Math 124 - Number Theory

Taught by Cliff Taubes
Notes by Dongryul Kim

Fall 2018

¡+instructor+¿ ¡+meetingtimes+¿ ¡+textbook+¿ ¡+enrolled+¿ ¡+grading+¿ ¡+courseassistants+¿

Contents

1 September 5, 2018

There will be two textbooks for the course: *The Higher Arithmetic* by H. Davenport, and *Elementary Number Theory—primes, congruences, and secrets* by W. Stein written in a more modern perspective. There are going to be weekly reading and assignments. The homework is there to make sure you learn number theory, so if you are stuck, email me or the course assistant. You are welcome to collaborate with other students. To hand in your homework late, you need to get my permission. There will be no exams in this course, but instead, there will be midterm and final writing assignments. You will write a lecture on a topic that I did not go over in class.

1.1 Overview

We are going to denote

$$\begin{aligned}
\mathbb{N} &= \{\text{natural numbers}\} = \{1, 2, 3, 4, \dots\}, \\
\mathbb{Z} &= \{\text{integers}\} = \{\dots, -2, -1, 0, 1, 2, \dots\}, \\
\mathbb{Q} &= \{\text{rational numbers}\} = \{\tfrac{p}{q} : p, q \text{ integers with } q \neq 0\}, \\
\mathbb{R} &= \{\text{real numbers}\}, \\
\mathbb{C} &= \{\text{complex numbers}\}.
\end{aligned}$$

Number theory deals with $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$. But what the big deal with $1, 2, 3, \dots$? There are even animals that can count, and one article says that even the Venus fly trap can count.

The integers carry a structure of addition, and also a structure of multiplication. If we look at these structures, some interesting thins happen. Not every number is divisible by 7, and not every number is divisible by 10. But every integer can be written as

$$7x + 10y$$

where x, y are integers. For instance, $15 = 7 \times 5 + 10 \times (-2)$. On the other hand, not every integer can be written as

$$6x + 10y.$$

This is not hard to see, because $6x + 10y$ is always an even number.

So given positive integers n, m, what numbers can be written as $nx + my$? That is, what is the set

$$\{nx + my : x, y \in \mathbb{Z}\}?$$

There is another curious pattern here. If a is an integer not divisible by 3, then $a^2 - 1$ is divisible by 3. For instance, $10^2 - 1 = 99$ and $11^2 - 1 = 120$. This is kind of cool, but why is this? Here is something better. If a is an integer not divisible by 5, then $a^4 - 1$ is divisible by 5. This is called Fermat's little theorem. So is it true that for any positive integer n, is the following true?

> If a is not divisible by n, then $a^{n-1} - 1$ is divisible by n.

This turns out to be false. When $n = 4$ and $a = 2$, we have $a^{n-1} = 2^3 - 1 = 7$ not divisible by $n = 4$. But if we modify the statement a little bit and put $a^{\varphi(n)}$ instead of a^{n-1}, we get a true statement. So there are all these patterns coming from playing with numbers.

Here is another curious pattern. Consider $x^2 + 1$ where x is an integer.

Theorem 1.1. *No number of the form $x^2 + 1$ is divisible by a number of the form $4k + 3$ where $k \geq 0$.*

You can check this all day. Pick any integer $x^2 + 1$ and try dividing it by 7 or 11.

There are also interesting questions about rational numbers. Recall that a rational number is a number of the form p/q where p and q are integers with $q \neq 0$. But are all numbers rational? If you have a square tile of side length 1, the length of a side is $\sqrt{2}$ by Pythagorean's theorem. We can prove that $\sqrt{2}$ is not rational, by proof by contradiction. Suppose that $\sqrt{2}$ is rational, so that we can write

$$\frac{p}{q} = \sqrt{2}.$$

Here we can assume that p and q are not both even, because then we can cancel out the 2 in both p and q. We square both sides and get

$$\frac{p^2}{q^2} = 2, \quad p^2 = 2q^2.$$

But then p has to be an even integer, because p^2 is an even number. So we can write $p = 2k$. Then

$$4k^2 = 2q^2, \quad 2k^2 = q^2.$$

By the same reason, q also has to be an even number, and this contradicts our assumption that p and q are not both even.

Number theory is also used in codes and cyphers.

Definition 1.2. A **prime number** is a positive number not divisible by any smaller number except 1. By convention, 1 is not a prime number.

So the list of prime numbers is $2, 3, 5, 7, 11, 13, 17, 19, \ldots$. RSA encryption is is based on prime numbers. For two big prime numbers p and q, you look at $N = pq$ and make the number N public, but keep p and q hidden. The RSA encryption system is designed so that if you know the number N, you can encrypt any message, but to decrypt it, you need to know the prime numbers p and q. Factoring an integer into primes is a computationally difficult job, so the message is secure.

There is also interesting number theory in geometry. An elliptic curve is the set of solutions of

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

in the (x, y)-plane. There is a way of defining addition on elliptic curves, and this satisfies commutativity and associativity.

If there is time, we are also going to talk about how many primes there are.

1.2 Addition and multiplication of integers

Let us look at the set integers $\mathbb{Z}$. There is addition on the set of integer, and they satisfy commutativity and associativity:

$$x + y = y + x, \quad x + (y + z) = (x + y) + z, \quad x + 0 = x, \quad x + (-x) = 0.$$

There is also multiplication, which is just addition. They also satisfy some rules:

$$xy = yx, \quad x(yz) = (xy)z, \quad 1x = x.$$

Then there is a rule about distribution:

$$x(y + z) = xy + xz.$$

Let's assume we all know these rules and not be too pedantic.

There are also cancellation laws:

If $x + y = x + z$ then $y = z$.

You can say that this follows from negative numbers, but in a sense, this law is why we can define negative numbers. There is also cancellation law for multiplication:

If $xy = xz$ and $x \neq 0$, then $y = z$.

1.3 Divisibility and primes

Definition 1.3. We say that a **divides** b if $b = ma$ for some integer a.

If a divides b, then $|a| \leq b$ unless $b = 0$.

Proposition 1.4. *If a divides b and b divides c, then a divides c.*

Proof. If $b = ma$ and $c = nb$ then $c = (mn)a$. □

Proposition 1.5. *If b and c are divisible by a, then $xb + yc$ is also divisible by a.*

Proof. Write $b = ma$ and $c = na$. Then

$$xb + yc = xma + yna = (xm + yn)a.$$ □

We defined a **prime number** as a number only divisible by n and 1. We say that n is **composite** if n is neither a prime or 1. So every positive integer n falls in exactly one of the following categories:

- primes,
- composites,
- 1.

Prime numbers are important because they form sort of an irreducible basis for multiplication of integers.

Theorem 1.6 (fundamental theorem of arithmetic)**.** *Every number can be written as a product of primes which is unique up to order.*

The factors that appear in the factorization of a are called the **prime factors** of a. The existence part is not hard to prove. Take any natural number n. If $n = 1$ or n is a prime, there is nothing to do. Otherwise, n is composite and we can write

$$n = m \cdot q, \quad 1 < m, q < n.$$

If m and q are both primes, we are done. Otherwise, we factor them further as

$$n = m_1 \cdot m_1 \cdot q \text{ or } n = m \cdot q_1 \cdot q_2.$$

This process should end at a point, so we get a factorization of n into prime numbers.

This has a nice consequence.

Proposition 1.7. *There are infinitely many primes.*

Proof. Again we do proof by contradiction. Suppose there are N prime, and let them be $n_1, \ldots, n_N$. Then consider

$$\ell = n_1 n_2 \cdots n_N + 1.$$

This is a number, but it is not divisible by any of $n_1, \ldots, n_N$. This contradicts the existence part of the fundamental theorem of arithmetic. □

2 September 10, 2018

We were talking last time about the fundamental theorem of arithmetic.

Theorem 2.1 (fundamental theorem of arithmetic)**.** *Every positive integer n can be written as a product of primes*

$$n = p_1 p_2 \cdots p_k,$$

where p_i are all primes. Moreover, this is unique up to ordering.

By convention, 1 is neither prime or composite, and it is the product of zero primes. If n is prime, we are done. If n is composite, we can write $n = qr$, and then we reduce the problem to smaller cases. So this proves existence of a prime factorization. Uniqueness is harder, and Davenport has a clever proof that is not too enlightening.

2.1 Davenport's proof of the fundamental theorem of arithmetic

Suppose we have

$$n = pq \cdots t = p'q' \cdots t'.$$

Here, write this so that p is the smallest prime in $pq \cdots t$ and p' is the smallest prime in $p'q' \cdots t'$.

Assume that n the smallest case when n has two different representations into products of primes. Then p does not appear in $p'q' \cdots t'$, because otherwise we can cancel out p on both sides and get two different representations of n/p. Similarly, p' does not appear in $pq \cdots t$. Because p is the smallest prime,

$$n = pq \cdots t \geq p \cdot p,$$

and hence $p \leq \sqrt{n}$. Likewise, $p' \leq \sqrt{n}$ and so

$$pp' < \sqrt{n} \cdot \sqrt{n} = n.$$

(Here, we have strict inequality because it cannot be that $p = p' = \sqrt{n}$.)

Let us now consider

$$0 < m = n - pp' = p(q \cdots t - p') = p'(q' \cdots t' - p).$$

Because we assumed that n was the smallest positive integer with non-unique representation, and m is smaller than n, there is a unique prime factorization of m. Therefore p and p' both appear in this representation. So we can write

$$m = n - pp' = pp's \cdots uv.$$

Now

$$n = pp'(1 + s \cdots uv) = pq \cdots t$$

and we can cancel both sides and get

$$\tfrac{n}{p} = q \cdots t = p'(1 + s \cdots uv).$$

But n/p is smaller than n, and so there should be a unique representation. But p' does not appear in $q \cdots t$ but does appear in $p'(1 + s \cdots uv)$. This is a contradiction.

2.2 Greatest common divisor

This was a clever proof, but it doesn't really tell us much. We are now going to develop a new technology that can take us further than that.

Definition 2.2. The **greatest common divisor** $\gcd(a, b)$ of two integers a and b is the largest integer that divides both a and b. (Davenport calls it the highest common factor, but I have never heard it called by that name.)

So for instance,

$$\gcd(12, 15) = 3, \quad \gcd(12, 66) = 6, \gcd(8, 20) = 4.$$

You can compute the greatest common divisor by listing all divisors on both sides, and finding numbers that match. Let p be a prime number. The only divisors of p are 1 and p. So its greatest common divisor with any number is

$$\gcd(p, n) = \begin{cases} 1 & \text{if } p \text{ does not divide } n, \\ p & \text{if } p \text{ divides } n. \end{cases}$$

Definition 2.3. We say that two integers a and b are **relatively prime** if $\gcd(a, b) = 1$.

The greatest common divisor is an important concept, and we will establish some properties.

Lemma 2.4. $\gcd(a, b) = \gcd(b, a) = \gcd(a, -b) = \gcd(-a, -b) = \gcd(-a, b)$.

Proof. This follows from the fact that the list of divisors of n is the same as the list of divisors of $-n$. □

Lemma 2.5. $\gcd(a, b) = \gcd(a, a + b) = \gcd(a, b - a)$.

Proof. If m divides both a and b, then we can write $a = ml$ and $b = mk$. Then $b - a = m(l - k)$, and so m divides $b - a$. That is, m divides a and $b - a$. Conversely, if m' divides both a and $b - a$, then we can write $b - a = m'q$ and $a = m'l$ and $b = m'(l + q)$. That is, m' divides both a and b. This shows that the list of common divisors of a and b is equal to the list of common divisors of a and $b - a$. So when we take the greatest one, we get $\gcd(a, b) = \gcd(a, b - a)$. □

Lemma 2.6. $\gcd(a, b) = \gcd(a, b - na)$ *for any* n.

Proof. We can apply the previous lemma iteratively and get $\gcd(a, b) = \gcd(a, b - a) = \gcd(a, b - 2a) = \cdots$. $\square$

You can find the greatest common divisor by using the Euclidean algorithm. Suppose I have integers a, b as $0 < b < a$. Then there are unique integers q and r such that

$$a = qb + r, \quad 0 \leq r < b.$$

So using the lemma above, we get that if $a = qb + r$, then

$$\gcd(a, b) = \gcd(b, a) = \gcd(b, a - qb) = \gcd(b, r).$$

Example 2.7. Take $a = 18$ and $b = 14$. Then

$$\gcd(18, 14) = \gcd(14, 4).$$

They are both equal to 2.

So we have a algorithm here, called the **Euclidean algorithm**. Given $0 < b < a$, we write

$$a = q_1 b + r_1, \quad 0 \leq r_1 < b = 0, \quad \gcd(a, b) = \gcd(b, r_1).$$

Then we can do this for r_1 and b, and write

$$b = q_2 r_1 + r_2, \quad 0 \leq r_2 < r_1, \quad \gcd(b, r_1) = \gcd(r_1, r_2).$$

Then

$$r_1 = q_3 r_2 + r_3, \quad 0 \leq r_3 < r_2, \quad \gcd(r_1, r_2) = \gcd(r_2, r_3),$$

and so on. These r_i decreases, so at some point, we will have $r_n = 0$. Then

$$\gcd(a, b) = \cdots = \gcd(r_{n-1}, r_n) = \gcd(r_{n-1}, 0) = r_{n-1}.$$

Example 2.8. We have

$$\begin{aligned} 79 &= 66 + 13 \\ 66 &= 5 \times 13 + 1 \\ 13 &= 13 \times 1 + 0. \end{aligned}$$

So $\gcd(79, 66) = 1$.

Lemma 2.9. *Suppose a, b, n are positive integers. Then*

$$\gcd(na, nb) = n \gcd(a, b).$$

Proof. We know that $n \gcd(a, b)$ divides na and nb, because $\gcd(a, b)$ divides both a and b. So

$$\gcd(na, nb) \geq n \gcd(a, b).$$

But it's not obvious how to show the other direction.

So we use the Euclidean algorithm on both sides. If we know the process of the Euclidean algorithm for a and b, we can multiply the entire thing by n and get

$$\begin{aligned} na &= q_1(nb) + nr_1, \quad 0 \leq nr_1 < nb, \\ nb &= q_2(nr_1) + nr_2, \quad 0 \leq nr_2 < nr_1, \\ nr_1 &= \cdots . \end{aligned}$$

This means that $\gcd(na, nb) = nr_{k-1} = n \gcd(a, b)$. □

Lemma 2.10. *Suppose n divides a and b. Then n divides $\gcd(a, b)$ as well.*

Proof. Consider the Euclidean algorithm

$$a = qb + r_1, \quad b = q_1 r_1 + r_2, \quad \dots .$$

If n divides a and b, it divides $r_1 = a - qb$. If n divides r_1 and b, it divides $r_2 = b - q_1 r_1$. You repeat this process until you see that n divides $r_{k-1} = \gcd(a, b)$. □

Theorem 2.11 (Euclid's theorem)**.** *If p is a prime, and if p divides ab, then p divides either a or b.*

Proof. If p divides a, then we are done. So assume that p doesn't divide a so that $\gcd(a, p) = 1$. Then

$$\gcd(ab, pb) = b.$$

Now p divides both ab and pb, so p divides their greatest common divisor, which is b. □

Of course, if we know the fundamental theorem of arithmetic, we can write out and see this. But we are trying to prove the fundamental theorem of arithmetic.

Altenrative proof of the fundamental theorem of arithmetic. Let us write

$$n = pqr \cdots, \quad n = p'q'r' \cdots .$$

By assumption, p divides $n = p'(q'r' \cdots)$, and so p divides either p' or $(q'r' \cdots)$. If p divides p', then $p = p'$, otherwise we can write $q'r' \cdot = q'(r' \cdots)$ and do the same thing over and over. So p appears in $p'q'r' \cdots$ and then we can cancel them out. □

2.3 Linear combinations

Suppose we are given integers a and b. We are interested in what integers we can get by taking integer linear combinations of a and b, i.e., for which m does

$$ax + by = m$$

have a solution $x, y \in \{\dots, -2, -1, 0, 1, 2\}$.

The first thing we observe is that anything that divides both a and b also has to divide m. So we should be able to write

$$m = \gcd(a,b)z, \quad z \in \{\dots, -2, -1, 0, 1, 2\}.$$

But can you get all the multiples of $\gcd(a,b)$?

Proposition 2.12 (Bezout)**.** *The set of linear combinations of two integers is*

$$\{ax + by\}_{x,y \in \{\dots,-1,0,1,\dots\}} = \{k \gcd(a,b)\}_{k \in \{\dots,-1,0,1,\dots\}}.$$

For example, every integer can be written as $3x + 17y$, like $21 = 3 \times (-10) + 17 \times 3$. For now, let us postpone proving this theorem and use it to prove the fundamental theorem of arithmetic.

Another alternative proof of the fundamental theorem of arithmetic. Suppose we have

$$n = pq \cdots r = p'q' \cdots r'.$$

If $p = p'$, we can divide by p and continue. If $p \neq p'$, we can write

$$px + p'y = 1,$$

and so multiplying both sides by $q' \cdots r'$ gives

$$p(xp'q' \cdots r') + p(yq \cdots r) = q' \cdots r'.$$

So p divides $q' \cdots r'$ and we can continue this process. □

We now need to prove this linear algebra proposition.

Proof of Bezout's theorem. Given a and b, then we want find x and y such that $ax + by = \gcd(a,b)$. Then we can write any multiple of $\gcd(a,b)$ as

$$a(kx) + b(ky)k = \gcd(a,b).$$

Write $a' = a/\gcd(a,b)$ and $b' = b/\gcd(a,b)$. Then $\gcd(a',b') = 1$ and it is enough to find x and y such that

$$a'x + b'y = 1.$$

Let m be the smallest positive integer such that we can write

$$a'x + b'y = m.$$

We will finish this next time. □

3 September 12, 2018

We introduced the notion of a greatest common divisor. This is

$$\gcd(a,b) = \text{greatest integer that divides both } a \text{ and } b.$$

Then we proved some interesting things about the greatest common divisor:

- $\gcd(a,b) = \gcd(b,a) = \gcd(\pm a, \pm b)$
- $\gcd(a,b) = \gcd(a, b - na)$
- $\gcd(na, nb) = |n| \gcd(a,b)$
- If n divides a and b, then n divides $\gcd(a,b)$

We also proved that if p is prime and it divides ab, then it divides either a or b.

We were looking at the linear combinations of two integers,

$$\{ax + by\}_{x,y\in\mathbb{Z}},$$

which is the set of k such that $ax + by = k$ has an integer solution.

Proposition 3.1 (Bezout)**.** *The equation $ax + by = k$ can be solved if and only if k is divisible by* $\gcd(a,b)$.

