number field and get a zeta function for that field:

$$\zeta_K(s) = \prod_v \left(1 - \frac{1}{n_v^s}\right)^{-1}.$$

where now we range over discrete valuations of that field. Let K be a number field and $\mathcal{O}_K$ its ring of integers. The valuations in this case are $\mathfrak{p}$-adic, where $\mathfrak{p}$ is a prime ideal of $\mathcal{O}_K$. It turns out that in this case, $n_v = N(\mathfrak{p})$ in general. So we could also write:

$$\zeta_K(s) = \prod_{\mathfrak{p}} \left(1 - \frac{1}{N(\mathfrak{p})^s}\right)^{-1}.$$

This is called the *Dedekind zeta function*. It is also expected to satisfy the Riemann hypothesis, once extended to the whole complex plane.

Now we turn to global fields, i.e. let K be the fraction fields of $\mathbb{F}_q[x,y]/(f(x,y))$. Elements of K are rational functions, and valuations count how often some linear term appears; i.e. the vanishing at a point. To be precise, let P be a point on the curve $C : f(x,y)$. Then $v_P(f) = ord_P(f)$, the order of vanishing or being a pole of f at point P.

Interestingly, sometimes different points give the same valuation. Suppose we consider the example (Chahal-Osserman) of

$$\mathbb{F}_3[x,y]/(y)$$

What curve is this? It has equation $y = 0$ in two variables; it's a line (the x-axis). We have

$$\mathbb{F}_3[x,y]/(y) \cong \mathbb{F}_3[x]$$

so

$$K \cong \mathbb{F}_3(x)$$

rational functions in one variable over $\mathbb{F}_3$. (We should actually projectivize as usual; this is a projective line.)

Note that there is no square root of -1 in $\mathbb{F}_3 \cong \mathbb{Z}/3\mathbb{Z}$, so

$$\mathbb{F}_3[x]/(x^2+1)$$

is a ring extension, and in fact a field extension. It has 9 elements, and is called $\mathbb{F}_9$ (there is only one finite field of each allowable cardinality, up to isomorphism). Call the square root of -1 by the moniker i. Then i is a point on our curve, and it gives a valuation. But I claim that i and $-i$ both give the same valuation. For, any rational function

vanishes (or has a pole) to the same order at i and $-i$ because the rational functions have coefficients in $\mathbb{F}_3$, i.e. factors $(x-i)$ and $(x+i)$ always come together to the same power.

Of course, the reason this happens is that i isn't in the base field $\mathbb{F}_3$. It is in an extension of degree 2, and there 2 points give the same valuation. In general, a point defined over an extension of degree m (and no smaller extension) will give the same valuation as $m-1$ other points, and the norm of such a valuation is q^m.

Proposition 61.1. *Let C be a curve in $\mathbb{P}^2$, defined over $\mathbb{F}_q$, with affine equation $f(x,y)=0$. Let K be the fraction field of $\mathbb{F}_q[x,y]/(f(x,y))$. Then the valuations of K are ord_P for each point P defined over $\overline{\mathbb{F}_q}$ and $n_{v_P} = q^m$, where $\mathbb{F}_{q^m}$ is the field of definition of P.*

Sketch of proof. The statement that these are the only valuations is an analogue to Ostrowski's theorem. Here we'll work out the norm, sketchily. We have

$$\begin{aligned}\mathcal{O}_v &= \{ \text{ rational functions having no poles at } P \ \},\\ \mathfrak{p}_v &= \{ \text{ rational functions vanishing at } P \ \}.\end{aligned}$$

Let's suppose $m=1$ for simplicity, and let $P=(P_x,P_y)$. Then $\mathcal{O}_v$ consists of rational functions, where the denominator, when considered modulo $x=P_x, y=P_y$, does not vanish. On the other hand, $\mathfrak{p}_v$ consists of such things where the numerator vanishes.

The maximal ideal of polynomials vanishing at P is $I_P = (x-P_x, y-P_y)$, so we have

$$\mathcal{O}_v/\mathfrak{p}_v \cong (\mathbb{F}_q[x,y]/(f(x,y)))/I_P \cong \mathbb{F}_q$$

In greater generality, where $m \geq 1$, I_P is defined as above only in $\mathbb{F}_{q^m}[x,y]/f(x,y)$ and its restriction to $\mathbb{F}_q[x,y]/f(x,y)$ is generated by elements of higher degree. Details left to the reader. □

Now we can construct the zeta function of this function field. Let $N_m(C)$ be the number of points of C over $\mathbb{F}_q$, and let $\tilde{N}_m(C)$ be the number of points of C over $\mathbb{F}_q$ *not defined over any smaller field.* Let