This implies, for instance, that if a and b are relatively prime (i.e., $\gcd(a,b) = 1$) then you can find integers x, y such that $ax + by = 1$. Before we prove this, let me ask a sloppy mathematical question. If you take two integers at random, what is the probability that they are relatively prime? The statement is that the probability is

$$\frac{6}{\pi^2} \approx 0.608.$$

More precisely, you are choosing two integers at random in $\{1, 2, \dots, n\}$, and take n to ∞. Let me give you a heuristic that shows that this somewhat makes sense. The odds of two numbers not being both divisible by 2 is

$$1 - \frac{1}{2^2}.$$

Then the odds of them being not simultaneously divisible by 3 is

$$1 - \frac{1}{3^2}.$$

Then we go on with all primes,

$$\Big(1 - \frac{1}{4}\Big)\Big(1 - \frac{1}{9}\Big)\Big(1 - \frac{1}{25}\Big)\cdots = \prod_p \Big(1 - \frac{1}{p^2}\Big).$$

So the number gets smaller and smaller, but maybe it converges to some one number. Cleverly, Euler realized that this converges to $6/\pi^2$. We might get to these cool statements near the end of the course.

Let us now prove what we wanted to prove.

Proof. We only need to show that $ax + by = k$ can be solved if k is divisible by $\gcd(a, b)$. If we let

$$a' = \frac{a}{\gcd(a,b)}, \quad b' = \frac{b}{\gcd(a,b)},$$

we have $\gcd(a', b') = 1$. So it is enough to show that

$$a'x + b'y = 1$$

has a solution.

Here is how you can prove this, although it doesn't give you the solution. Consider the set

$$\{a'x + b'y\}_{x,y\in\mathbb{Z}}$$

and look at the smallest positive number m that can be written as $a'x+b'y = m$. Suppose that m is not 1. Then m cannot divide both a' and b', because a' and b' are relatively prime. Assume without loss of generality that m does not divide a'. We have $m < a'$, because otherwise $a'(x-1) + b'y = m - a'$ is a smaller positive integer than m. Now write

$$a' = mq + r, \quad 0 < r < m.$$

(We have $0 < r$ because m does not divide a'.) Then we get

$$a' = (a'x + b'y)q + r, \quad a'(1 - qx) + b'(-qy) = r.$$

This contradicts the fact that m is the minimal positive integer, because $r < m$. The only way out of this contradictory loop is when $m = 1$. $\square$

3.1 Finding primes

There is something called the **prime sieve** that lets you to list primes. Suppose we want to compute all primes less than n. There is how you do this.

1. [Initialize] We know that 2 is a prime number. We set

   $$X = \{3, 5, 7, \ldots \leq n\}, \quad P = \{2\}.$$

   X is the set where we search for primes, and P is the set of primes that we found.

2. Let p be the smallest element remaining in X. If $p > \sqrt{n}$, add all of X to P and terminate the program. If $p \leq \sqrt{n}$, add p to P.
3. From X, remove all elements divisible by p.
4. Go to Step 2.

When we pick p in Step 2, it has to be a prime, because in Step 3 we always throw away all the things divisible by smaller primes. So p in Step 2 cannot be divisible by any smaller prime, which means that it is a prime. But why do we terminate the program when $p > \sqrt{n}$? The reason is this. Let the set X be

$$X = \{p, q, r, \ldots \leq n\}$$

where $p > \sqrt{n}$. Then r cannot be divisible by anything smaller than p. It could be that $r = ab$, but then $ab = r \leq n$ so either $a \leq \sqrt{n} < p$ or $b \leq \sqrt{n} < p$. This means that r has to be eliminated before X reached this state. So r cannot be composite, so it has to be a prime.

There is another famous theorem of Dirichlet.

Theorem 3.2 (Dirichlet)**.** *There are infinitely many primes of the form $ax + b$ (for fixed a and b), if a and b are relatively prime.*

This really requires a lot of technology. But here is one case we can prove. Almost all primes are odd, and they are either of the form $4x - 1$ or $4x + 1$.

Theorem 3.3. *There are infinitely many primes of the form $4x - 1$.*

For instance, $3, 7, 11, 19, 23, 31, \ldots$.

Proof. Let $p_1, p_2, \ldots, p_n$ be primes of the form $4x_n - 1$. Now look at

$$K = 4p_1p_2 \cdots p_n - 1.$$

If this is prime, then it is bigger than any of $p_1, \ldots, p_n$, so we get a new prime. But may be it is composite, and has prime factorization

$$K = (4x_1 + 1)(4x_2 + 1) \cdots (4x_t + 1).$$

This cannot happen, because if you expand the right side, it takes the form of $4X + 1$. So this means that there is a prime of the form $4x_{n+1} - 1$ dividing K, and it cannot be any of $p_1, \ldots, p_n$ because p_i and K are relatively prime. That is, given any n primes we can find a new prime. $\square$

The prime number theorem tells us how rare or common primes are. If we define

$$\pi(x) = \#\text{prime numbers less than or equal to } x,$$

then the theorem states the following asymptotic behavior.

Theorem 3.4 (prime number theorem)**.** *We have*

$$\lim_{x \to \infty} \frac{\pi(x)}{x / \log x} = 1.$$

So primes are not too rare, and also not to common.

3.2 Groups and rings

There is applied number theory. There is something called a perfect shuffle of playing cards. When you have 52 cards, you split it into exactly 26 and 26 cards, and you place exactly one between another. This seems like a perfectly randomized shuffle, but the theorem is that if you do the perfect shuffle eight times, you get back to the original position.

We don't have the technology right now, so let me introduce the notion of a group.

Definition 3.5. A **group** is a (finite) set G, with a distinguished element $1 \in G$ and a binary operation

$$G \times G \to G, \quad (a, b) \mapsto ab,$$

such that

- $(ab)c = a(bc)$ for all $a, b, c \in G$,
- $1a = a1 = a$ for all $a \in G$,
- for any $a \in G$ there is b so that $ab = 1$ and $ba = 1$.

Here is an example that might be confusing. If you take $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ with addition $+$, we have

$$a + 0 = 0 + a = a, \quad a + (-a) = 0, \quad a + (b + c) = (a + b) + c.$$

So this is a group. Here is another example. We can look at the group of pemutations the deck of cards. You can compose permutations, and this is multiplication we use. Every permutation has an inverse, so it becomes a group. This is a bit scary if you think about it. If you do a shuffle, and then do the inverse shuffle, you get the original deck so you can cheat.

Definition 3.6. We say that a group G is an **abelian group** if $ab = ba$ for all $a, b \in G$.

There are groups that are not abelian. Permutation groups are generally not abelian. For instance, take the permutation group on three elements a, b, c. You can actually check this.

Definition 3.7. A **commutative ring** is a set R with two distinguished elements $0, 1 \in R$ with two binary operations

$$+, \cdot : R \times R \to R$$

such that

- $+$ with the identity 0 makes it an abelian group,
- $1 \cdot a = a \cdot 1 = a$ for all $a \in R$,
- $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in R$,

- $a \cdot b = b \cdot a$ for all $a, b \in R$,
- $a \cdot (b + c) = a \cdot b + a \cdot c$ for all $a, b, c \in R$.

But there are no multiplicative inverses.

So let me give you the simplest possible nontrivial ring. This is $R = \{0, 1\}$, and addition and multiplication are defined as

$$0 + 1 = 1, \quad 0 + 0 = 0, \quad 1 + 1 = 0,$$

and

$$0 \cdot 0 = 0, \quad 0 \cdot 1 = 0, \quad 1 \cdot 0 = 0, \quad 1 \cdot 1 = 1.$$

This is called $\mathbb{Z}/2\mathbb{Z}$, and 1 sort of represents "odd numbers" while 0 represents "even numbers".

4 September 17, 2018

Last time we introduced the concept of a group. This is a set G with a rule $G \times G \to G$ satisfying

$$a(bc) = (ab)c, \quad 1a = a1 = a,$$

and for every $a \in G$ there exists a b such that $ab = ba = 1$. A group is also called abelian if $ab = ba$. There are groups that are not abelian, for instance, the group of invertible 2×2 real matrices. But in this class, we are mostly look at abelian groups.

Example 4.1. Consider the positive rationals $\mathbb{Q}_{>0} = \{\frac{a}{b} : a, b \text{ positive integers}\}$. This is a group under multiplication, with inverse of $\frac{a}{b}$ being $\frac{b}{a}$. The integers $\mathbb{Z}$ also forms a group under addition, with inverse of a being $-a$.

We also got to the notion of rings. This is a set R with two binary operations $+$ and $\times$. We require that $(R, +, 0)$ is an abelian group, so that we have things like $a + (-a) = 0$. But multiplication is not going to be a group. We are going to require that

$$a \times (b \times c) = (a \times b) \times c, \quad 1 \times a = a \times 1 = a, \quad a \times (b + c) = a \times b + a \times c.$$

Multiplication need not be commutative, but we can further require that multiplication is commutative, in which case the ring is called a commutative ring. The point is that there need not be inverses for multiplication.

Example 4.2. $\mathbb{Z}$ is a commutative ring. We have addition like $3 + (-5) = -2$ and multiplication like $2 \times (-4) = -8$. This satisfies all the condition we listed above.

4.1 The congruence ring $\mathbb{Z}/n\mathbb{Z}$

The definition of a ring won't be very interesting if there is only one ring on earth. Fix a positive integer n. We are going to construct rings $\mathbb{Z}/n\mathbb{Z}$, also called the **congruence modulo n ring**.

Definition 4.3. If a and b are integers such that $a - b$ is divisible by n, then we say that a is **congruent** modulo n and write $a \equiv b \pmod{n}$.

We then define

$$\mathbb{Z}/n\mathbb{Z} = \{\text{equivalence class of integer congruent modulo } n\}.$$

So for $n = 3$, there are three classes

$$[0] = \{\dots, -3, 0, 3, 6, \dots\}, \quad [1] = \{\dots, -5, -2, 1, 4, \dots\}, \quad [2] = \{\dots, -4, -1, 2, 5, \dots\}.$$

In general, $\mathbb{Z}/n\mathbb{Z}$ has n congruence classes,

$$\mathbb{Z}/n\mathbb{Z} = \{[0], [1], [2], \dots, [n-1]\}.$$

We actually do this everyday. The hours are like $\mathbb{Z}/12\mathbb{Z}$. We are only trying to formalize this.

Proposition 4.4. *The set $\mathbb{Z}/n\mathbb{Z}$ with $(+, [0])$ and $(\times, [1])$ forms a ring.*

Let us now analyze the perfect shuffle. We have 26 cards on each side, and if we do the out-shuffle, we get

$$\begin{gathered} 1, 2, 3, 4, \dots, 26, 27, 28, 29, \dots, 52 \\ \downarrow \\ 27, 1, 28, 2, 29, 3, \dots, 51, 25, 52, 26. \end{gathered}$$

So the mth card becomes the $2m$th card, modulo 53 if we suppose that there is a Joker at the end of the deck. So the perfect out-shuffle is just

$$m \mapsto 2m \pmod{53}.$$

If we pick the representatives $\{0, 1, \dots, n-1\}$ of $\mathbb{Z}/n\mathbb{Z}$, then addition is

$$a + b = \begin{cases} a+b & \text{if } a+b \leq n-1, \\ a+b-n & \text{if } a+b \geq n. \end{cases}$$

We can also define multiplication as doing ordinary multiplication and looking at the remainder when divided by n,

$$a \cdot b = q \cdot n + r, \quad r \in \{0, \dots, n-1\}.$$

Then in $\mathbb{Z}/n\mathbb{Z}$, we have $a \cdot b = r$. The remainder of the product only depends on the remainders of each number, because

$$(a + xn)(b + yn) = ab + n(xb + ya + nxy).$$

Let us do some practice. We can try to compute the multiplication table in $\mathbb{Z}/5\mathbb{Z}$, and the result is in Table 1. Note that in this case, every nonzero element has a multiplicative inverse. If a ring satisfies this property, we call it a **field**. I don't know the etymology of this word. If it is a ring with a nice property, why not call it a bracelet?

$\times$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Table 1: Multiplication table for $\mathbb{Z}/5\mathbb{Z}$

It is not true that all $\mathbb{Z}/n\mathbb{Z}$ are fields. Let us take $n = 8$ for instance. Then we get Table 2. Here, some numbers like 1, 3, 5, 7 have inverses, but some numbers like 2, 4, 6 don't. So we need to figure out which have inverses and which don't.

$\times$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	0	2	4	6
3	0	3	6	1	4	7	2	5
4	0	4	0	4	0	4	0	4
5	0	5	2	7	4	1	6	3
6	0	6	4	2	0	6	4	2
7	0	7	6	5	4	3	2	1

Table 2: Multiplication table for $\mathbb{Z}/8\mathbb{Z}$

Let me do another example. We did the out shuffle, but let's now do the in-shuffle. In this case, it is convenient to renumber the cards to $0, 1, \ldots, 51$ instead of $1, 2, \ldots, 52$. Then the shuffle looks like

$$\begin{gathered} 0, 1, 2, \ldots, 25, 26, \ldots, 51 \\ \downarrow \\ 0, 26, 1, 27, 2, 28, \ldots, 25, 51. \end{gathered}$$

Here, we see that the first and last card never changes, and the rule is that the mth card becomes the $2m$th card modulo 51. That is, the in shuffle is

$$m \mapsto 2m \pmod{51}.$$

Now we can prove that the deck comes back to itself after 8 shuffles. If we iterate the in-shuffle k times, we get

$$m \mapsto 2m \mapsto 4m \mapsto \cdots \mapsto 2^k m \pmod{51}.$$

Here, we note that

$$2^8 = 256 = 5 \times 51 + 1 \equiv 1 \pmod{51}.$$

So if we shuffle it 8 times, the shuffle is $m \mapsto m \pmod{51}$.

If we did this for the out-shuffle, we needed to find when

$$2^k \equiv 1 \pmod{53}.$$

The smallest k happens to be 52. We have

$$2^{52} = 4503\,5996\,2737\,0496 = 8497357784815 \times 53 + 1.$$

(This is like a Visa card number.)

4.2 Linear equations modulo a number

Consider n a number, which we also call the modulus in the following situation. Suppose we want to solve the linear equation

$$ax \equiv b \pmod{n}.$$

If we change $x \mapsto x' = x + qn$, this doesn't change anything. So really we are looking for $x \in \{0, 1, \dots, n-1\}$. Actually, we already know the answer. This equation is equivalent to

$$ax + ny = b.$$

This is solvable for x and y if and only if b is divisible by $\gcd(a, n)$. For instance, if $\gcd(a, n) = 1$ then we can solve the equation for any b.

This is relevant to the question about when you can invert an element.

Lemma 4.5. *If p is prime, then every nonzero element in $\mathbb{Z}/p\mathbb{Z}$ has a multiplicative inverse.*

This means that $\{1, 2, \dots, p-1\}$ is an abelian group under multiplication, so $\mathbb{Z}/p\mathbb{Z}$ is a field. Do you believe this? I don't, so let us check this for $p = 7$. As we see in Table 3, everything can be multiplied by something to give anything else.

$\times$	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Table 3: Multiplication table for $\mathbb{Z}/7\mathbb{Z}$

On the other hand, suppose that n is not a prime. We know that $ax \equiv b \pmod{n}$ can be solved if and only if $\gcd(a, n)$ divides b. This means that if $b = 1$, it can be solve if and only if a and n are relatively prime. So the only elements in $\mathbb{Z}/n\mathbb{Z}$ that has a multiplicative inverse are exactly those that are relatively prime to n. We can see this for $n = 6$.

In general, we can look at

$$(\mathbb{Z}/n\mathbb{Z})^\times = \{a \in \mathbb{Z}/n\mathbb{Z} \text{ that are relatively prime to } n\}.$$

Now this does form a group under multiplication, and this is called the **group of units** in $\mathbb{Z}/n\mathbb{Z}$. You can ask why this set is even closed under multiplication. If a and b are relatively prime to n, then ab is relatively prime to n. You can see this if you apply the fundamental theorem of arithmetic to a, b, n, and compare factors.

Example 4.6. For $n = 8$, we have

$$(\mathbb{Z}/8\mathbb{Z})^\times = \{1, 3, 5, 7\}.$$

In this group, $3 \cdot 5 = 7$, $7 \cdot 3 = 5$, $a \cdot a = 1$ for all $a \in \mathbb{Z}/8\mathbb{Z}$.

4.3 Euler's totient function

Definition 4.7. For a positive integer n, we define

$$\varphi(n) = \text{number of elements in } \{1, \dots, n\} \text{ that are relatively prime to } n.$$

This is equal to the number of elements in the group $(\mathbb{Z}/n\mathbb{Z})^\times$.

For instance, if p is a prime, then $\varphi(p) = p - 1$. What about $n = pq$, where p and q are distinct primes? We need to count the numbers among

$$1, 2, \dots, pq$$

that are not divisible by both p and q. First throw away pq. Then numbers divisible by p are

$$p, 2p, 3p, 4p, \dots, p(q-1),$$

and the numbers divisible by q are

$$q, 2q, 3q, \dots, (p-1)q.$$

These numbers don't overlap, and so we get

$$\varphi(pq) = pq - 1 - (p-1) - (q-1) = pq - p - q + 1 = (p-1)(q-1).$$

5 September 19, 2018

If we look at integers modulo n, we say that $a \sim b$ are equivalent if $a - b$ is divisible by n. So this is modeling things that are cycling, like hours on a clock. If we take

$$\{0, 1, 2, \dots, n-1\},$$

this is a **complete set of residues**. This means that any integer is equivalent to exactly one element in the set modulo n. We can take other sets, for instance,

$$\{n, n+1, n+2, \dots, 2n-1\} \text{ or } \{n, 2n+1, 3n+2, \dots, n^2 + (n-1)\}$$

are complete set of residues, in the sense that no two elements are equivalent modulo n and every integer is equivalent to some element modulo n.

Adding numbers modulo n is something we do all the time. If we try to calculate what day of the week it is, 30 days from today, we are doing addition modulo 7. Multiplication might not be so familiar, but we can do this similarly. If $a \equiv 3 \pmod{12}$ and $b \equiv 5 \pmod{12}$, then we have $ab \equiv 3 \times 5 = 15 \equiv 3 \pmod{12}$. We also showed that a number $a \in \mathbb{Z}/n\mathbb{Z}$ has a (multiplicative) inverse if and only if a is relatively prime to n. Addition in $\mathbb{Z}/52\mathbb{Z}$ also can be thought of as a shuffling. If we break the deck into two sets and exchange them, like

$$\begin{gathered} 0, 1, 2, \dots, a-1, a, a+1, \dots, 51 \\ \downarrow \\ a, a+1, \dots, 51, 0, 1, 2, \dots, a-1, \end{gathered}$$

we are just applying the transformation

$$m \mapsto m - a \pmod{52}.$$

Even if we do this shuffle k times, it is just $m \mapsto m - ka \pmod{52}$. So if ka is a multiple of 52, the deck will come back to itself.

So with all these operations, we can say that $\mathbb{Z}/n\mathbb{Z}$ is a ring. A field was a ring with inverses for every nonzero number, so $\mathbb{Z}/p\mathbb{Z}$ is a field. (Actually, the notion of rings and fields will not be too important in this course. Don't try too hard to remember the precise definition.) The number $a \in \mathbb{Z}/n\mathbb{Z}$ has an inverse if and only if it is relatively prime. We defined

$$(\mathbb{Z}/n\mathbb{Z})^\times = \text{group of units in } \mathbb{Z}/n\mathbb{Z},$$

and we defined $\varphi(n)$ to be the size of this group. We computed

$$\varphi(p) = p - 1, \quad \varphi(pq) = (p-1)(q-1)$$

for distinct prime p, q last time.

Let me do another example. What is $\varphi(p^3)$, for instance? Among the numbers

$$1, 2, 3, \dots, p^3 - 1,$$

the numbers not relatively prime to p^3 are the numbers divisible by p. These are

$$p, 2p, 3p, \dots, p^3 - p = (p^2 - 1)p.$$

If we throw them away, we get

$$\varphi(p^3) = (p^3 - 1) - (p^2 - 1) = p^3 - p^2 = p^2(p - 1).$$

More generally, we will get

$$\varphi(p^k) = (p^k - 1) - (p^{k-1} - 1) = p^k - p^{k-1} = p^{k-1}(p - 1).$$

Now we know $\varphi(n)$ for n a power of a prime. There is this theorem, which we will prove later.

Theorem 5.1. *If a and b are relatively prime, then*

$$\varphi(ab) = \varphi(a)\varphi(b).$$

Because any integer is a product of prime powers, relatively prime to each other, this theorem tells us exactly how to calculate $\varphi(n)$ for all n.

5.1 Order of an element

Recall what happened for the perfect shuffle. To find how many times we need to perform a shuffle to get back, we needed to find the k such that

$$2^k \equiv 1 \pmod{51}, \quad 2^k \equiv 1 \pmod{53}.$$

We can do this is greater generality.

Let G be a group, so that there is multiplication $a \cdot b$ satisfying the axioms $a(bc) = (ab)c$ and $1a = a1 = a$ and existence of inverses. (Think of $G = (\mathbb{Z}/n\mathbb{Z})^\times$ if you're confused.) For each element $a \in G$, we can try to look at the smallest positive integer m such that

$$a^m = 1 \in G.$$

This smallest m is called the **order** of a. This number might not exist of the group G is infinite, but if G is finite, the order is always defined and is finite.

Example 5.2. Consider $(\mathbb{Z}/10\mathbb{Z})^\times = \{1, 3, 7, 9\}$. This is a group; everybody has an inverse, because $1 \cdot 1 = 3 \cdot 7 = 7 \cdot 3 = 9 \cdot 9 = 1$. Now we can look for the order of each element. We have

$$3^1 = 3, \quad 3^2 = 9, \quad 3^3 = 9 \cdot 3 = 7, \quad 3^4 = 7 \cdot 3 = 1.$$

This means that the order of $3 \in (\mathbb{Z}/10\mathbb{Z})^\times$ is 4. If we do this for all the elements, we get

$$\text{order}(1) = 1, \quad \text{order}(3) = 4, \quad \text{order}(7) = 4, \quad \text{order}(11) = 2.$$

Example 5.3. What about $(\mathbb{Z}/12\mathbb{Z})^\times = \{1, 5, 7, 11\}$? Here, we have that

$$\mathrm{order}(1) = 1, \quad \mathrm{order}(5) = 2, \quad \mathrm{order}(7) = 2, \quad \mathrm{order}(11) = 2.$$

So everything sort of cycles around.

Example 5.4. What if we take a prime, like $(\mathbb{Z}/7\mathbb{Z})^\times = \{1, 2, 3, 4, 5, 6\}$? We can calculate

$$\mathrm{order}(1) = 1, \quad \mathrm{order}(2) = 3, \quad \mathrm{order}(3) = 6, \quad \mathrm{order}(4) = 3, \quad \mathrm{order}(5) = 6, \quad \mathrm{order}(6) = 2.$$

Now let me make an argument that every element has a finite order, inside a finite group. Take any element $a \in G$, and look at the sequence

$$1, a, a^2, a^3, a^4, \dots.$$

Because the group is finite, this sequence should visit something that it visited before, at some point. Assume that

$$a^k = a^p$$

for $p < k$. Then we can multiply $(a^{-1})^p$ on both sides, and this gives

$$a^{k-p} = 1.$$

So the order of a must be finite. But there is a theorem of Lagrange gives more than that.

5.2 Lagrange's theorem

Theorem 5.5 (Lagrange)**.** *The order of any element of a finite group divides the size of the group.*

In $(\mathbb{Z}/53\mathbb{Z})^\times$, we had that $\mathrm{order}(2) = 52$. Lagrange's theorem tells us that $\mathrm{order}(2)$ has to divide 52, and it just turned out that the order is the maximal possible number. In principle, it could have been $\mathrm{order}(2) = 4$ maybe, but this did not happen. On the other hand, in $(\mathbb{Z}/51\mathbb{Z})^\times$, the order $\mathrm{order}(2)$ divides $\varphi(51) = 32$. We actually got $\mathrm{order}(2) = 8$ in this case, and so we were able to conveniently cheat by doing only 8 shuffles.

Proof. Let us write $a \in G$ and $\mathrm{order}(a) = m$. Then we can look at the subset

$$\{1, a, a^2, \dots, a^{m-1}\} \subseteq G.$$

This is this not all of G, we can pick $b_1 \notin \{1, \dots, a^{m-1}\}$ and look at the subset

$$\{b_1, ab_1, a^2b_1, \dots, a^{m-1}b_1\}.$$

These are all distinct, and there is no overlap because if $a^k b_1 = a^p$ then $b_1 = a^{p-k}$ contradicts our assumption. If these two subset exhaust all of G, then we can pick another element $b_2 \in G$ not listed above. Then we again look at

$$\{b_2 ab_2, a^2b_2, \dots, a^{m-1}b_2\},$$

and do this over an over again. Once are done, we see that we divided G into groups of size m. So m divides the size of G. $\square$

Here is another very clever proof, assuming that G is abelian, e.g., for $G = (\mathbb{Z}/n\mathbb{Z})^\times$.

Proof. Let us write
$$G = \{1 = a_1, a_2, a_3, \dots, a_N\},$$
where N is the size of N. For $x \in G$, I want to look at the product
$$P = (xa_1) \cdot (xa_2) \cdot \cdots \cdot (xa_N) = x^N (a_1 a_2 \cdots a_N) \in G.$$
But on the other hand, we have
$$xa_1 = a_k, \quad xa_2 = a_l, \quad xa_3 = a_m, \quad \dots$$
while something like $a_k = a_l$ cannot happen because then $a_1 = x^{-1}a_k = x^{-1}a_l = a_2$. This means that if I multiply them all together, we get
$$P = (xa_1) \cdot (xa_2) \cdot \cdots \cdot (xa_N) = a_1 \cdot a_2 \cdot \cdots \cdot a_N.$$
Combining the two equations for P, and canceling out $a_1 a_2 \cdots a_N$, we get
$$x^N = 1.$$
This shows that order(m) divides N, because otherwise we can write $N = mq + r$ and get $x^r = x^N (x^m)^{-q} = 1$. □

If we apply this to $(\mathbb{Z}/n\mathbb{Z})^\times$, we get the following fact. If x is relatively prime to n, then
$$x^{\varphi(n)} \equiv 1 \pmod{n}.$$
If n is prime, we get the following theorem.

Theorem 5.6 (Fermat's little theorem)**.** *If p is prime and x does not divide p, then*
$$x^{p-1} \equiv 1 \pmod{p}.$$

6 September 24, 2018

We can think of $(\mathbb{Z}/n\mathbb{Z})^\times$ as the numbers that are relatively prime to n. If $n = p$ is prime, this is just

$$\{1, 2, \dots, p-1\},$$

which is an interesting case. If $a \in (\mathbb{Z}/n\mathbb{Z})^\times$, we call the smallest m with $a^m \equiv 1 \pmod{n}$ the order of a. We showed that

$$a^{\varphi(n)} \equiv 1 \pmod{n},$$

which implies that the order of every a divides $\varphi(n)$. If n is a prime, we get Fermat's little theorem, that

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

if p is a prime that does not divide a. So for instance, $10^6 - 1 = 999\,999$ is divisible by 7.

Corollary 6.1. *We have*

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

for all $a \in \{1, 2, \dots, p-1\}$ if and only if p is prime.

Proof. We know this for p a prime. Suppose p is composite, so that $p = rs$ with $1 < r, s < p$. Is it possible that $r^{p-1} - 1 = qp$? No, because both r^{p-1} and qp are divisible by r. $\square$

This is important in coding. There is a Rabin–Miller algorithm that takes a number and tells you whether it is probably a prime.

6.1 Wilson's theorem

There is another theorem about primes.

Theorem 6.2 (Wilson)**.** *For $p > 2$ an integer*

$$(p-1)! = 1 \cdot 2 \cdot \dots \cdot (p-1) \equiv (-1) \pmod{p}$$

if and only if p is a prime.

Actually this does not give an efficient algorithm for primality testing, because to compute $(p-1)!$, you need to actually perform $p - 2$ multiplications.

Proof. Suppose p is prime. Then for every element $a \in \{2, 3, \dots, p-2\}$, there is an inverse. So these things pair up with the inverse. Here, note that none of them are their own inverses, because $x \equiv x^{-1} \pmod{p}$ means $x^2 - 1 \equiv 0 \pmod{p}$, and $x^2 - 1 = (x+1)(x-1)$ implies that this is possible only for $x \equiv \pm 1 \pmod{p}$. Then when we pair up, they vanish, so

$$(p-2)! \equiv 1 \cdot (p-1) \cdot (1 \cdot \dots \cdot 1) \equiv -1 \pmod{p}.$$

On the other hand, if p is not a prime, so that $p = rs$, then r divides $(p-1)!$ because r is going to be one factor. It cannot be that r divides p and p divides $(p-1)! + 1$, because r divides $(p-1)!$. $\square$

We can check this:

$$4! = 24 \equiv -1 \pmod 5, \quad 5! = 120 \equiv 0 \pmod 6,$$
$$6! = 720 \equiv -1 \pmod 7, \quad 7! = 5040 \equiv 0 \pmod 8.$$

So you can impress people at the party using this.

6.2 Rabin–Miller and Euclidean algorithms

Here is how you run the **Rabin–Miller algorithm**, which tests if p is prime.

1. Look at the number p, and write it as $p-1 = 2^k \cdot m$, where m is odd. (Assume that p is odd, because that is when primality testing is interesting. Then $k \geq 1$.)
2. Pick an integer $1 \leq a \leq p-1$ at random.
3. Compute $b = a^m \pmod p$. If $b \equiv \pm 1 \pmod p$, then output "probably prime" and stop.
4. Compute b^{2^r} for $r = 1, 2, \dots, k-1$. If $b^{2^r} \equiv -1 \pmod p$ for some r, then output "probably prime" and stop.
5. Otherwise, output "not prime".

All steps seems to take at most $\log_2 p$ computations, except for computing $b = a^m \pmod p$. This seems to take m steps, but is there a more efficient way? We first write m in the base 2 expansion,

$$m = 2^k + \epsilon_{k-1} 2^{k-1} + \cdots + \epsilon_1 2^2 + \epsilon_0$$

with $\epsilon_j \in \{0, 1\}$. (Then k is around $\log_2 m$.) To compute a^m, we first compute the powers

$$a^1, a^2, a^4, a^8, \dots, a^{2^k} \pmod p$$

by iterative squaring. This can be done using not much memory, because we can take the remainder modulo p after each squaring. So this only requires something like $k \sim \log_2 m$ multiplications. Then we take

$$a^m = (a^{2^k}) \cdot (a^{2^{k-1}})^{\epsilon_{k-1}} \cdot \cdots \cdot (a^2)^{\epsilon_1} \cdot (a)^{\epsilon_0}.$$

This takes at most $k \sim \log_2 m$ additional multiplications. So the entire process takes at most $C \log_2 p$ time, where C is some constant not depending on p. This is pretty fast. This whole subject of unbreakable codes is based on calculation of primes.

Here is another algorithm that was in the book. Recall that we proved that there exist integers x and y such that

$$ax + by = \gcd(a, b).$$

Stein's book gives a proof by finding an algorithm for doing this. This is based on the Euclidean algorithm. Given $a < b$, we look at

$$\begin{aligned} a &= qb + r_1, \\ b &= q_2 r_1 + r_2, \\ r_1 &= q_3 r_2 + r_3, \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n, \\ r_{n-1} &= q_{n+1} r_n. \end{aligned}$$

Now we can read this as r_1 is a linear combination of a and b, and then r_2 is a linear combination of b and r_1 so it is also a linear combination of a and b. We can use this to inductively go down the equations and see that $r_n = \gcd(a, b)$ is a linear combination of a and b.

This algorithm is really efficient. To see this, note that $r_1 \leq \frac{1}{2}a$, because $r_1 \leq a - b$ and $r_1 \leq b$. (Adding these give $2r_1 \leq a$.) Now we can do this for each equation, and we get

$$r_1 \leq \frac{1}{2}a, \quad r_2 \leq \frac{1}{2}b, \quad r_3 \leq \frac{1}{2}r_1, \quad \cdots.$$

So there are at most $2\log_2 a$ steps in the Euclidean algorithm. This means that it is pretty fast.

6.3 Sunzi's remainder theorem

Now I am going to introduce a theorem about relatively prime numbers. It is also called the **Chinese remainder theorem**, because it is unclear if Sunzi first discovered it or just wrote down common knowledge.

Theorem 6.3 (Sunzi's remainder theorem)**.** *Let a and b be integers, and n and m be positive integers with $\gcd(n, m) = 1$. Then there is an integer x such that*

$$x \equiv a \pmod{m} \quad \text{and} \quad x \equiv b \pmod{n}.$$

Moreover x is unique modulo mn (up to adding multiples of mn).

So you can solve double congruence equations.

Proof. For the first congruence, we can just write $x = a + mq$. Now the question is, can we choose q so that the second congruence is satisfied? The second equation can be written as

$$a + mq + sn = b.$$

Then because m and n are relatively prime, we can find q and s such that $mq + sn = b - a$. Moreover, the solution is unique up to $q \to q + cn$ and $s \to s - cm$. This shows that $x = a + mq$ is unique up to adding multiples of mn. $\square$

What this means is that if I look at

$$\mathbb{Z}/21\mathbb{Z} = \{0, 1, 2, \ldots, 20\},$$

then specifying an element of $\mathbb{Z}/21\mathbb{Z}$ is the same as specifying an element of $\mathbb{Z}/3\mathbb{Z} = \{0, 1, 2\}$ and specifying an element of $\mathbb{Z}/7\mathbb{Z} = \{0, 1, 2, 3, 4, 5, 6\}$. We can make a table out of this:

	0	1	2	3	4	5	6
0	0	15	9	3	18	12	6
1	7	1	16	10	4	19	13
2	14	8	2	17	11	5	20

Table 4: Chinese remainder theorem for $\mathbb{Z}/21\mathbb{Z} \cong (\mathbb{Z}/3\mathbb{Z}) \times (\mathbb{Z}/7\mathbb{Z})$

It is saying in general, in terms of rings, there is a one-to-one map

$$\mathbb{Z}/nm\mathbb{Z} \to (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z}); \quad x \mapsto (a, b) = ([x]_{\text{ mod } m}, [x]_{\text{ mod } n}).$$

This is even a homomorphism of rings, i.e., they carry around the ring structure. So we also get

$$(\mathbb{Z}/nm\mathbb{Z})^\times \cong (\mathbb{Z}/n\mathbb{Z})^\times \times (\mathbb{Z}/m\mathbb{Z})^\times$$

as groups. If we count the number of elements on both sides, we learn a new thing. If n and m are relatively prime, then

$$\varphi(mn) = \varphi(m)\varphi(n).$$

So if we write any integer n as

$$n = p_1^{k_1} p_2^{k_2} \cdots,$$

then we can compute

$$\varphi(n) = \varphi(p_1^{k_1})\varphi(p_2^{k_2}) \cdots = p_1^{k_1-1}(p_1 - 1)p_2^{k_2-1}(p_2 - 1) \cdots.$$

7 September 26, 2018

We looked at Sunzi's theorem, which characterizes $\mathbb{Z}/ab\mathbb{Z}$ for a and b relatively prime. We showed that as a ring, this is isomorphic to

$$\mathbb{Z}/ab\mathbb{Z} \cong (\mathbb{Z}/a\mathbb{Z}) \times (\mathbb{Z}/b\mathbb{Z}).$$

This means that given integers y and z, there is a unique $x \pmod{ab}$ such that

$$x \equiv y \pmod{a}, \quad x \equiv z \pmod{z}.$$

This even implies that

$$(\mathbb{Z}/ab\mathbb{Z})^\times \cong (\mathbb{Z}/a\mathbb{Z})^\times \times (\mathbb{Z}/b\mathbb{Z})^\times; \quad x \mapsto ([x]_a, [x]_b)$$

is an isomorphism. If we count elements, we get $\varphi(ab) = \varphi(a)\varphi(b)$.

7.1 Primitive roots

We know that $(\mathbb{Z}/p\mathbb{Z})^\times = \{1, 2, \dots, p-1\}$ for p a prime, but we don't know a lot about the multiplicative group structure.

Theorem 7.1. *If p is a prime, there are $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ that generate $(\mathbb{Z}/p\mathbb{Z})^\times$, that is,*

$$(\mathbb{Z}/p\mathbb{Z})^\times = \{1, a, a^2, \dots, a^{p-2}\}.$$

In such a case, a is called a **primitive root**. This also can be stated as that there is an element of $(\mathbb{Z}/p\mathbb{Z})^\times$ with order equal to $p-1$. (Recall that it always divides $p-1$.) So this group is really simple, given that we can find this element a. Let us look at examples.

- Take $(\mathbb{Z}/2\mathbb{Z})^\times = \{1\}$. That works.
- What about $(\mathbb{Z}/3\mathbb{Z})^\times = \{1, 2\}$? We have $2^2 = 1$ so $a = 2$ satisfies this.
- For $(\mathbb{Z}/5\mathbb{Z})^\times = \{1, 2, 3, 4\}$, we have $2^2 = 4, 2^3 = 3$ so we have everybody for $a = 2$. We find that $a = 3$ also works.
- For $(\mathbb{Z}/7\mathbb{Z})^\times = \{1, 2, 3, 4, 5, 6\}$, we have $2^2 = 4, 2^3 = 1$, so it fails. We have $3^2 = 2, 3^3 = 6, 3^4 = 4, 3^5 = 5$ and so $a = 3$ works.

The algorithm is basically, try 2, then try 3, and so on.

The proof is based on that $\mathbb{Z}/p\mathbb{Z}$ is a field. Basically if $ab = 0$ then $a = 0$ or $b = 0$.