$\tilde{C}(\mathbb{F}_{q^m})$ be the points defined over $\mathbb{F}_{q^m}$ and no smaller field.

$$\begin{aligned}
\zeta_K(s) &= \prod_v \left(1 - \frac{1}{N_v}\right)^{-1} \\
&= \prod_m \left(\prod_{P \in \tilde{C}(\mathbb{F}_{q^m})} \left(1 - \frac{1}{q^{ms}}\right)^{-\frac{1}{m}} \right) \\
&= \prod_m \left(1 - \frac{1}{q^{ms}}\right)^{-\frac{\tilde{N}_m(C)}{m}} \\
&= \exp\left(\sum_m \frac{\tilde{N}_m(C)}{m} \log\left(1 - \frac{1}{q^{ms}}\right) \right) \\
&= \exp\left(\sum_m \tilde{N}_m(C) \sum_n -\frac{q^{-mns}}{nm} \right) \\
&= \exp\left(\sum_d \sum_{m|d} \tilde{N}_m(C) \frac{q^{-ds}}{d} \right) \\
&= \exp\left(\sum_d N_m(C) \frac{q^{-ds}}{d} \right).
\end{aligned}$$

For function fields, the usual notation is $t = q^{-s}$ and Z instead of ζ. Back to our projective line over $\mathbb{F}_3$, we know that

$$N_m(C) = q^m + 1$$

so we get

$$\begin{aligned}
Z_C(t) &= \exp\left(\sum_m (q^m + 1) \frac{t^m}{m} \right) \\
&= \exp\left(\sum_m \left(\frac{(qt)^m}{m} + \frac{t^m}{m} \right) \right) \\
&= \exp\left(-\log(1 - qt) - \log(1 - t) \right) \\
&= \frac{1}{(1-t)(1-qt)}
\end{aligned}$$

So the zeta function is a simple rational function! Wow!

Finally, we can explain why Hasse's Theorem is the Riemann Hypothesis for Elliptic Curves. If one does the same work for an elliptic

curve, we find that

$$Z_E(t) = \frac{1 - a_q t + qt^2}{(1-t)(1-qt)}$$

(I'm not giving a proof; there's work to do here).

The Riemann Hypothesis is that this has zeroes only at $Re(s) = 1/2$. Let's put it back in terms of q:

$$Z_E(q) = \frac{1 - a_q q^{-s} + q^{1-2s}}{(1-q^{-s})(1-q^{1-s})}$$

So the zeroes of $Z_E(q)$ are the zeroes of the numerator. The zeroes are both on $Re(s) = 1/2$ if their t are of absolute value $q^{-1/2}$ (the complex part of s giving a rotation). The RH is therefore equivalent to the zeroes of $1 - a_q t + qt^2$ having absolute value $q^{-1/2}$, or the zeroes of $X^2 - a_q X + q$ both having absolute value $q^{1/2}$, or the zeroes of $X^2 - a_q X + q$ being complex conjugates, i.e. discriminant zero, i.e. $a_q^2 < 4q$ i.e. Hasse Bound. Voila!

62. Weil Conjectures

Now that we have the zeta function in terms of point counting, we could discuss the Weil Conjectures, but I won't head very far in that direction. Suffice it to say that they include the Riemann Hypothesis for smooth projective varieties, and say roughly that the Riemann Hypothesis holds, the zeta function has a functional equation, and the zeta function is always a nice rational function like we saw for elliptic curves. These have been proven.

63. Elliptic Curves over Complex Numbers

Let Λ be a lattice in the complex plane $\mathbb{C}$. That is, a rank-2 $\mathbb{Z}$-module. The quotient $\mathbb{C}/\Lambda$ is topologically a torus, i.e. genus 1. It turns out that it is an elliptic curve, in the sense that there is a meromorphic function $\wp(z)$ which is periodic modulo Λ such that

$$\mathbb{C}/\Lambda \mapsto E : y^2 = x^3 + ax + b$$

by the map

$$z \mapsto (\wp(z), \wp'(z)).$$

The function $\wp$ has a pole on Λ, which maps Λ to the point at infinity; note that the definition of $\wp$ depends on Λ so one sometimes writes $\wp(z, \Lambda)$. The coefficients a and b depend on Λ also.

The group law on the elliptic curve is inherited from addition on $\mathbb{C}$. For example, the two-torsion points are all the points $\frac{1}{2}\Lambda$.

I'm not going into the theory that develops this, but it is perhaps enlightening to compare it to a completely analogous case that you are familiar with. That is, consider a rank-1 $\mathbb{Z}$-module Λ in $\mathbb{R}$, i.e. $\Lambda = c\mathbb{Z} \subset \mathbb{R}$. Then $\mathbb{R}/\Lambda$ is a circle, topologically. There's a map

$$\mathbb{R}/\Lambda \mapsto C : x^2 + y^2 = 1$$

which is given by

$$r \mapsto (\sin(2\pi r/c), \cos(2\pi r/c))$$

The map depends on Λ.

The trigonometric functions, which are periodic in one dimension, can be discovered as inverses to arclength integrals in calculus class:

$$\sin^{-1}(x) = \int_0^x \frac{dt}{\sqrt{1-t^2}}.$$

Similarly, the elliptic curve case arises from the study of the arclength of ellipses, i.e.

$$\int \frac{dt}{\sqrt{(t-a)(t-b)(t-c)}}.$$

That's the reason for the (somewhat unfortunate) name.

What I'm interested in for now is that a lattice $\Lambda \subset \mathbb{C}$ gives an elliptic curve. In fact, every elliptic curve over $\mathbb{C}$ arises from some lattice, and two lattices give the same curve if and only if they are *homothetic*, meaning one is a (complex) scalar multiple of the other.

Theorem 63.1. *Write*

$$\mathcal{L} = \{\textit{lattices in } \mathbb{C}\}, \quad \mathcal{E} = \{\textit{elliptic curves over } \mathbb{C}\}.$$

Then there is a bijection

$$\frac{\mathcal{L}}{\mathbb{C}^*} \leftrightarrow \frac{\mathcal{E}}{\mathbb{C}-\text{isomorphism}}.$$

For our purposes, we have defined elliptic curves by their Weierstrass equations; two are isomorphic if

$$a' = u^4 a, \quad b' = u^6 b$$

The isomorphism classes are exactly parametrized by the j-invariant,

$$j = -1728\frac{(4a)^3}{\Delta}, \quad \Delta = -16(4a^3 + 27b^2).$$

One of the powerful tools to study elliptic curves is to study them as a collection, in the form of the collection of lattices modulo homothety.

64. The study of lattices in $\mathbb{C}$

We could describe a lattice by giving a basis:

$$\omega_1\mathbb{Z} + \omega_2\mathbb{Z}$$

But many ordered bases give the same lattice; any change of variables (element of $\mathrm{GL}_2(\mathbb{Z})$ will give a new basis. We can endow the lattice with an *orientation* endowed by its basis: if the basis vectors, in the order given, have angle $0 < \theta < 2\pi$, then it is *positively oriented*. If we wish to study all lattices without orientation, we may as well restrict to bases that give *negative* orientation[7], and then consider them up to $\mathrm{SL}_2(\mathbb{Z})$.

Thus, as we have seen before,

$$\frac{\{\text{lattices}\}}{\mathbb{C}^*} \leftrightarrow \frac{\{\text{negatively oriented bases}\}}{\mathbb{C}^* \cdot \mathrm{SL}_2(\mathbb{Z})}.$$

Here, the notation on the right hand side means we are quotienting out by homothety ($\mathbb{C}^*$) and by orientation changing change of basis ($\mathrm{SL}_2(\mathbb{Z})$).

By taking a homothety, we can rotate and scale the lattice. Hence we may assume its basis is of the form

$$\tau\mathbb{Z} + \mathbb{Z}$$

where τ is some element of the upper half plane:

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

We obtain

$$\frac{\{\text{lattices}\}}{\mathbb{C}^*} \leftrightarrow \frac{\{\tau \in \mathbb{H}\}}{\mathrm{SL}_2(\mathbb{Z})}.$$

But now we have to describe how $\mathrm{SL}_2(\mathbb{Z})$ acts on τ. It acts on $1, \tau$ by basis change. There are various ways to accomplish a basis change, all pretty much isomorphic[8]. Let's say it gives a new basis

$$a\tau + b, c\tau + d.$$

But now we wish to take a homothety so this is of the form

$$\tau', 1$$

In other words,

$$\tau' = \frac{a\tau + b}{c\tau + d}.$$

[7] negative is more convenient than positive

[8] I mean, for example, that you could apply the transpose first, which is an automorphism of SL_2, etc. etc.