Theorem 7.2. *Suppose k is a field. Consider solving the polynomial equation*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$$

for $x \in k$, where $a_i \in k$ and $a_n \neq 0$. Then there are at most n solutions.

This works because k is a field. For instance, $x^2 - 1 = 0$ has 4 solutions in $\mathbb{Z}/8\mathbb{Z}$, which is not a field.

Proof. We do induction on $n = \deg f$. If $n = 1$, we have $a_1 x + a_0 = 0$, so we have $x = a_1^{-1} a_0$. (Recall that $a_1 \neq 0$ by assumption and any nonzero element of k has an inverse.)

Let's now look at $a_n x^n + \cdots + a_1 x + a_0 = 0$. If there are no solutions, we are done. Otherwise, let α be one solution to this equation. Then $f(x) = 0$ is equivalent to

$$a_n(x^n - \alpha^n) + a_{n-1}(x^{n-1} - \alpha^{n-1}) + \cdots + a_1(x - \alpha) = 0.$$

But then we can factor $x^q - y^q = (x-y)(x^{q-1} + \cdots + y^{q-1})$. Then we can write equation as

$$(x-\alpha)\big[a_n(x^{n-1} + \alpha x^{n-2} + \cdots) + a_{n-1}(x^{n-2} + \cdots) + \cdots + a_1\big] = 0.$$

Here, the thing in the bracket is a polynomial in x of degree $n-1$. By the induction hypothesis, that polynomial has at most $n-1$ solutions.

If x is a solution to $f(x) = 0$, then either $x - \alpha = 0$ or $[\cdots] = 0$ because k is a field. That is, either $x = \alpha$ or x is one of the roots of $[\cdots] = 0$. This shows that $f(x)$ has at most $1 + (n-1) = n$ solutions. □

By the way, there are all kinds of interesting fields number theorists introduce. For instance

$$\{a + bi \in \mathbb{C} : a, b \in \mathbb{Q}\} \text{ or } \{a + b\sqrt{2} : a, b \in \mathbb{Q}\}$$

are fields. We have

$$(a + b\sqrt{2})(a' + b'\sqrt{2}) = (aa' + bb') + (ab' + a'b)\sqrt{2}.$$

Now we have this algebraic fact. We know that $x^{p-1} - 1 = 0$ has exactly $p-1$ solutions in $\mathbb{Z}/p\mathbb{Z}$.

Lemma 7.3. *If p is a prime and d divides $p-1$, then*

$$x^d - 1 = 0$$

has exactly d solutions in $\mathbb{Z}/p\mathbb{Z}$.

Proof. There is some e such that $d \cdot e = p - 1$. So we write $x^{p-1} - 1$ as

$$0 = x^{p-1} - 1 = (x^d)^e - 1 = (x^d - 1)((x^d)^{e-1} + (x^d)^{e-2} + \cdots + 1).$$

Here, $x^d - 1$ has at most d solutions, $(\cdots)$ has at most $d(e-1)$ solutions, but are exact $p-1$ solutions to $x^{p-1} - 1$. This shows that $x^d - 1$ has to have exactly d solutions and $(\cdots)$ has to have exactly $d(e-1)$ solutions. □

So let's see how we can prove something. Let's assume that $p - 1 = 2q$ where q is an odd prime, for instance. The solutions to the equation $x^2 - 1 = 0$ are ± 1, so there are two. There are going to be q solutions to $x^q - 1 = 0$, and 1 is a common solutions. So there are going to be $q + 2 - 1 = q + 1$ numbers that are solutions to either $x^2 - 1$ or $x^q - 1$. This leaves out $q - 1$ numbers, whose order must be $2q = p - 1$.

Example 7.4. If we look at $p = 7$ and $q = 3$. Then the solutions of $x^2 - 1 = 0$ are $\{1, 6\}$, and the solutions of $x^3 - 1 = 0$ are $\{1, 2, 4\}$. So 5, which is left out, should be a primitive root of 7.

In the general case, let us write $p - 1 = q_1^{k_1} \cdots q_l^{k_l}$. Then the equation

$$x^{q^k} - 1 = 0$$

has q^k solutions, and $x^{q^{k-1}} - 1 = 0$ has q^{k-1} solution. So there are $q^{k-1}(q-1)$ many x such that $x^{q^k} - 1 = 0$ but $x^{q^{k-1}} - 1 \neq 0$. That is, there are elements $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that the order of a is $q_i^{k_i}$.

Lemma 7.5. *In an abelian group, if a has order m and b has order n and m and n are relatively prime, then ab has order mn.*

Proof. Note that

$$(ab)^{mn} = (a^m)^n (b^n)^m = 1^n 1^m = 1,$$

so the order divides mn. So we can write the order as rs, where r divides m and s divides n. Write $m = rr'$ and $n = ss'$. Then

$$1 = 1^{r'} = ((ab)^{rs})^{r'} = a^{ms} b^{ms} = b^{ms}.$$

Because b has order n, this implies that n divides ms, and so n divides s. This shows that $n = s$. Similarly, we get $m = r$. Therefore the order of ab is mn. □

We showed above that we can find elements

$$a_1, a_2, \ldots, a_l \in (\mathbb{Z}/p\mathbb{Z})^\times$$

such that a_i has order $q_i^{k_i}$. So then the order of their product

$$a = a_1 a_2 \cdots a_l$$

is $q_1^{k_1} \cdots q_l^{k_l} = p - 1$, by the above lemma. This shows that a is a primitive root. In fact, you can show that there are exactly $\varphi(p-1)$ primitive roots, if you are careful with counting.

So what is the probability that a random number in $(\mathbb{Z}/p\mathbb{Z})^\times$ will be a primitive root? We can calculate this by

$$\begin{aligned} \frac{\varphi(p-1)}{p-1} &= \frac{q_1^{k_1-1}(q_1 - 1) q_2^{k_2-1}(q_2 - 1) \cdots q_l^{k_l-1}(q_l - 1)}{q_1^{k_1} q_2^{k_2} \cdots q_l^{k_l}} \\ &= \frac{q_1 - 1}{q_1} \cdots \frac{q_l - 1}{q_l} = \Big(1 - \frac{1}{q_1}\Big)\Big(1 - \frac{1}{q_2}\Big) \cdots \Big(1 - \frac{1}{q_l}\Big). \end{aligned}$$

This is going to be pretty big, so you will probably find a primitive root pretty soon.

Theorem 7.6. *For an odd prime p, the group $(\mathbb{Z}/p^k\mathbb{Z})^\times$ is a cyclic group as well, that is, there is primitive root for p^k.*

Proof. I'm not going to prove it. $\square$

Now we're going to move to an applied topic. There is a cryptography called the RSA, and it uses the fact that if p and q are primes, then

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

for any $a \in (\mathbb{Z}/pq\mathbb{Z})^{\times}$.

8 October 1, 2018

Let p be a prime. A primitive root is, we defined, an element $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ that has order $p-1$, so that

$$(\mathbb{Z}/p\mathbb{Z})^\times = \{1, a, a^2, \dots, a^{p-2}\}.$$

There is a famous conjecture called Artin's conjecture.

Artin's conjecture. *For every integer $a \neq -1$ that is not a perfect square, there are infinitely many primes p such that a is a primitive root for p.*

Even if you can prove this for $a = 2$, you will become famous. The reason we are excluding -1 is because $(-1)^2 = 1$, and the reason we are excluding $a = x^2$ is because $a^{(p-1)/2} = x^{p-1} \equiv 1 \pmod{p}$.

8.1 Secret key exchange

Suppose my son wants to use my Amazon account, and I want to send him my password. But someone might be reading my emails. So here is what I do. First choose a prime number $p \approx 100^{600}$ of 600 digits, and a small number g, and share them over email.

- Now I choose a number m and my son (H) chooses n.
- I send H the number $g^m \pmod{p}$.
- H sends me the number $g^n \pmod{p}$.
- I compute the number $g^{mn} = (g^n)^m \pmod{p}$.
- H computes the number $g^{mn} = (g^m)^n \pmod{p}$.

So at the end, we share the number $g^{mn} \pmod{p}$. What the eavesdropper know is

$$p, \quad g, \quad g^n, \quad g^m.$$

But it is hard to compute n from g^n. This is called the **discrete logarithm problem**. Given $g \in G$ and $b = g^n \in G$, we can't efficiently calculate the number n. One thing you can try is enumerate the power of g,

$$g, g^2, g^3, \dots, g^n = b$$

until b appears, but it will take n time, where n is of the order 10^{600}. On the other hand, taking the nth power has an efficient algorithm. This is because we can do binary expansion and compute $g^1, g^2, \dots, g^{2^k}$.

In calculus, we have $g^{x+\Delta x} \sim g^x + O(\Delta x)$. But in the discrete logarithm problem, the number $g^{x+\Delta x}$ changes drastically for a small Δx. For instance, if $p = 17$ and $g = 3$, we get the sequence

$$3, 9, 10, 13, 5, 15, 11, 16, 14, 8, 7, 4, 12, 2, 6.$$

So it just bouncing around, almost randomly.

Now suppose there are three people, let's say me and H and A. Again, choose a prime p of 600 digits and a number $g \in \{1, 2, \dots, p-1\}$, and then share them over email with everyone.

- I pick m, H picks n, and A picks k.
- I send H the number g^m, then H sends A the number $(g^m)^n = g^{mn}$, and then A calculates the number $(g^{mn})^k = g^{mnk}$.
- H sends A the number g^n, then A sends me the number g^{nk}, then I compute $(g^{nk})^m = g^{mnk}$.
- A sends me the number g^k, then I send H the number g^{mk}, and then H computes $(g^{mk})^n = g^{mnk}$.

At the end, we all share the number $s = g^{mnk}$. That's pretty clever.

But here is what the eavesdropper (E) can do, if E can intercept the message rather than just eavesdropping.

- E picks a number t.
- I think I am sending H the number g^m, but actually E receives the number.
- E gets the number g^m, and sends g^t instead to H, so that H thinks g^t came from me.
- H thinks that the secret key is g^{tm}.
- Similarly, E intercepts H's message and sends g^t instead to me.
- I think that the secret key is g^{tn}.

$$C \underset{g^t}{\overset{g^m}{\rightleftarrows}} E \underset{g^n}{\overset{g^t}{\rightleftarrows}} H$$

But at this point, E knows both the number g^{tm} and g^{tn}. So when we try to encode messages using the secret key s, what E can do is to intercept all messages, decode with respect to one key and encode with respect to another key.

8.2 RSA cryptography

This is based on the fact that it is really hard to factorize integers. Pick two primes p and q of the order 10^{600}, and compute the public key $N = pq$. This number N is displayed in public, although p and q are kept secret. I also pick a number c that is relatively prime to $(p-1)(q-1)$, and make c public as well.

Now we consider

$$(\mathbb{Z}/pq\mathbb{Z})^\times = \{\text{relatively prime to } N \text{ between } 1 \text{ and } pq-1\}.$$

If a is relatively prime to N, we have

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

by Euler's theorem. Because c is chosen to be relatively prime to $(p-1)(q-1)$, I can solve

$$cd + (p-1)(q-1)y = 1$$

using the Euclidean algorithm, so I can compute the inverse $cd \equiv 1 \pmod{(p-1)(q-1)}$ relatively efficiently. This d is kept secret, and people can't do the same computation efficiently because $(p-1)(q-1)$ is kept secret (although $N = pq$ is public).

If A wants to encode some message and send this to me, A can first change the message to a number a smaller than 10^{600} (just to ensure that it is relatively prime to N), and then send

$$a^c \pmod{N}$$

to me. This is the encoded message. To decode it, I can receive the number $a^c \pmod{N}$ and then raise it to the power of d, which gives

$$(a^c)^d \equiv a^1 = a \pmod{N}.$$

So here is how this works in entirety, if A wants to send C some message.

- C picks large primes p and q.
- C picks a number c that is relatively prime to $(p-1)(q-1)$.
- C sends A the number $N = pq$ and c.
- A chops the message into pieces and turns them into numbers smaller than N (or maybe $\min(p, q)$).
- A computes the numbers $a^c \pmod{pq}$, and then sends C these numbers.
- C computes the integer d such that $cd \equiv 1 \pmod{(p-1)(q-1)}$.
- Then C computes $(a^c)^d \equiv a \pmod{pq}$. This process recovers the original message a.

If the eavesdropper figures out what p and q are, then E can do the decoding as fast as I can. Or if E can compute the order of b in $(\mathbb{Z}/pq\mathbb{Z})^\times$ then E can decode the message.

Here is how we could encode the message (in English) in $(\mathbb{Z}/pq\mathbb{Z})^\times$. First attach to each alphabet a number,

$$\text{space} = 0, \quad A = 1, \quad B = 2, \quad \dots, \quad Z = 26.$$

Then we can look at any text and write it in base 27, like

$$a_k 27^k + a_{k-1} 27^{k-1} + \cdots + a_0$$

where the original message was $a_0 a_1 \cdots a_k$. As long as $27^{k+1} < N$, this number fits in $(\mathbb{Z}/pq\mathbb{Z})^\times$. If this is not relatively prime, just add a number of spaces at the end, and this will be relatively prime with very high probability.

If you can get either p or q, you can decode the message. This is equivalent to knowing $\varphi(N) = (p-1)(q-1)$, because $p+q = pq - \varphi(N) + 1$. As we will see on Wednesday, if b is an encoded message and I know some r such that $b^r \equiv 1 \pmod{pq}$, (not necessarily the smallest!) then I can decode the message easily.

9 October 3, 2018

Let us recall the idea of RSA. We pick two primes p, q, and make their product $N = pq$ public. Also, we pick c a number relatively prime to $(p-1)(q-1)$, and make that public as well. The message is encoded by raising to the power of c. The decoding is done by finding a d such that $cd \equiv 1 \pmod{(p-1)(q-1)}$ and then raising to the power of d.

Suppose that the message is b, and suppose we know the order r of the message b, so that $b^r \equiv 1 \pmod N$. What can we do with that? Then r is an order of a as well, because

$$a^r \equiv (b^d)^r = (b^r)^d \equiv 1^d = 1 \pmod N.$$

Because r is an order, it divides $(p-1)(q-1)$ and so c is relatively prime to r. Then we can find a d' such that

$$d'c \equiv 1 \pmod r, \quad d'c = 1 + mr.$$

Then

$$b^{d'} \equiv (a^c)^{d'} = a^{cd'} = a \cdot a^{mr} = a \cdot 1^m = a \pmod N.$$

That is, we can decode the message by raising to the power of d'. This shows that if we know the order of the message b, then we can decode the message pretty fast. But again, finding the order of an element is not an easy task.

If you know the order, it is also pretty easy to factorize the integer. Suppose that we have

$$h^m \equiv 1 \pmod N.$$

Keep dividing m by 2, until we have something like

$$h^m \equiv 1 \pmod N, \quad h^{m/2} \not\equiv \pm 1 \pmod N.$$

Then $h^{m/2}$ is not 1, but its square is 1. We saw that

$$\mathbb{Z}/pq\mathbb{Z} \cong (\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/q\mathbb{Z}),$$

so there are four numbers squaring to 1, namely $(1, 1)$ and $(1, -1)$ and $(-1, 1)$ and $(-1, -1)$. Because we are assuming that it is not $(1, 1)$ or $(-1, -1)$, it is either $(1, -1)$ or $(-1, 1)$. Then

$$\gcd(h^{m/2} - 1, N)$$

will give one prime factor of N.

9.1 Equations with powers

Previously, we looked at how to solve the equation

$$ax \equiv b \pmod p.$$

Now let's try to solve equations that look like

$$ax^k \equiv b \pmod{p}.$$

Because we can find a d such that $ad \equiv 1 \pmod{p}$, we can multiply d on both sides and write the equation equivalently as

$$x^k \equiv bd = c \pmod{p}.$$

where we define $c = bd$.

We know that, because $\mathbb{Z}/p\mathbb{Z}$ is a field, the equation $x^k \equiv c \pmod{p}$ has at most k solutions.

Definition 9.1. The number c is called a kth **power residue** if $x^k \equiv c$ has a solution. It is called a **quadratic residue** if $x^2 \equiv c$ has a solution, and a **quadratic non-residue** if there is not.

Let's try to solve $x^k \equiv c \pmod{p}$ in general. Pick a primitive root g, and then we can write

$$(\mathbb{Z}/p\mathbb{Z})^\times = \{1, g, g^2, \dots, g^{p-2}\}.$$

Then we can write $c = g^m$ for some m and $x = g^z$ for some z. This z is what we are looking for. Then our equation can be written as

$$g^{kz} \equiv g^m \pmod{p},$$

which is equivalent to

$$kz \equiv m \pmod{p-1},$$

because g is a primitive root. So this reduces to a linear equation, not mod p but mod $p - 1$.

If k is relatively relatively prime to $p - 1$, then there exists a k' such that $kk' \equiv 1 \pmod{p-1}$. then we have

$$z \equiv k'm \pmod{p-1}.$$

Corollary 9.2. *If k is relatively prime to $p - 1$, then every nonzero element in $\mathbb{Z}/p\mathbb{Z}$ is a kth power residue.*

Example 9.3. Take $\mathbb{Z}/17\mathbb{Z}$. Then 3 turns out to be a primitive root, and so we can write $(\mathbb{Z}/17\mathbb{Z})^\times$ as

$$(\mathbb{Z}/17\mathbb{Z})^\times = \{1, 3, 9, 10, 13, 5, 15, 11, 16, 14, 8, 7, 4, 12, 2, 6\}.$$

Now suppose that we want to solve $x^3 = c$. Everything is going to be a cube. If we list $(3^k)^3$, we will get

$$1, 10, 15, 14, 4, 6, 9, 5, 16, 7, 2, 3, 13, 11, 8, 12.$$

This is a complete list.

But for quadratic residues, it is not true. If we try to solve

$$g^{2z} = g^m$$

for instance, we get

$$2z \equiv m \pmod{p-1}.$$

Because $p-1$ is an even number, this will have a solution if and only if m is even.

Example 9.4. In $(\mathbb{Z}/17\mathbb{Z})^\times$, the quadratic residues are

$$1, 9, 13, 15, 16, 8, 4, 12, 6$$

if we list $(3^k)^2$.

In fact, this list is not going to depend on the choice of our primitive root g. You can see this from the definition of a power residue, which does not depend on any primitive root. Or you can see that if g and h are primitive roots, then $h = g^l$ for some l relatively prime to $p-1$, and so

$$kz \equiv m \pmod{p-1}$$

has a solution if and only if

$$kz \equiv ml \pmod{p-1}.$$

9.2 Linear equations in an abelian group

Here is a different way to look at the equation $x^k = c$. Consider the abelian group

$$G = (\mathbb{Z}/p\mathbb{Z})^\times.$$

We can now define a map

$$\varphi : (\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times; \quad x \mapsto x^k.$$

You can check that this map preserves products and inverses, i.e., $\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2)$ and $\varphi(g_1^{-1}) = \varphi(g_1)^{-1}$.

Now we look at the kernel of this map,

$$\begin{aligned} \ker(\varphi) &= \{g \in G : \varphi(g) = 1\} \\ &= \{x \in (\mathbb{Z}/p\mathbb{Z})^\times : x^k = 1\}. \end{aligned}$$

So this is just the set of solutions for $x^k = 1$. So far, we know that

- there is exactly one solution if k is relatively prime to $p-1$,
- if k is a divisor of $p-1$ then it has k solutions.

We are going to try and prove the following theorem.

Theorem 9.5. *The equation*

$$x^k \equiv c \pmod{p}$$

has a unique solution if k is relatively prime to $p-1$. If k is a divisor of $p-1$, then there is exactly 0 *of k solutions.*

We already know this, but we are going to do this using group theory only. Recall that we defined

$$\varphi : G \to H, \quad K = \ker(\varphi) = \{g \in G : \varphi(g) = 1\},$$

and this is going to be a group inside G. This is because if $g_1 \in \ker(\varphi)$ and $g_2 \in \ker(\varphi)$ then

$$\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) = 1 \cdot 1 = 1$$

so $g_1 g_2 \in \ker(\varphi)$. Similarly, you can check that inverses of things in $\ker(\varphi)$ are also in $\ker(\varphi)$.

We also can think of the subset

$$K_c = \{g \in G : \varphi(g) = c\} \subseteq G.$$

This is closely related to K, because if if pick any element $g \in K_c$, then for any $h \in K$ we have

$$\varphi(gh) = \varphi(g)\varphi(h) = c \cdot 1 = c$$

so $g' = gh \in K_c$. Conversely, if we have any two $g, g' \in K_c$, we can divide $h = g^{-1}g'$ and we get $h \in K$. This means that if K_c is nonempty, then there is a one-to-one correspondence between

$$K_c = \{g : \varphi(g) = c\} \quad \longleftrightarrow \quad K = \{g : \varphi(g) = 1\}.$$

Proposition 9.6. *If $x^k = c$ has a solution, then the number of solutions is same as the number of solutions to $x^k = 1$.*

The last thing we can do with this is to count how many c we can solve the equation. Each c accounts for k solutions in $(\mathbb{Z}/p\mathbb{Z})^\times$. So in the case where k divides $p-1$, the number of kth power residues is

$$\frac{\#(\mathbb{Z}/p\mathbb{Z})^\times}{k} = \frac{p-1}{k}.$$

This is the number of c such that $x^k = c$ has a solution. For instance, there are $\frac{p-1}{2}$ number of c such that $x^2 = c$ has a solution.

10 October 10, 2018

We were looking at equations of the form

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

Here, if we write $x = g^z$ and $a = g^m$, this equation just becomes

$$g^{kz} \equiv g^m \pmod{p},$$

which is equivalent to the linear equation

$$kz \equiv m \pmod{p-1}.$$

If k divides $p-1$, there are either 0 or $\frac{p-1}{k}$ solutions to this equations. (If m is divisible by k then there are $\frac{p-1}{k}$ solutions, if not, there are no solutions.) So there are exactly $\frac{p-1}{k}$ elements $c \in (\mathbb{Z}/p\mathbb{Z})^\times$ for which the equation

$$x^k - c \equiv 0$$

is solvable, in which case there are exactly k solutions.

10.1 Quadratic residues

We are mostly interested in the case $k = 2$. Assume that p is odd. We know that there are $\frac{p-1}{2}$ elements $c \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that $x^2 - c$ is solvable, and $\frac{p-1}{2}$ elements c such that $x^2 - c$ is not solvable.

Definition 10.1. Fix p an (odd) prime. We call an element $c \in (\mathbb{Z}/p\mathbb{Z})^\times$ a **quadratic residue** if the equation

$$x^2 \equiv c \pmod{p}$$

is solvable. If $c \in (\mathbb{Z}/p\mathbb{Z})^\times$ is not a quadratic residue, we call it a **quadratic non-residue**.

There are exactly $\frac{p-1}{2}$ quadratic residues and $\frac{p-1}{2}$ quadratic non-residues. But which number is a quadratic residue and which number is a quadratic non-residue? For instance, we can ask when -1 is a quadratic residue, that is,

$$x^2 + 1 \equiv 0 \pmod{p}$$

has a solution.

Lemma 10.2. *Let p be an odd prime. A number $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ is a quadratic residue if and only if*

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Proof. If a is a quadratic residue, then $a \equiv x^2$ for some x. Then

$$a^{\frac{p-1}{2}} \equiv x^{p-1} \equiv 1 \pmod{p}$$

by Fermat's little theorem.

Now we can count numbers of quadratic residues and solutions of the equation. There are exactly $\frac{p-1}{2}$ solutions to the equation $a^{\frac{p-1}{2}} \equiv 1$, and there are exactly $\frac{p-1}{2}$ quadratic residues. Because all quadratic residues are solutions, it must be the case that the set of quadratic residues is equal to the set of solutions. Therefore we get if and only if. □

This answers the question of when -1 is a quadratic residue. By this lemma, -1 is a quadratic residue if and only if

$$(-1)^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Corollary 10.3. *For an odd prime p, the number -1 is a quadratic residue if and only if $p = 4k + 1$.*

Now we can ask the question of, after fixing an odd prime p, whether a given number is a quadratic residue or a quadratic non-residue. We have seen that a is a quadratic residue if and only if $a^{(p-1)/2} \equiv 1 \pmod{p}$. But it is always true that $(a^{(p-1)/2})^2 \equiv 1 \pmod{p}$, so we always have $a^{(p-1)/2} \equiv \pm 1$. That is,

$$a^{\frac{p-1}{2}} \equiv \begin{cases} 1 & a \text{ is a quadratic residue} \\ -1 & a \text{ is a quadratic non-residue} \end{cases} \pmod{p}.$$

This implies the following properties:

- If a and b are quadratic residues, then ab is also a quadratic residue. (You can see this by $a \equiv x^2$ and $b \equiv y^2$ implying $ab \equiv (xy)^2$.)
- If a is a quadratic residue and b is a quadratic non-residue, then ab is a quadratic non-residue. (If $a^{(p-1)/2} \equiv 1$ and $b^{(p-1)/2} \equiv -1$ then $(ab)^{(p-1)/2} \equiv -1$.)
- If a and b are quadratic non-residues, then ab is a quadratic residue. (If $a^{(p-1)/2} \equiv -1$ and $b^{(p-1)/2} \equiv -1$ then $(ab)^{(p-1)/2} \equiv -$.)

This means that there is a group homomorphism

$$(\mathbb{Z}/p\mathbb{Z})^\times \to \{1, -1\}$$

that sends quadratic residues to 1 and quadratic non-residues to -1.

Definition 10.4. We define the **Legendre symbol** as

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & a \text{ is a quadratic residue} \\ -1 & a \text{ is a quadratic non-residue.} \end{cases}$$

This is the homomorphism $(\mathbb{Z}/p\mathbb{Z})^\times \to \{\pm 1\}$ we described above.

So we immediately have

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

For instance, if we have a prime factorization $a = q_1^{r_1} q_2^{r_2} \cdots$ then we have

$$\left(\frac{a}{p}\right) = \left(\frac{q_1}{p}\right)^{r_1}\left(\frac{q_2}{p}\right)^{r_2} \cdots .$$

10.2 Statement of quadratic reciprocity

There is this celebrated theorem of Gauss, which is one of the early major achievements of number theory. The proof is really complicated, maybe the hardest proof we will do in this semester. But I spent my Columbus day trying to understand the proof, so I will give you a distilled version.

Theorem 10.5 (Gauss, quadratic reciprocity 1)**.** *If p and q are odd primes, then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}}.$$

What is this thing on the right hand side saying? If we write $p = 2k+1$ and $q = 2l+1$, then

$$(-1)^{\frac{(p-1)(q-1)}{4}} = (-1)^{kl}.$$

So it is -1 if both k and l are odd, and 1 if at least one of k or l is even. We then see that

$$(-1)^{\frac{(p-1)(q-1)}{4}} = \begin{cases} 1 & \text{at least one of } p \text{ or } q \text{ is of the form } 4k+1, \\ 1 & \text{both } p \text{ and } q \text{ are of the form } 4k+3. \end{cases}$$

This theorem doesn't say anything about the prime 2. So there is a part for this as well.

Theorem 10.6 (Gauss, quadratic reciprocity 2)**.** *For p an odd prime,*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & p \textit{ is of the form } 8k \pm 1, \\ -1 & p \textit{ is of the form } 8k \pm 3. \end{cases}$$

This theorem is really useful in computing the Legendre symbols.

Example 10.7. Suppose we want to compute

$$\left(\frac{3}{389}\right).$$

One ways is to just compute $3^{194} \pmod{389}$, because $194 = \frac{389-1}{2}$. This is doable, but it is a pain to try and do this. If we use Gauss's theorem, we can just do

$$\left(\frac{3}{389}\right)\left(\frac{389}{3}\right) = 1, \quad \left(\frac{389}{3}\right) = \left(\frac{2}{3}\right) = -1.$$

So we see that 3 is a quadratic non-residue of 389.

Example 10.8. Let's try another example. Is 29 a square modulo 23? First, we can reduce

$$\left(\frac{29}{23}\right) = \left(\frac{6}{23}\right) = \left(\frac{2}{23}\right)\left(\frac{3}{23}\right).$$

Here, $(\frac{2}{23}) = 1$ because $23 = 8 \times 3 - 1$, and

$$\left(\frac{3}{23}\right) = -\left(\frac{23}{3}\right) = -\left(\frac{2}{3}\right) = 1.$$

So 29 is a quadratic residue of 23.

10.3 First proof: counting points I

The first idea is this. Take any number a relatively prime to p. We look at the numbers

$$S = \{a, 2a, 3a, \dots, (\tfrac{p-1}{2})a\}$$

If we multiply them, we will get

$$1 \cdot 2 \cdot \dots \cdot (\tfrac{p-1}{2}) \cdot a^{\frac{p-1}{2}}.$$

On the other hand, we can take all the number in S and make them lie in the range

$$-(\tfrac{p-1}{2}), \dots, -1, 0, 1, \dots, (\tfrac{p-1}{2})$$

modulo p. Here, different elements of S become different numbers in this range, because $-(\frac{p-1}{2}), \dots, \frac{p-1}{2}$ forms a complete set of residues. Moreover, there can't be something landing in x and something landing in $-x$, because that will mean that

$$ka \equiv -x, \quad la \equiv x \pmod p$$

and adding them gives $(k+l)a \equiv 0 \pmod p$, which is impossible because $2 \leq k + l \leq p - 1$ and a is relatively prime to p.

$$\begin{array}{c} a, 2a, 3a, \dots, \frac{p-1}{2}a \\ \Big\downarrow \text{mod} p \\ \epsilon_1, 2\epsilon_2, \dots, \frac{p-1}{2}\epsilon_{\frac{p-1}{2}} \end{array}$$

Figure 1: First proof of quadratic reciprocity: reducing numbers modulo p

In the above picture, each ϵ_i is either $+1$ or -1. So when we multiply everything, we get

$$\left(\frac{p-1}{2}\right)! \cdot a^{\frac{p-1}{2}} \equiv \left(\frac{p-1}{2}\right)! \cdot (\epsilon_1 \cdots \epsilon_{\frac{p-1}{2}}) \pmod p.$$

So to compute $a^{\frac{p-1}{2}}$, we only need to count how many ϵ_j are -1. More precisely,

$$a^{\frac{p-1}{2}} = \begin{cases} 1 & \text{an even number of } \epsilon_j \text{ are } -1, \\ -1 & \text{an odd number of } \epsilon_j \text{ are } -1. \end{cases}$$

Graphically, here is what we are doing. We draw a real line, and color the intervals $[kp, (k+\frac{1}{2})p]$ red and color the intervals $[(k+\frac{1}{2})p, (k+1)p]$ blue. Then we take a ruler with markings at $a, 2a, \ldots, (\frac{p-1}{2})a$, and count how many points lie in the blue part of the real line. We will try to do this next class.

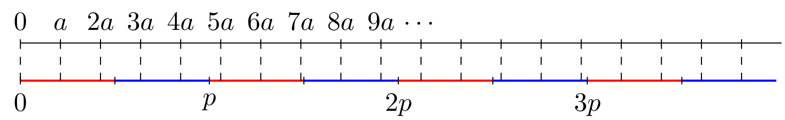

Figure 2: First proof of quadratic reciprocity: geometric interpretation

11 October 15, 2018

If p is an odd prime, then $x^2 \equiv a \pmod{p}$ has exactly two solutions or no solutions. So there are exactly $\frac{p-1}{2}$ numbers a such that $x^2 \equiv a$ is solvable, and these are called quadratic residues.

Lemma 11.1. *An integer a is a quadratic residue if and only if*

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

So we defined the Legendre symbol $(\frac{a}{p})$ so that it is 1 if a is a quadratic residue and -1 if a is not a quadratic residue. We had these properties

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

and there was Gauss's theorem that

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

For example,

$$\left(\frac{5}{29}\right)\left(\frac{29}{5}\right) = 1, \quad \left(\frac{29}{5}\right) = \left(\frac{4}{5}\right) = 1,$$

so we have that 5 is a square modulo 29. You could impress people at parties with this.

11.1 First proof: counting points II

We looked at the first step in the proof. We looked at the set

$$\{a, 2a, \dots, \tfrac{p-1}{2}a\}$$

and then looked at its distribution in the set $\{-\frac{p-1}{2}, \dots, -1, 0, 1, \dots, \frac{p-1}{2}\}$, how may things appear with a plus sign or a minus sign. If we write this set as

$$\{\epsilon_1, 2\epsilon_2, 3\epsilon_3, \dots, \epsilon_{\frac{p-1}{2}}\tfrac{p-1}{2}\} \subseteq \{-\tfrac{p-1}{2}, \dots, \tfrac{p-1}{2}\},$$

then $a^{\frac{p-1}{2}}$ is just how many $-$ signs we pick up from $\epsilon_1, \epsilon_2, \dots$. We also had a geometric interpretation of this. You color the real line by red and blue, with each segment having length $\frac{p}{2}$ appear in alternating colors, and then we are using a yard stick of length a and see where it lands.

So this involves a clever counting. Let's consider the example when $a = 2$. Then our set is

$$S = \{2, 4, 6, \dots, p-1\},$$

and we want to see how many lie in the interval $[\frac{p}{2}, p]$.

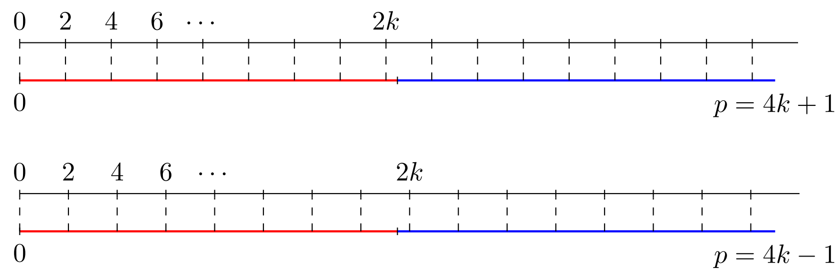

Figure 3: Gauss's theorem for $a = 2$

If $p = 4k+1$, the even numbers are $2, \dots, 2k, 2k+2, \dots, 4k$. So the numbers in the interval are

$$S \cap [\tfrac{p}{2}, p] = \{2k+2, 2k+4, \dots, 4k\}, \quad \#(S \cap [\tfrac{p}{2}, p]) = k.$$

This means that there are k minus signs, so we have

$$\left(\frac{2}{p}\right) = (-1)^k \quad \text{if } p = 4k+1.$$

We also have the case when $p = 4k - 1$. In this case, we can see that

$$S \cap [\tfrac{p}{2}, p] = \{2k, 2k+2, \dots, 4k-2\}, \quad \#(S \cap [\tfrac{p}{2}, p]) = k.$$

So again, we get

$$\left(\frac{2}{p}\right) = (-1)^k \quad \text{if } p = 4k-1.$$

If we put all of this together, we get

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & p = 8k \pm 1 \\ -1 & p = 8k \pm 3. \end{cases}$$

Now take a deep breath, as we are going to do the general situation. We want to show that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \begin{cases} 1 & p = 4k+1 \text{ or } q = 4l+1 \\ -1 & p = 4k-1 \text{ and } q = 4l-1. \end{cases}$$

Lemma 11.2. *Suppose that p and q are prime, and assume*

$$q \equiv \pm p \pmod{4a}.$$

Then

$$\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right).$$

If we have this statement, we can prove quadratic reciprocity. For instance, suppose $p - q = 4a$. Then we get

$$\left(\frac{p}{q}\right) = \left(\frac{p-q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right),$$
$$\left(\frac{q}{p}\right) = \left(\frac{q-p}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{p}\right).$$

So when $p \equiv q \pmod 4$, we get something like

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

Similarly if $p + q = 4a$, we can do the same thing

$$\left(\frac{p}{q}\right) = \left(\frac{p+q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right),$$
$$\left(\frac{q}{p}\right) = \left(\frac{q+p}{p}\right) = \left(\frac{4a}{p}\right) = \left(\frac{a}{p}\right).$$

That is,

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = 1$$

always, if $p \equiv -q \pmod 4$. So if we have the lemma, we are done.

Proof. If we try to compute $(\frac{a}{p})$, we saw that this is equivalent to counting how many of

$$\{a, 2a, \dots, \tfrac{p-1}{2}a\}$$

lie in the intervals $[(k - \frac{1}{2})p, kp]$. This is equivalent to counting how many of the integers

$$\{1, 2, \dots, \tfrac{p-1}{2}\}$$

lie in the intervals $[(k - \frac{1}{2})\frac{p}{a}, k\frac{p}{a}]$. We can interpret this as computing the sum

$$\sum_{1 \le k \le \frac{a}{2}} \#(\mathbb{Z} \cap [(k - \tfrac{1}{2})\tfrac{p}{a}, k\tfrac{p}{a}]).$$

But if we have $p = 4a + q$, then the intervals can be written as

$$[(k - \tfrac{1}{2})\tfrac{p}{a}, k\tfrac{p}{a}] = [(k - \tfrac{1}{2})\tfrac{q}{a} + 4k - 2, k\tfrac{q}{a} + 4k] = 4k + [(k - \tfrac{1}{2})\tfrac{q}{a} - 2, k\tfrac{q}{a}].$$

So you're looking at the interval $[(k - \frac{1}{2})\frac{q}{a}, k\frac{q}{a}]$, extending the smaller endpoint by 2, and then shifting by an integer. So we always have

$$\#(\mathbb{Z} \cap [(k - \tfrac{1}{2})\tfrac{p}{a}, k\tfrac{p}{a}]) = 2 + \#(\mathbb{Z} \cap [(k - \tfrac{1}{2})\tfrac{q}{a}, k\tfrac{q}{a}]).$$

If we add up over all the possible k, we get

$$\sum_{1\le k\le \frac{a}{2}} \#(\mathbb{Z}\cap[(k-\tfrac{1}{2})\tfrac{p}{a}, k\tfrac{p}{a}]) \equiv \sum_{1\le k\le \frac{a}{2}} \#(\mathbb{Z}\cap[(k-\tfrac{1}{2})\tfrac{q}{a}, k\tfrac{q}{a}]) \pmod 2.$$

Because $(\frac{a}{p})$ is (-1) to the power of the number on the left hand side, and $(\frac{a}{q})$ is (-1) to the power of the number of the right hand side, they are equal. $\square$

We have one more case, $p = 4k - q$, but here you do the same thing but with reflection.