To check that we've preserved negative orientation, we should check that

$$\mathrm{Im}(\tau') = \mathrm{Im}\left(\frac{a\tau+b}{c\tau+d}\right) = \frac{(ad-bc)\,\mathrm{Im}(\tau)}{|c\tau+d|^2}.$$

So indeed SL_2 acting in this way preserves orientation (and its complement in GL_2 would reverse it). The map

$$z \mapsto \frac{az+b}{cz+d}$$

is called a *Möbius transformation*.

To reiterate: $\mathbb{H}$ parametrizes bases up to homothety. But *lattices* up to homothety are parametrized by

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

We write $\mathrm{SL}_2(\mathbb{Z})$ on the left to keep track of the fact that it's a left action (before I wrote it as a fraction, which was a bit ambiguous).

To sum up:

Proposition 64.1. *Let $\Lambda = \omega_1\mathbb{Z}+\omega_2\mathbb{Z} \subset \mathbb{C}$ where ω_1, ω_2 are negatively oriented.*

(1) Any other negatively oriented basis for Λ is of the form

$$\omega_1' = a\omega_1 + b\omega_2$$
$$\omega_2' = c\omega_1 + d\omega_2$$

(2) Two lattices $\tau\mathbb{Z}+\mathbb{Z}$ and $\tau'\mathbb{Z}+\mathbb{Z}$ are homothetic if and only if

$$\tau' = \frac{a\tau+b}{c\tau+d}, \quad \text{for some } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

(3) There is a $\tau \in \mathbb{H}$ such that Λ is homothetic to $\tau\mathbb{Z}+\mathbb{Z}$.

Since

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

acts trivially on $\mathbb{H}$, we should actually be using $\mathrm{PSL}_2(\mathbb{Z})$ all along.

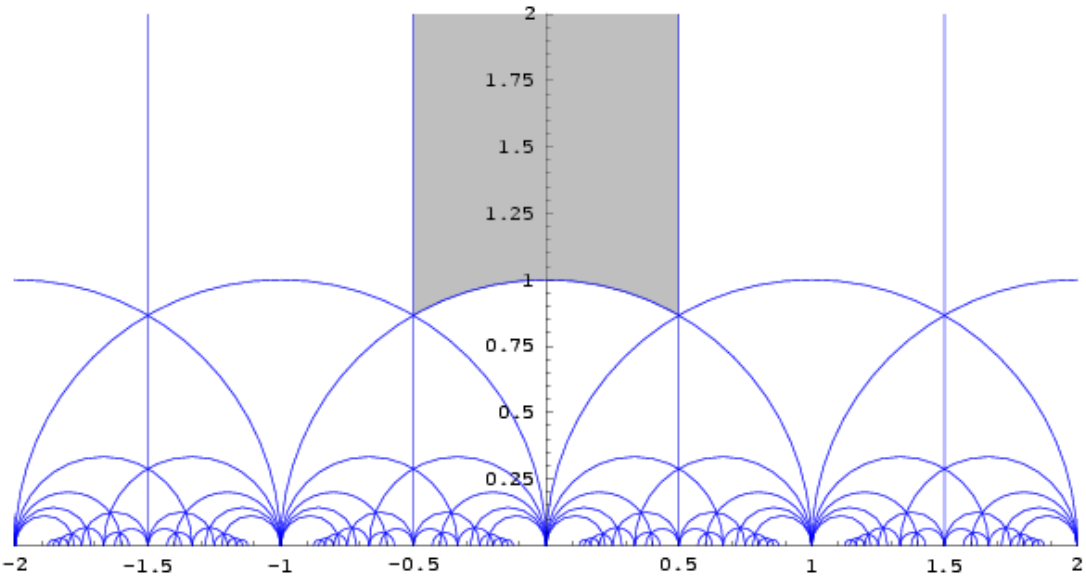
Theorem 64.2. *We have a bijection*

$$\frac{\{\textit{lattices}\}}{\mathbb{C}^*} \leftrightarrow \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}.$$

Recall from before that $\mathrm{PSL}_2(\mathbb{Z})$ is generated by the elements

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Here's the upper half plane, with the action of $\mathrm{PSL}_2(\mathbb{Z})$ (known in this context as the *modular group*) shown:



The image is from Wikimedia Commons. It shows in grey the fundamental domain; all the other white regions are images under the modular group. The element T represents translation by 1, while S takes the grey region to the adjacent white region inside the unit circle by inversion in that circle.

I won't prove that these are the regions; for a proof, see the first chapter of *Advanced Topics in the Arithmetic of Elliptic Curves*.

Remark: The topograph can be drawn over this picture by putting a vertex in each region and connecting two vertices if the regions are adjacent across a blue line. This is another picture of the topograph as a graph associated to $\mathrm{PSL}_2(\mathbb{Z})$.

65. The Modular Curve $X(1)$ – sketchily

It is traditional to write

$$\Gamma(1) = \mathrm{PSL}_2(\mathbb{Z})$$

in the context at hand. The quotient space $\Gamma(1)\backslash\mathbb{H}$ classifies lattices up to homothety, or, equivalently, elliptic curves over $\mathbb{C}$. Topologically, this looks like the grey region above; it's a punctured sphere in the sense that there is one point missing up 'at infinity'. Equivalently, one could use the white region below it; the missing point is on the real line, which is not included in the upper half plane. Without further explanation, I'll just add a point and call the resulting space $X(1)$, the *modular curve*. It is topologically a sphere. Each point on the modular

curve represents a single elliptic curve, except the point I added, called a cusp. The j-invariant gives an isomorphism

$$X(1) \to \mathbb{P}^1(\mathbb{C})$$

taking a point representing a curve to the j-invariant of that curve.

To study this space one studies the following objects:

Definition 65.1. *A* modular form of weight $2k$ *is a function on* $\mathbb{H}$ *which is everywhere holomorphic (on* $\mathbb{H}$ *and at infinity), and satisfies*

$$f(\gamma\tau) = (c\tau + d)^{2k} f(\tau)$$

for any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1)$.

For example, the coefficients of the Weierstrass equation, which depend on τ, are actually modular forms of weight 4 and 6. The j-invariant is the only modular form of weight 0 – and hence the only modular form which is actually a function on $\Gamma(1)\backslash\mathbb{H}$. The others are really differential forms on $X(1)$.

66. Fermat's Last Theorem

For more, look at Glenn Stevens' Overview from *Modular Forms and Fermat's Last Theorem* edited by Cornell, Silverman and Stevens.

Theorem 66.1 (Fermat's Last Theorem)**.** *Let* n *be an integer. Then the only integer solutions* (a, b, c) *to*

$$a^n + b^n = c^n$$

satisfy $abc = 0$.

Following Stevens, let's write $FLT(n)$ for the statement above, restricted to that integer n. It is clear that whenever $d \mid n$, $FLT(d) \implies FLT(n)$. So we just need to prove it for primes. We'll rearrange the equation to

$$a^p + b^p + c^p = 0.$$

Fermat (1640): FLT(4)

Euler (1760s): FLT(3)

There's a ton of interesting stuff to do with the history of FLT before the modern proof. I won't go into it.

Let (a, b, c) be a solution to FLT(p). Let $A = a^p$, $B = b^p$, $C = c^p$; then $A + B + C = 0$. Define an elliptic curve:

$$E_{A,B,C} = y^2 = x(x - A)(x - B)$$

We compute some of the basic quantities associated to this curve, like the discriminant, and it gives us arithmetic information about the curve. The arithmetic information is so wild that it can't be true.

67. Proof of Fermat's Last Theorem

There are other subgroups of $\Gamma(1)$ of interest, called *congruence subgroups*, e.g.

$$\Gamma(N) = \{\gamma \in \Gamma(1) : \gamma \equiv I \pmod{N}\}.$$

These (the above and others) give rise to other modular curves. Various modular curves parametrise such things as elliptic curves together with N-torsion points or cyclic subgroups of order N, etc.

In particular, if we define

$$\Gamma_0(N) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : c \equiv 0 \pmod{N}\}$$

Then the resulting curve, $X_0(N)$ parametrises elliptic curves together with a cyclic subgroup of order N. This modular curve has an explicit polynomial equation. An elliptic curve is called *modular* if this modular curve, for some N, covers the elliptic curve. Note that this is weird: the modular curve parametrizes the collection of elliptic curves, and then maps to our one elliptic curve of interest.

The main statement needed for Fermat's Last Theorem is that *all elliptic curves are modular*[9]. Being modular can also be interpreted in an analytic way. There's a sort of zeta function for the elliptic curve over $\mathbb{Q}$, called an L-function. And any modular form also has a function associated to it, a fourier expansion. An elliptic curve is modular if its L-function's coefficients are the same as the fourier expansion coefficients of some modular form. (The modular form gives us the covering map referred to above.)

Then one shows that the curve $E_{A,B,C}$ defined above cannot be modular, because the modular form it would be associated to has properties that can't exist.

68. Arithmetic Dynamics

For the last part of the course, I'm going to use notes of Silverman from the 2010 Arizona Winter School, and teach you some basics of Arithmetic Dynamics. The notes are available via a link on our website.

[9]actually, just *semistable* ones are needed, which is what Wiles proved, but we haven't even defined that