12 October 17, 2018

We proved quadratic reciprocity when $p \equiv q \pmod 4$. We looked at the set $\{a, 2a, \dots, \frac{p-1}{2}a\}$ and then we looked at these things modulo p. We found out that we can determine $a^{\frac{p-1}{2}}$ by looking at how many things land in $\{-\frac{p-1}{2}, \dots, -1\}$. Using this, we were able to prove this lemma saying that if $p \equiv \pm q \pmod{4a}$ then

$$\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right).$$

Quadratic reciprocity followed from this lemma by doing things like

$$\left(\frac{p}{q}\right) = \left(\frac{p+q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right),$$
$$\left(\frac{q}{p}\right) = \left(\frac{p+q}{p}\right) = \left(\frac{4a}{p}\right) = \left(\frac{a}{p}\right),$$

and so the two are equal.

Here is the main idea used to prove this lemma. Instead of counting $a, \dots, \frac{p-1}{2}a$, we decided that we would count how many integers lie in the intervals

$$\left[\frac{1}{2}\frac{p}{a}, \frac{p}{a}\right], \left[\frac{3}{2}\frac{p}{a}, 2\frac{p}{a}\right], \dots, \left[\left(b - \frac{1}{2}\right)\frac{p}{a}, b\frac{p}{a}\right].$$

Let me call these the p-intervals. Now let us assume that $p = -q + 4a$. (This is the case we didn't cover last class.) Then these can also be written as

$$\left[-\frac{kq}{a} + \frac{q}{2a} - 2 + 4k, -\frac{kq}{a} + 4k\right].$$

Because we can shift by an integer value and the number of integers in the interval doesn't change, we can write this as

$$\left[-\frac{kq}{a} + \frac{q}{2a} - 1, -\frac{kq}{a}\right].$$

On the other hand, the q-intervals are, after reflection,

$$\left[-\frac{kq}{a}, -\frac{kq}{a} + \frac{q}{2a}\right].$$

If you add the p-interval and the q-interval together, we get the interval

$$\left[-\frac{kq}{a} + \frac{q}{2a} - 2, -\frac{kq}{a} + \frac{q}{2a}\right]$$

which has always 2 integers. This shows that the parity of the numbers of integers in the p-interval is equal to the number of integers in the q-interval. So we get quadratic reciprocity.

12.1 Finding the square root

So how do you find the square root? If a is a quadratic residue, then we have

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Then we have

$$(a^{\frac{p+1}{4}})^2 = a^{\frac{p+1}{2}} \equiv a \pmod{p}$$

if $p = 4k - 1$, so we find this square root $a^{\frac{p+1}{4}}$.

But what if $p = 4k + 1$? This is a magic of ring theory. Given a quadratic residue a, we want to find the square root of a. Let us introduce a new symbol α, and just declare that $\alpha^2 = a$. (This is like introducing this thing called i and declaring that $i^2 = -1$.) Now we consider the numbers of the form $u + v\alpha$ and define addition and multiplication like

$$\begin{aligned}(u + v\alpha) + (u' + v'\alpha) &= (u + u') + (v + v')\alpha,\\ (u + v\alpha)(u' + v'\alpha) &= (uu' + avv') + (uv' + vu')\alpha.\end{aligned}$$

Now what happens if we take

$$(u + v\alpha)^{\frac{p-1}{2}}?$$

This is computable, because every time we see an α^2, we can change it into a. So let us compute

$$(1 + z\alpha)^{\frac{p-1}{2}} \equiv u + v\alpha \pmod{p}.$$

Because b should satisfy the same rules as α, we would also have

$$(1 + zb)^{\frac{p-1}{2}} \equiv u + vb \pmod{p}.$$

Also, this is a $\frac{p-1}{2}$-th power modulo p, so it is either 0 or 1 or -1, although we don't know that it is. Still, we can try all the possibilities:

- if it is 0, then $b = -u/v$,
- if it is 1, then $b = (1 - u)/v$,
- if it is -1, then $b = (-1 - u)/v$.

Now we can try which is the correct solution by just taking the square and comparing them with a. This doesn't work if $v = 0$, but then we can choose another z.

Example 12.1. Let's try to find the solution of $x^2 \equiv -2 \pmod{17}$. If we pick $z = 5$, then we can find

$$\begin{aligned}(1 + 5\alpha)^2 &= -49 + 10\alpha = 2 + 10\alpha,\\ (1 + 5\alpha)^4 &= 4 - 40\alpha + 100\alpha^2 = 8 + 6\alpha,\\ (1 + 5\alpha)^8 &= -8 - 6\alpha.\end{aligned}$$

Then $u = -8$ and $v = -6$, and $v^{-1} = -3$. So the possibilities are

$$b = 24, 21, 27.$$

We see that $27^2 \equiv -2 \pmod{17}$.

So what is happening here? We are forming this new ring

$$(\mathbb{Z}/p\mathbb{Z})[\alpha]/(\alpha^2 - a) = \{x + y\alpha\} \text{ with } \alpha^2 = a,$$

and then there is a ring homomorphism

$$(\mathbb{Z}/p\mathbb{Z})[\alpha]/(\alpha^2 - a) \to \mathbb{Z}/p\mathbb{Z}$$

by sending α to b. We are then looking at the $\frac{p--1}{2}$-th power, so that we know what the image is regardless of what b is.

Next time, we are going to look at another proof of quadratic reciprocity, using something called Gauss sums. This is a sum of complex numbers. Define

$$\zeta = e^{2\pi i/z} = \cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p}$$

so that $\zeta^p = 1$, and the complete set of solutions of $x^p = 1$ is

$$1 = \zeta^0, \zeta, \zeta^2, \zeta^3, \ldots, \zeta^{p-1}.$$

13 October 22, 2018

Let p and q be distinct odd primes. Quadratic reciprocity says that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}}.$$

There is another proof using complex numbers, that somehow comes out of the blue.

13.1 Gauss sums

For $n \geq 2$, there is a complex number $\zeta \in \mathbb{C}$ such that $\zeta^n = 1$ but $\zeta, \zeta^2, \dots, \zeta^{n-1} \neq 1$. Explicitly, we can set

$$\zeta = e^{2\pi i/n} = \cos\frac{2\pi}{n} + i \sin\frac{2\pi}{n}.$$

On the complex plane, this is a point on the unit circle, with angle $\frac{2\pi}{n}$ from the real axis.

We can compute

$$\sum_{a=0}^{n} \zeta^n = \frac{\zeta^n - 1}{\zeta - 1} = \frac{0}{\zeta - 1} = 0.$$

But Gauss said, let's plug in the Legendre symbol and see what we get. So he introduced the **Gauss sum**

$$\sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \zeta^{ak} = G_a.$$

If you are used to Fourier series, you might know that we can get $(\frac{a}{p})$ from G_b, by

$$\left(\frac{a}{p}\right) = \frac{1}{p} \sum_{b=0}^{p-1} \zeta^{-ab} G_b.$$

To prove this, we need some facts about these complex numbers.

Proposition 13.1. *We have*

$$\sum_{k=0}^{p-1} \zeta^{ak} = \begin{cases} p & a \equiv 0 \pmod{p}, \\ 0 & a \not\equiv 0 \pmod{p}. \end{cases}$$

Proof. If $a \equiv 0 \pmod{p}$, then we have

$$\sum_{k=0}^{p-1} \zeta^{ak} = \sum_{k=0}^{p-1} 1^k = p.$$

On the other hand if $a \not\equiv 0 \pmod{p}$, then we can do this geometric sum

$$\sum_{k=0}^{p-1} \zeta^{ak} = \frac{\zeta^{pa} - 1}{\zeta^a - 1} = \frac{0}{\zeta^a - 1} = 0$$

because $\zeta^a - 1 \neq 0$. □

Now we can use this to prove that

$$\begin{aligned}\frac{1}{p}\sum_{b=0}^{p-1} \zeta^{-ab} G_b &= \frac{1}{p}\sum_{b=0}^{p-1} \zeta^{-ab} \sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \zeta^{bk} \\ &= \frac{1}{p}\sum_{k=0}^{p-1} \left(\frac{k}{p}\right) \sum_{b=0}^{p-1} \zeta^{b(k-a)} = \frac{1}{p}\left(\frac{a}{p}\right) p = \left(\frac{a}{p}\right),\end{aligned}$$

because $\sum_{b=0}^{p-1} \zeta^{b(k-a)}$ is only nonzero for $k = a$.

We also know that

$$G_a = \sum_{b=0}^{p-1} \zeta^{ab} \left(\frac{b}{p}\right) = \left(\frac{a}{p}\right) \sum_{b=0}^{p-1} \zeta^{ab} \left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right) G_1,$$

because summing over ab is the same as summing over b. So we now have three fundamental formulas:

$$G_a = \sum_{b=0}^{p-1} \zeta^{ab} \left(\frac{b}{p}\right), \quad \left(\frac{a}{p}\right) = \frac{1}{p}\sum_{b=0}^{p-1} \zeta^{-ab} G_b, \quad G_a = \left(\frac{a}{p}\right) G_1.$$

Using this, we get

$$p\left(\frac{-1}{p}\right) = \sum_{b=0}^{p-1} \zeta^b G_b = \left[\sum_{b=0}^{p-1} \zeta^b \left(\frac{b}{p}\right)\right] G_1 = G_1^2.$$

That is,

$$G_1 = (-1)^{\frac{p-1}{2}} p.$$

13.2 Second proof: Gauss sums

Using this, we can prove quadratic reciprocity. First, consider

$$G_1^{q-1} = (G_1^2)^{\frac{q-1}{2}} = (-1)^{\frac{(p-1)(q-1)}{4}} p^{\frac{q-1}{2}}.$$

This is starting to look like quadratic reciprocity, because $p^{\frac{q-1}{2}} \equiv (\frac{p}{q})$ modulo q.

On the other hand, we have

$$G_1^q = \left[\sum_{b=0}^{p-1} \left(\frac{b}{p}\right) \zeta^b\right]^q \equiv \sum_{b=0}^{p-1} \left(\frac{b}{p}\right) \zeta^{bq} = G_q = \left(\frac{q}{p}\right) G_1 \pmod{q},$$

because $(\alpha+\beta)^q \equiv \alpha^q + \beta^q \pmod q$. So we have computed G_1^q in two ways:

$$\left(\frac{q}{p}\right)G_1 \equiv G_1^q \equiv (-1)^{\frac{(p-1)(q-1)}{4}}\left(\frac{p}{q}\right)G_1 \pmod q$$

If we cancel G_1 from both sides, we get

$$\left(\frac{q}{p}\right) = (-1)^{\frac{(p-1)(q-1)}{4}}\left(\frac{p}{q}\right),$$

which is quadratic reciprocity.

I have to admit, this proof doesn't give us extra insight about the primes. It's like some alien came down and handed us a piece of paper with quadratic reciprocity written on it.

13.3 Sum of two squares

Now let us change topics and look at the equation

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

These are sometimes called Diophantine equations. The question is, for which N does there exist a solution? For small numbers, we get

$$\begin{aligned}&1^2+0^2=1, \quad 1^2+1^2=2, \quad 2^2+0^2=4, \quad 1^2+2^2=5, \quad 2^2+2^2=8,\\ &3^2+0^2=9, \quad 3^2+1^2=10, \quad 3^2=2^2=13, \quad 4^2+0^2=16, \quad 4^2+1^2=17,\\ &3^2+3^2=18, \quad \ldots.\end{aligned}$$

Do you see a pattern? If we look at the primes N with solutions, we observe that all primes of the form $N \equiv 1 \pmod 4$ do have a solution and primes of the form $N \equiv 3 \pmod 4$ do not have a solution.

Theorem 13.2. *The equation $x^2 + y^2 = p$ cannot be solved if p is a prime of the form $4k+3$.*

Proof. If there is a solution, then $x^2 \equiv -y^2 \pmod p$, and y cannot be a multiple of p because $y^2 \le p$ implies that $y \le \sqrt{p}$, and then we should have $y = 0$. This shows that $hy \equiv 1 \pmod p$ for some h. Then

$$h^2x^2 \equiv -1 \pmod p$$

and so -1 is a quadratic residue of p. This contradicts that p is a prime of the form $4k+3$. □

There is a stronger statement.

Theorem 13.3. *The equation $x^2 + y^2 = N = p_1^{m_1} \cdots p_l^{m_l}$ cannot be solved if for some k, the prime p_k is of the form $4t+3$ and m_k is odd.*

Proof. We have seen that if $x^2 + y^2 = N$ and $p \mid N$, where p is of the form $4t + 3$, then we get
$$x^2 \equiv -y^2 \pmod{p}.$$
If y is not divisible by p, we were able to invert y and get a contradiction from -1 being a quadratic non-residue. This shows that p has to divide y, and then p also divides x.

So we can write $x = px'$ and $y = py'$, and then our equation becomes
$$N = x^2 + y^2 = p^2(x'^2 + y'^2).$$
So if p^2 does not divide N, we see that there is no solution, and if p^2 does divide N, we can cancel out p^2 from both sides.

If N has a prime factor p of the form $4t + 3$ with odd exponent, then we can cancel out p^2 from N until p divides N but p^2 doesn't divide N. Then applying this fact again, we see that the equation $x^2 + y^2 = N$ does not have a solution. $\square$

This contains the fact that $x^2 + y^2 = N$ does not have a solution if $N \equiv 3 \pmod 4$. If $N \equiv 3 \pmod 4$, then there has to be a factor $p_k^{m_k}$ that is 3 modulo 4. But then $p_k \equiv 3 \pmod 4$ and m_k has to be odd.

The marvelous fact is that this is a sufficient and necessary condition. That is, if N does not take such a form, we always have a solution.

Theorem 13.4 (Euler)**.** *If m_k is even for all p_k of the form $4t + 3$, then the equation*
$$x^2 + y^2 = N = p_1^{m_1} \cdots p_l^{m_l}$$
has a solution.

We will prove this next time.

14 October 24, 2018

We were looking at which integers can be written as $N = x^2 + y^2$. Some necessary conditions were

- $N \equiv 0, 1, 2 \pmod 4$,
- if $N = p_1^{m_1} \cdots p_l^{m_l}$ and p_r is of the form $4k + 3$, then m_r is even.

This had to do with the fact that -1 is not a quadratic residue modulo p_r if p_r is of the form $4k + 3$. In fact, Euler's theorem says that this second condition is also sufficient.

Before proving this, let me give a nice theorem related to this.

Theorem 14.1. *There are infinitely many primes of the form $4k + 1$.*

Recall that we have previously shown that there are infinitely many primes of the form $4k + 3$. To show this, we assume there were finitely many primes $p_1, \ldots, p_N$, looked at

$$K = 4p_1p_2 \cdots p_N - 1,$$

and then got a contradiction. We will do something similar.

Proof. Suppose that $p_1, p_2, \ldots, p_N$ are all the primes of the form $4k + 1$. Look at

$$Q = 4p_1^2p_2^2 \cdots p_N^2 + 1,$$

and take a prime divisor q of Q. Then it is clear that q is at least distinct from $p_1, \ldots, p_N$. So q is a prime of the form $4k + 3$. But then,

$$-1 \equiv 4p_1^2 \cdots p_N^2 = (2p_1 \cdots p_N)^2 \pmod q$$

contradicts the fact that -1 is a quadratic non-residue of q, since q is of the form $4k + 3$. □

14.1 Euler's theorem

But anyways, let us prove Euler's theorem.

Theorem 14.2 (Euler)**.** *If m_k is even for all p_k of the form $4t + 3$, then the equation*

$$x^2 + y^2 = N = p_1^{m_1} \cdots p_l^{m_l}$$

has a solution.

Let us change notation and write

$$N = 2^l p_1^{m_1} \cdots p_r^{m_r} q_1^{2z_1} \cdots q_s^{2z_s},$$

where $p_1, \ldots, p_r$ are of the form $4k + 1$ and $q_1, \ldots, q_s$ are of the form $4k + 3$. Because we can set

$$x = q_1^{z_1} \cdots q_s^{z_s} x', \quad y = q_1^{z_1} \cdots q_s^{z_s} y',$$

it is enough to solve

$$(x')^2 + (y')^2 = 2^l p_1^{m_l} \cdots p_r^{m_r}.$$

Lemma 14.3. *The set of integers that can be written as a sum of two squares is closed under multiplication. More precisely, if $n = x^2 + y^2$ and $m = a^2 + b^2$, then*
$$mn = (xa + yb)^2 + (xb - ya)^2.$$

So if we can solve
$$x^2 + y^2 = 2, \quad x_i^2 + y_i^2 = p_i$$
for all p_i of the form $4k+1$, we are done, because we can apply this lemma over and over again to build a solution for $x^2 + y^2 = N$. Well, $x^2 + y^2 = 2$ has a solution
$$1^2 + 1^2 = 2.$$
So now let us show that $x^2 + y^2 = p$ has a solution for p prime of the form $4k + 1$.

We make use of the fact that -1 is a quadratic residue modulo p. So there is an x and m such that
$$-1 = x^2 - mp, \quad -\frac{1}{2}p < x < \frac{1}{2}p.$$
So we can write
$$x^2 + 1^2 = mp.$$

Now Euler's brilliance is that if we have anything like
$$x^2 + y^2 = mp$$
with $-\frac{1}{2}p < x, y < \frac{1}{2}p$ (this implies $m < \frac{1}{2}p$), and $m > 1$, then we can find a smaller $m' < m$ and integers x', y' such that
$$(x')^2 + (y')^2 = m'p.$$
So we begin with this initial seed $(x, 1)$ and keep finding smaller solutions, until we get $x^2 + y^2 = p$.

So how does Euler achieve this? If we have $x^2 + y^2 = mp$, then we can write
$$x = u + lm, \quad y = v + km, \quad -\frac{1}{2}m \le u, v \le \frac{1}{2}m.$$
Then we get
$$u^2 + v^2 = rm, \quad 1 \le r \le \frac{1}{2}m < m.$$
Now we can take the product of the two things and get
$$(xu + yv)^2 + (xv - yu)^2 = rm^2p.$$

Let's look at what we've got. If we look modulo m, we see that
$$\begin{aligned} xu + yv &\equiv u \cdot u + v \cdot v = rm \equiv 0 \pmod{m}, \\ xv - yu &\equiv u \cdot v - v \cdot u = 0 \pmod{m}. \end{aligned}$$

So we can write
$$xu + yv = mX, \quad xu + yv = mY.$$
Then
$$(mX)^2 + (mY)^2 = rm^2p, \quad X^2 + Y^2 = rp.$$
Because $r < m$, we get a smaller multiple of p written as a sum of two squares.

If you think about this, this proof is constructive and algorithmically efficient. Every time you do this process, the value of m decreases by a half. So it will only take around $\log p$ time to run this algorithm.

Theorem 14.4. *For p a prime of the form $4k+1$, the representation*
$$x^2 + y^2 = p$$
is essentially unique. (This means, unique up to putting minus signs or switching the numbers.)

Proof. Suppose we have
$$x^2 + y^2 = p, \quad X^2 + Y^2 = p.$$
Then if we write $h^2 \equiv -1 \pmod p$ for the square root of -1, then
$$x^2 \equiv h^2y^2, \quad X^2 \equiv h^2Y^2 \pmod p,$$
so after possibly changing signs of y and Y, we can arrange so that
$$x \equiv hy, \quad X \equiv hY \pmod p.$$
Then we get
$$p^2 = (x^2 + y^2)(X^2 + Y^2) = (xX + yY)^2 + (xY - yX)^2.$$
But $xX + yY \equiv xX + (hx)(hX) \equiv xX - xX = 0 \pmod p$ and similarly $xY - yX \equiv x(hX) - (hx)X = 0 \pmod p$. So both are divisible by p. Then
$$1 = \left(\frac{xX + yY}{p}\right)^2 + \left(\frac{xY - yX}{p}\right)^2$$
where both are integer squares. The only possibilities are then $(\pm 1)^2 + 0^2$ and $0^2 + (\pm 1)^2$. If $a = 1$ and $b = 0$, this is saying
$$xX + yY = p, \quad xY = yX.$$
Because $x^2 + y^2 = X^2 + Y^2 = p$, if you combine these equations, we get $(x, y) = (X, Y)$. $\square$

14.2 Quadratic forms

Now we know how to solve $x^2 + y^2 = N$. But can we solve more general equations? For instance, when can we represent a given number N as

$$ax^2 + bxy + cy^2 = N,$$

when we are given integers a, b, c? What we have shown before was the case $(a, b, c) = (1, 0, 1)$.

Definition 14.5. We say that n is **properly represented** by a quadratic form $ax^2 + bxy + cy^2$ if there are relatively prime integers x, y such that

$$ax^2 + bxy + cy^2 = n.$$

There are interesting quadratic equations like Pell's equations,

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

I won't say much about this equation, maybe we will get to talk about this later in the course. There is this boring solution $(x, y) = 1$, but the interesting thing is that if we have two solutions

$$x^2 - Ny^2 = 1, \quad X^2 - NY^2 = 1,$$

then we can multiply the two solutions and get

$$(xX + NyY)^2 - N(xY + yX)^2 = 1.$$

So with this multiplication operation, this forms a group.

Let us get back to quadratic forms. We can think of a quadratic form as a symmetric matrix,

$$\mathbb{A} = \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix}.$$

Then for any vector $\vec{v} = \left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)$, we can form

$$ax^2 + bxy + cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \vec{v}^T \mathbb{A} \vec{v}.$$

So we are really just solving the equation

$$v^T \mathbb{A} v = n$$

for $v \in \mathbb{Z}^2$.

This is a useful thing to do, because we can make sense of change of basis. If we write

$$x = pX + qY, \quad y = rX + sY,$$

then we can think of this as a change of basis,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Then we can write the equation as

$$n = \begin{pmatrix} X & Y \end{pmatrix} \begin{pmatrix} p & r \\ q & s \end{pmatrix} \begin{pmatrix} a & \frac{b}{2} \\ \frac{b}{2} & c \end{pmatrix} \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

So if we can invert the matrix

$$\mathbb{M} = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$$

with integer entries,

$$\mathbb{M}^{-1} = \frac{1}{ps - qr} \begin{pmatrix} s & -q \\ -r & p \end{pmatrix}$$

with $ps - qr = \pm 1$, then we see that the change of basis

$$\mathbb{A} \quad \rightsquigarrow \quad \mathbb{M}^T \mathbb{A} \mathbb{M}$$

really doesn't change the quadratic form in a significant way. To be more precise, the set of integers properly representable by $\mathbb{A}$ is equal to the set of integers properly representable by $\mathbb{M}^T \mathbb{A} \mathbb{M}$.

Now we can ask the following interesting question. What are all the quadratic forms, up to equivalence? Here, we say that $\mathbb{A} \sim \mathbb{A}'$ if and only if

$$\mathbb{A}' = \mathbb{M}^T \mathbb{A} \mathbb{M}$$

for some 2×2 integer matrix $\mathbb{M}$ with $\det \mathbb{M} = ps - qr = 1$. To answer this question, we need to look at quantities that distinguishes quadratic forms.

Definition 14.6. The **discriminant** of a quadratic form $\mathbb{A}$ is given by

$$\operatorname{disc}(\mathbb{A}) = d = b^2 - 4ac = -4 \det \mathbb{A}.$$

This is a useful thing to define, because if $\mathbb{A}'$ is equivalent to $\mathbb{A}$ by $\mathbb{A}' = \mathbb{M}^T \mathbb{A} \mathbb{M}$, then

$$\operatorname{disc}(\mathbb{A}') = -4 \det \mathbb{A}' = -4(\det \mathbb{M}^T)(\det \mathbb{A})(\det \mathbb{M}) = -4 \det \mathbb{A} = \operatorname{disc}(\mathbb{A}),$$

because $\det \mathbb{M} = \det \mathbb{M}^T = 1$. So equivalent quadratic forms have the same discriminant.

Unfortunately, having the same discriminant does not mean that the quadratic forms are equivalent. For instance,

$$2x^2 + 3y^2, \quad x^2 + 6y^2$$

both have discriminant -24, but they are not equivalent because $2x^2 + 3y^2$ cannot represent 1, but $x^2 + 6y^2$ can represent 1. But discriminants are still going to be part of the story.

15 October 29, 2018

A quadratic form was represented by three integers a, b, c. We wanted to see which integers n can be represented as

$$n = ax^2 + bxy + cy^2.$$

We showed that we can make a change of basis

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}$$

if $ps - rq = 1$. This does help, because we might be able to simplify the quadratic form using this procedure. On the quadratic form, this corresponded to doing

$$\mathbb{A}' = \mathbb{M}^T \mathbb{A} \mathbb{M}.$$

So how can we tell if two quadratic forms are equivalent?

15.1 Discriminant of quadratic forms

We introduced an invariant called the discriminant. This was defined by

$$d = -4 \det \mathbb{A} = b^2 - 4ac.$$

This is invariant under change of basis, because $\operatorname{disc}(\mathbb{A}') = \operatorname{disc}(\mathbb{A})$. But this is not enough to distinguish all equivalence classes.

Example 15.1. The two quadratic forms $x^2 + 6y^2$ and $2x^2 + 3y^2$ both has discriminant -24, but they are not equivalent. You can see this by noting that the first represents 1 but the second does not.

There are positive discriminants as well. For instance, $x^2 - 2y^2$ has $d = 8 > 0$. These things are like Pell's equation. If the discriminant is negative, $d < 0$, then the quadratic form is either positive or negative. You can see this from completing the square. We have

$$ax^2 + bxy + cy^2 = \frac{1}{4a}\left(4a^2\left(x + \frac{by}{2a}\right)^2 + (4ac - b^2)y^2\right),$$

so if $4ac - b^2 = -d > 0$, then this always has the same sign as a. We also see from this expression that discriminant $d = 0$ is not interesting; this just becomes

$$\frac{1}{4a}(2ax + by)^2.$$

For this reason, we are only going to look at the case $d \neq 0$. Assume that n is properly representable by some $x = p$ and $y = r$. Then p and q are relatively prime, (this is the definition of properly representable) and so there are integers s and q such that

$$pr - qs = 1.$$

Then we can use this to make a linear transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

Then our quadratic form becomes

$$nX^2 + hXY + lY^2,$$

because $(X, Y) = (1, 0)$ gives $(x, y) = (p, r)$, and hence should evaluate to n. If we look at the discriminant, we see that

$$d = h^2 - 4nl, \quad d \equiv h^2 \pmod{4n}.$$

That is, if n is properly represented by a quadratic form, its discriminant d has to be a quadratic residue of $4n$.

Here is one other thing. We always have

$$d = b^2 - 4ac \equiv b^2 \pmod 4.$$

So we get the following.

Proposition 15.2. *Any discriminant of a quadratic form is* 0 *or* 1 *modulo* 4*.*

In fact any number of the form $4k$ or $4k + 1$ can be expressed as $b^2 - 4ac$, so every such number is the discriminant of some quadratic form.

Lemma 15.3. *If* n *is property represented by* $ax^2 + bxy + cy^2$*, then there is an equivalent quadratic form* $nx^2 + hxy + ly^2$*. (It follows that* $d \equiv h^2 \pmod{4n}$*.)*

Conversely, if d is a quadratic residue modulo $4n$, then we can write $d = h^2 - 4nl$. Then n can be properly represented by

$$nx^2 + hxy + ly^2,$$

which is a quadratic form of discriminant d. Of course, this does not have to be equivalent to the form we started out with. So we haven't answered the original question we asked.

We probably won't have time to discuss this, but we might want to study how many equivalence classes of quadratic forms there are, given a discriminant d. This is called the **class number** of d, written as $C(d)$. If $d = -p$ where $p = 4k + 3$, this becomes

$$C(d) = \frac{B - A}{p}, \quad B = \text{sum of quad. non-res.}, \quad A = \text{sum of quad. res.}.$$

15.2 Special quadratic forms

Let us play around with quadratic forms. The quadratic form

$$x^2 + y^2$$

has discriminant $d = -4$. So what are the quadratic form of discriminant -4, up to equivalence?

Theorem 15.4. *All quadratic forms of discriminant $d = -4$ is equivalent to $x^2 + y^2$ or $-x^2 - y^2$.*

Proof. Consider $ax^2 + bxy + cy^2$, where $b^2 - 4ac = -4$. If we make a change of basis $x \to X \pm Y$ and $y \to Y$, then we get to

$$aX^2 + (b \pm 2a)XY + c'Y^2.$$

So by continuously doing this, we can arrange so that $|b| \leq |a|$.

Now we ask, which of $|a|$ and $|c|$ is bigger? If $|c|$ is smaller than $|a|$, we can switch them (by $x \to Y$ and $y \to X$) and do the above process, and repeat. I always just make b smaller, so this process should end at some point. At the end, we get a quadratic form $ax^2 + bxy + cy^2$ with

$$|b| \leq |a| \leq |c|.$$

Up to this point, we didn't even use the fact that the discriminant is -4; this can be done for every discriminant.

The reduction can't go any further. So let's just see what we have got. We have

$$4|ac| - 4 \leq |4ac - 4| = |b|^2 \leq |a|^2 \leq |ac|.$$

So either $|ac| = 0$ or $|ac| = 1$, but then $b^2 - 4ac = -4$ shows that $ac > 0$. So the only possibility is $ac = 1$, which implies $b = 0$. This shows that the quadratic form at the end is either $x^2 + y^2$ or $-x^2 - y^2$. □

This can be used to prove Euler's theorem.

Corollary 15.5. *A positive integer n is properly represented by $x^2 + y^2$ if and only if -1 is a quadratic residue of n.*

Proof. If n is properly represented by $x^2 + y^2$, then we have seen that

$$-4 = d \equiv h^2 \pmod{4n},$$

and then because h has to be even, we can write $h = 4m$ and get $-1 \equiv m^2 \pmod{n}$.

Conversely, if -1 is a quadratic residue of n, then $-1 \equiv m^2 \pmod{n}$, so we can write

$$-4 = (2m)^2 - 4nl$$

for some integer l. Then n is properly representable by the quadratic form

$$nx^2 + (2m)xy + ly^2,$$

which has discriminant -4. Then this quadratic form is equivalent to either $x^2 + y^2$ or $-x^2 - y^2$, and positivity of n tells us that it has to be $x^2 + y^2$. Therefore n has to be properly represented by $x^2 + y^2$ as well. □

Proposition 15.6. *For $n = 2^s p_1^{l_1} \cdots p_k^{l_k}$ a positive integer, -1 is a quadratic residue modulo n if and only if $s \leq 1$ and all p_i are of the form $4k + 1$.*

Proof. By the Chinese remainder theorem, -1 is a quadratic residue modulo n if and only if -1 is quadratic residue modulo 2^s and modulo $p_i^{l_i}$ for all i, individually.

It it straightforward to check that if -1 is a quadratic residue, each p_i have to be of the form $4k+1$ and $s \leq 1$. (This is because -1 is not a quadratic residue of primes of forms $4k+3$, and also of 4.) Conversely, -1 is a quadratic residue of 2. So it suffice to check that -1 is a quadratic residue of $p_i^{k_i}$ if p_i is of the form $4k+1$. One way is to do this by induction. If $z^2 \equiv -1 \pmod{p^k}$, then we can find a suitable t such that

$$(z+tp^k) \equiv z^2 + 2tp^k \equiv -1 \pmod{p^{k+1}},$$

because this is equivalent to $2t \equiv (1-z^2)/p^k \pmod{p}$. □

So if we combine these, we get the following.

Corollary 15.7. *A positive integer n is properly represented by x^2+y^2 if and only if n is of the form $n = 2^s p_1^{l_1} \cdots p_k^{l_k}$, with each p_i a prime of the form $4k+1$.*

16 October 31, 2018

Today I want to talk a bit about Pell's equations and then change to elliptic curves.

16.1 Pell's equation

Let N be a positive integer. Then we are going to look at the equation

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

Then the discriminant is $d = 4N$. Note that this is the equation for the hyperbola. It always has two solutions, $(\pm 1, 0)$, and so it passes through the two points.

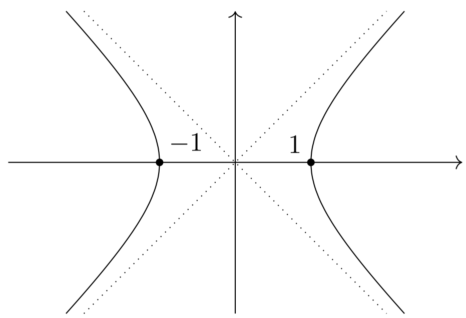

Figure 4: Solutions to Pell's equations $x^2 - Ny^2 = 1$

There is a multiplication structure on the set of solutions. If we have

$$x^2 - Ny^2 = r, \quad a^2 - Nb^2 = s,$$

then we can write the product of the two equations as

$$(ax + Nby)^2 - N(bx + ay)^2 = (x^2 - Ny^2)(a^2 - Nb^2) = rs.$$

We can also think of this as a unimodular transformation

$$\mathbb{M} = \begin{pmatrix} a & Nb \\ b & a \end{pmatrix}, \quad \mathbb{M}^T \begin{pmatrix} 1 & 0 \\ 0 & -N \end{pmatrix} \mathbb{M} = \begin{pmatrix} 1 & 0 \\ 0 & -N \end{pmatrix}.$$

This unimodular transformation then preserves the quadratic form $x^2 - Ny^2 = 1$, and so this is an **automorphism** of the quadratic form. This is why it forms a group.

The inverse solution of (x, y) is given by $(x, -y)$. If we try to multiply the two, we get

$$(x, y) * (x, -y) = (x^2 - Ny^2, xy - yx) = (1, 0).$$

So this is kind of cool, and it is an abelian group.

Example 16.1. Consider the equation $x^2 - 2y^2 = 1$. There is a solution $(x, y) = (3, 2)$. If we multiply it with itself, we get

$$(3, 2) * (3, 2) = (17, 12).$$

Then we can multiply it with $(3, 2)$ to get bigger solutions,

$$(3, 2) * (17, 12) = (99, 70).$$

Example 16.2. For $x^2 - 3y^2 = 1$, there is a solution $(x, y) = (2, 1)$. Then

$$(2, 1) * (2, 1) = (7, 4), \quad (2, 1) * (7, 4) = (26, 15).$$

So these are solutions.

16.2 Elliptic curves

We are not going to get too deep, because we would need a lot a machinery. But we can still say some interesting things. Let a and b be integers. We are going to look at solutions to

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

More generally, we can take $x^3 + \alpha x^2 + \beta x + \gamma$ on the right hand side, but we can shift it back to $x^3 + ax + b$ by making a change of variables $x \to x + c$. This graph lies in some range $x \geq -C$.

We can look for integer solutions $x, y \in \mathbb{Z}$, but we can also look for solutions in the reals, $x, y \in \mathbb{R}$, or other fields. In fact, we can choose a prime p and consider the solutions

$$E : y^2 \equiv x^3 + ax + b \pmod{p},$$

for $x, y \in \mathbb{Z}/p\mathbb{Z}$. This is called an **elliptic curve** over a finite field. Then we can consider

$$E = \{(x, y) \in \mathbb{Z}/p\mathbb{Z} : y^2 = x^3 + ax + b\} \cup \{O\}.$$

So people usually study elliptic curves over finite fields, or also over the rational numbers.

There is a way to add solutions together. That is, given two solutions (x, y) and (x', y') of the equation

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

we can add them to get another solution (x'', y''). This is assuming that we have

$$D = -16(4a^3 + 27b^2) \neq 0$$

in our field $k = \mathbb{Z}/p\mathbb{Z}$ or $k = \mathbb{Q}$. (This D is called the **discriminant** of the curve.) If $D = 0$, let's say over the reals, then the curve is not going to be smooth. Here is how you see this. Consider our curve $f(x,y) = y^2 - x^3 - ax - b = 0$, and let's try to compute the tangent vector. If we compute the gradient, this is

$$\nabla f = (-3x^2 - a, 2y).$$

If $f = 0$ and $\nabla f = 0$, this means that the level set of $f = 0$ has some singularity. First $\nabla f = 0$ happens when $y = 0$ and $x = \sqrt{-a/3}$. Then this satisfies $f = 0$ when

$$(-\tfrac{a}{3})^{3/2} + a(-\tfrac{a}{3})^{1/2} + b = 0,$$

which is equivalent to $4a^3 + 27b^2 = 0$. So this the bad condition.

Now we are going to define addition on the points. Here is the rule. We are given two points P_1 and P_2 on the curve.

- We always have $O + P_1 = P_1 + O = P_1$ and $O + P_2 = P_2 + O = P_2$.
- If P_2 and P_1 have the same x-coordinate and are reflections of each other along the x-axis, then $P_1 + P_2 = O$.
- If $P_1 \neq P_2$, you take the line connecting P_1 and P_2, and look at the third intersection of the line with the curve. Reflect this point along the x-axis, and set $P_1 + P_2$ as this reflection.
- If $P_1 = P_2$, we instead take the tangent line at P_1 to the curve, and then reflect the third intersection to get $P_1 + P_1 = 2P_1$.

So this make sense as long as there are tangent vectors, and that is why we need this condition $4a^3 + 27b^2 = 0$.

17 November 5, 2018

I was talking about elliptic curves, maybe we will get to applications. This is defined over a field K (with addition with additive unit 0 and multiplication with multiplicative unit 1), like $\mathbb{R}$ or $\mathbb{Q}$ or $\mathbb{Z}/p\mathbb{Z}$. So this is given by an equation

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

for $a, b \in K$, and we are looking for solutions in $x, y \in K$. Then we write

$$E(K) = \{(x, y) : y^2 = x^3 + ax + b\} \cup \{O\},$$

as the set of solutions plus this unnatural "point at infinity" which denote by O. An elliptic curve looks something like this:

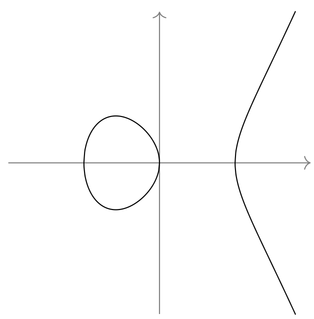

Figure 5: The elliptic curve $y^2 = x^3 - x$ over $\mathbb{R}$

Given an point, it is not hard to write down an elliptic curve that passes through this point. You can just play around with a and b. But the other direction is hard. Given an elliptic point, it is really hard to find rational points on that curve, or even determine if there is a rational point. But over finite fields, this is doable.

Example 17.1. Let us take $K = \mathbb{Z}/5\mathbb{Z}$. Consider the elliptic curve $E : y^2 = x^3 + 1$. Then we have $x \in \{0, 1, 2, 3, 4\}$, so we can just see for which x is $x^3 + 1$ a quadratic residue, i.e., in $\{0, 1, 4\}$. If we do this, we can list all the solutions, and this is

$$E(K) = \{(0, 1), (0, 4), (2, 2), (2, 3), (4, 0), O\}.$$

If we take $E : y^2 = x^3 - x$, then we see

$$E(K) = \{(0, 0), (1, 0), (2, 1), (2, 4), (3, 2), (3, 3), (4, 0), O\}.$$

We can do the same thing for $K = \mathbb{Z}/7\mathbb{Z}$.

17.1 Group structure on elliptic curves

The cool thing about elliptic curves is that points on an elliptic curve has a group structure. Let us take two points P_1, P_2 in $E(K)$. Here are the rules for addition:

- $O + P = P$, $P + O = P$,
- for $P = (x, y)$, if we denote $\bar{P} = (x, -y)$, then $P + \bar{P} = O$,
- if $P_1 \neq P_2$, so $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, then

  $$P_1 + P_2 = (\lambda^2 - x_1 - x_2, -\lambda(\lambda^2 - x_1 - x_2) - y_1 + \lambda x_1), \quad \lambda = \frac{y_1 - y_2}{x_1 - x_2},$$

  (Here what we are doing is looking at the third intersection point of the elliptic curve with the line connecting P_1 and P_2, and then reflecting it about the x-axis.)
- if $P_1 = P_2 = P$, then we can formally think that λ is the tangent line, so

  $$P + P = (\lambda^2 - 2x, -\lambda(\lambda^2 - 2x) - y + \lambda x), \quad \lambda = \frac{3x^2 + a}{2y}.$$

In order for this to make sense, we need that $-16(27a^3 + 4b^2) \neq 0$ in the field K, since we don't want to divide by zero.

So why is there this third point on the curve? If I have any line, I can parametrize it as

$$y = \lambda x + c.$$

(Let's assume that we don't have a vertical line, since this is the case where we add P and $\bar{P}$.) Our elliptic curve equation is just $y^2 = x^3 + ax + b$, so if we substitute for y, we get

$$\lambda^2 x^2 + 2\lambda xc + c^2 = x^3 + ax + b,$$

which is a cubic equation in x. But we already have two solutions, and so we see that $\lambda^2 - x_1 - x_2$ has to be a third solution, since the three solutions add up to λ^2.

Theorem 17.2. *Elliptic curve addition obeys the group axioms, and is commutative. So it is an abelian group.*

It is not hard to see that this is commutative. We see that λ does not change when we switch P_1 and P_2, since

$$\lambda = \frac{y_1 - y_2}{x_1 - x_2} = \frac{y_2 - y_1}{x_2 - x_1}.$$

Then the only other term that seems to depend on the order is $y_1 - \lambda x_1$, but this also does not matter because $y_1 - \lambda x_1 = y_2 - \lambda x_2$. So the addition formula is actually symmetric.

The hard part is checking associativity,

$$P_1 + (P_2 + P_3) = (P_1 + P_2) + P_3.$$

You can just do the algebra and check this. There is a geometric interpretation, but this is still hard to see. But we can still check something simpler. Let's try to figure out why

$$P_2 + (\bar{P}_2 + P_1) = P_1.$$

If we look at the intersection point of the curve and the line ℓ connecting P_1 and $\bar{P}_2$ and reflect, this is the same as reflecting the line and then intersecting with the curve. So $\bar{P}_2 + P_1$ is the intersection point Q of the curve with the line $\bar{\ell}$ connecting P_2 and $\bar{P}_1$. Then if we try to add P_2 to this point, we are looking at the line connecting P_2 and Q, which is just $\bar{\ell}$, taking the third intersection with the curve, which is at $\bar{P}_1$, and then reflect, which gives us back P_1.

Example 17.3. Take $y^2 = x^3 + 1$ with the point $P = (2, 3)$. If we add this point with itself, we get

$$\lambda = \frac{3 \cdot 4}{2 \cdot 3} = 2, \quad 2P = (4 - 4, 1) = (0, 1).$$

If we add $(2, 3)$ again to $(0, 1)$, then we get

$$3P = (-1, 0), \quad 6P = (-1, 0) + (-1, 0) = O.$$

Sometimes, nP generates infinitely many points on the elliptic curves, and sometimes $nP = O$ for some n, so there are only a finitely many points generated by P. These are called **torsion points** on the elliptic curve. Over finite fields, there are only finitely many points, so all points are torsion points.

18 November 7, 2018

Elliptic curves were defined over fields $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{Z}/p\mathbb{Z}$, and then we defined

$$E(K) = \{(x, y) : y^2 = x^3 + ax + b\} \cup \{O\}.$$

In order for this to be a group, we needed a condition $16(27a^2 - 4b^3) \neq 0$.

Example 18.1. Take $y^2 = x^3 + 8$. There are solutions like $(1, \pm 3)$, $(2, \pm 4)$, $(-2, 0)$. If we add $(1, 3)$ and $(2, 4)$, we get $(-2, 0)$ which is torsion. But if we try to add $(2, 4)$ to itself, we have

$$2(2, 4) = (-\tfrac{7}{4}, \tfrac{13}{8}).$$

If you add this more, you will get more complicated fractions.

There is a notion of equivalence of elliptic curves. Suppose I have

$$y^2 = x^3 + ax + b, \quad y^2 = x^3 + Ax + B.$$

Now we ask if there is a change of coordinates that make them equivalent. If n has an inverse in K, then we can consider the change of coordinates $x = x'/n^2$ and $y = y'/n^3$. Then this becomes

$$y'^2 = x'^3 + n^4 a x' + n^6 b.$$

18.1 Factorization using elliptic curves

First we look at Pollard's $p - 1$ method for factoring.

Definition 18.2. Let B be a large integer. An integer n is said to be **B-power-smooth** if in the factorization $n = p_1^{a_1} \cdots p_k^{a_k}$, each $p_i^{a_i}$ are at most B.

Example 18.3. The number $90 = 2 \times 3^2 \times 5$ is 9-power-smooth, but $25 = 5^2$ is not 9-power-smooth.

Now we can factor integers $N = pq$ if $p - 1$ or $q - 1$ are B-power-smooth. We first choose B, and then let

$$m_B = \text{least common multiple of } \{1, 2, \dots, B\}.$$

This can be done by

- listing all primes $\leq B$,
- for each prime p, looking at the largest power p^k that does not exceed B,
- multiplying all these powers.

We can even get a bound on this number m_B.

So we want to factor $N = pq$. We can first take m the least common multiple of $\{1, 2, \ldots, B\}$, and assume that $p - 1$ divides M. This shows that we have $m = (p - 1)l$ and so

$$a^m \equiv 1 \pmod{p}$$

by Fermat's little theorem. This means that $p \mid a^m - 1$. So if we take the $\gcd(a^m - 1, N)$, there is a good change that we will recover p.

Let us now look at **Lenstra's method**. Consider an elliptic curve $E : y^2 = x^3 + ax + b$ over $\mathbb{Z}/p\mathbb{Z}$. Similarly, denote by m the least common multiple of $\{1, 2, \ldots, B\}$.

Proposition 18.4. *The size of $E(\mathbb{Z}/p\mathbb{Z})$ is $p \pm s$ for some $0 \leq s < 2\sqrt{p}$.*

So by changing a, you can change the size of the elliptic curve, but you can't get too far way from p. Now we consider the elliptic curve over $\mathbb{Z}/N\mathbb{Z}$, which is still fine as long as $16(27b^2 + 4a^3) \in (\mathbb{Z}/N\mathbb{Z})^\times$. Then we can still make sense of

$$E(\mathbb{Z}/N\mathbb{Z}) = \{(x, y) : y^2 = x^3 + ax + b\}$$

as a group. This actually splits into

$$E(\mathbb{Z}/p\mathbb{Z}) \times E(\mathbb{Z}/q\mathbb{Z})$$

as a group, by the Chinese remainder theorem.

Now we look at the curve $E : y^2 = x^3 + ax + 1$. This has one solution $(0, 1)$. So we can try to compute

$$m(0, 1) \in E(\mathbb{Z}/p\mathbb{Z}) \times E(\mathbb{Z}/q\mathbb{Z}).$$

So if we look at this and it becomes the unit element $E(\mathbb{Z}/p\mathbb{Z})$, we can similarly take the greatest common divisor to recover p. The advantage of the elliptic curve method over Pollard's method is that we can play around with a, and change the sizes of $E(\mathbb{Z}/p\mathbb{Z})$ and $E(\mathbb{Z}/q\mathbb{Z})$.

19 November 12, 2018

There is another thing I want to talk about factoring. If G is a finite group, then the order $|G|$ always is a multiple of any order of an element. That is, if m is the smallest positive element such that $g^m = 1$, then m divides $|G|$.

In Pollack's method, we wanted to factor $\gcd(g^m - 1, N)$ for some m we chose. Here, m was chosen to guarantee a correct vectoring, if N has a prime p such that $p - 1$ has small factors.

The elliptic curve method was sort of similar. We chose an elliptic curve

$$E : y^2 = x^3 + ax + 1,$$

with a point we know, namely $P = (0, 1)$. Then we started computing multiples of P, and mP was guaranteed to factor.

19.1 Elliptic curve cryptography

Again, you want to share your secret key with someone else. We first publicly choose a prime p, an elliptic curve

$$E : y^2 = x^3 + ax + b,$$

and a point P on E.

If we want to choose a password, here is how we do it. I secretly choose an integer m, and you secretly choose an integer n. I send you publicly the point mP, and you send me publicly the point nP. Then we can both compute mnP, so this is going to be our secrete key. The key to this being secret is that it is difficult to get mnP from P, mP, and nP. Again, this is the discrete logarithm problem.

This is similar to $(\mathbb{Z}/p\mathbb{Z})^\times$. We fixed some $g \in (\mathbb{Z}/p\mathbb{Z})^\times$, we exchanged g^m and g^n and then used g^{mn} as the secret key. But this will be unclever if we used the additive group $\mathbb{Z}/p\mathbb{Z}$. It is easy to solve the discrete logarithm problem, because if we have g and ng and mg, it is easy to compute g^{-1} and recover n and m.

So now how do we do cryptography? Again, take prime p and an elliptic curve E and a point B on $E(\mathbb{Z}/p\mathbb{Z})$. This is going to all be public. I also choose an integer n that is kept secret, but I publish nB.

Here is how to encrypt the message. If somebody wants to send some point P (containing the message), then the person picks an integer r in secret, and sends the two points

$$rB, \quad P + r(nB).$$

When I receive the message, I can decrypt it by simply taking

$$P = [P + r(nB)] - n[rB],$$

which I can do because I know n. If there are interceptors seeing rB and $P + rnB$, they would not be able to recover P.

19.2 Rational points on elliptic curves

Given an elliptic curve, we can look at the group of rational points,

$$E(\mathbb{Q}) = \{x, y \in \mathbb{Q} : y^2 = x^3 + ax + b\} \cup \{O\}.$$

There is a famous theorem. We are not going to prove it, because it will require too much technology.

Theorem 19.1 (Mordell)**.** *The group $E(\mathbb{Q})$ is finitely generate, i.e., there are a finite number of points $P_1, \dots, P_N$ such that every point P can be written as*

$$P = a_1P_1 + \cdots + a_NP_N$$

for some integers $a_1, \dots, a_N$.

On the other hand, there was an interesting problem in the book.

Proposition 19.2. *The group of real points $E(\mathbb{R})$ is not finitely generated.*

But anyways, any rational point on the curve is the sum of some integer multiples of points,

$$P = a_1P_1 + a_2P_2 + \cdots + a_nP_n.$$

There is also this notion of a torsion point. This is a point $P \in E(\mathbb{Q})$ having the property that $nP = O$ for some positive integer n. By Mordell's theorem, there has to be only finitely many torsion elements. So what torsion groups can we get?

Theorem 19.3 (B. Mazur)**.** *There are only 15 possible groups for $E(\mathbb{Q})_{\mathrm{tors}}$,*

$$\mathbb{Z}/n\mathbb{Z} \text{ for } n \le 10 \text{ or } n = 12, \quad \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2n\mathbb{Z} \text{ for } n \le 4.$$

So now that we know about $E(\mathbb{Q})_{\mathrm{tors}}$, let us now look at the quotient group

$$E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tors}},$$

which is the group of equivalent classes $[g]$ under $g \sim g'$ if $g - g'$ is a torsion element. Then this quotient group is going to be a torsion-free abelian group, and there is a structure theorem.

Proposition 19.4. *Any finitely-generated torsion-free abelian group is isomorphic to some finite product $\mathbb{Z} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$.*

Proof. Take a minimal generating set $\{x_1, \dots, x_N\}$. But what if a number g can be written in two different ways, say

$$g = m_1x_1 + \cdots + m_Nx_N = p_1x_1 + \cdots + p_Nx_N?$$

Then we would get a relation

$$0 = (m_1 - p_1)x_1 + \cdots + (m_N - p_N)x_N$$

between the generators. Let us write

$$0 = r_1 x_1 + \cdots + r_N x_N.$$

If $r_1 = 1$, then we should be able to write x_1 in terms of $x_2, \ldots, x_N$, and so we can throw away x_1 from the set $\{x_1, \ldots, x_N\}$ and still get a generating set. This contradicts minimality.

So here is the general idea. Consider a relation that looks like

$$r_1 x_1 + r_2 x_2 + \cdots + r_n x_n = 0$$

with $1 < r_1 \leq r_2 \leq \cdots$ attained by relabeling, such that n is minimal. Then by changing variables like $x_1 \to y_1 y_2^{-1}$ and $x_1 \to y_2$, you get smaller relations. Using this, you can either make r_i smaller or you get zero, so that we contradict minimality of n or get $r_1 = 1$. $\square$

Fermat's last theorem was proved using elliptic curves. If you have a solution to

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

then Frey showed that you can make an elliptic curve with strange properties. Then Wiles with the help of Taylor showed that such an elliptic curve cannot exist.

20 November 14, 2018

I want to say one more thing about elliptic curves and then move on. There is this **congruent number problem**. Given n a rational number, is there a right angle with area n, such that all the sides have rational length? That is, are there rational numbers a, b, c such that

$$a^2 + b^2 = c^2, \quad \frac{1}{2}ab = n?$$

If we look at integer n, here is the sequence of possible n:

$$5, 6, 7, 13, 14, 15, 20, 21, 22, 23, 24, 28, 29, 30, 31, \dots.$$

If you write $x = \frac{n(a+c)}{b}$ and $y = \frac{2n^2(a+c)}{b^2}$, then the equation we have becomes

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

So if n is a congruent number, then this elliptic curve has a nontrivial solution.

How did anyone come up with this? Let us assume $a \le b < c$, and let $A = \frac{c^2}{4} = (\frac{c}{2})^2$. Then

$$A + n = \frac{c^2}{4} + \frac{ab}{2} = \frac{a^2+b^2}{4} + \frac{1}{2}ab = \left(\frac{a+b}{2}\right)^2,$$
$$A - n = \frac{a^2+b^2}{4} - \frac{1}{2}ab = \left(\frac{a-b}{2}\right)^2.$$

So we have three squares, and then if we multiply them, we get

$$y^2 = A(A+n)(A-n) = A^3 - n^2 A.$$

It's not obvious how to go back, but it can be done.

20.1 Continued fractions

Now we want to move to a different topic. We are going to look at fractions like

$$[a_0, a_1, a_2, \dots] = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

for $a_0, a_1, a_2, \dots$ positive, and we call them **continued fractions**. If this is finite, so that we have $[a_0, a_1, \dots, a_n]$, we can compute this. We say that this is a simple continued fraction if $a_0, a_1, a_2, \dots$ are all positive integers.

Theorem 20.1. *Every real number greater than 1 can be written uniquely as simple continued fraction. The number is rational if and only if it is finite. (Here is what we mean by uniquely. If the continued fraction is finite, then the last nonzero number has to be greater than 1. Otherwise, we can erase that and add 1 to the number before.)*

How do we represent a real number? We can define

$$[a_0, a_1, a_2, \ldots] = r = \lim_{n\to\infty} [a_0, a_1, \ldots, a_n].$$

This always converges if all a_k are positive integers. Here is an example. We have

$$\frac{1+\sqrt{5}}{2} = [1, 1, 1, 1, \ldots], \quad e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, \ldots].$$

There is something iterative about continued fractions. We have

$$a_0 + \frac{1}{a_1} = \frac{a_0a_1 + 1}{a_1}, \quad a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_0(a_1 + \frac{1}{a_2}) + 1}{a_1 + \frac{1}{a_2}} = \frac{a_0a_1a_2 + a_0 + a_2}{a_1a_2 + 1}.$$

So we see that if we have $[a_0, a_1, a_2 + r]$ and know the continued fraction of $\frac{1}{r}$, then we can write

$$[a_0, a_1, a_2 + r] = [a_0, a_1, a_2, \tfrac{1}{r}].$$

Proposition 20.2. *For a sequence $a_0, a_1, \ldots, a_n$, define a sequence*

$$\begin{gathered}(p_{-2}, q_{-2}) = (0, 1), \quad (p_{-1}, q_{-1}) = (1, 0), \quad (p_0, q_0) = (a_0, 1),\\ (p_k, q_k) = (a_kp_{k-1} + p_{k-2}, a_kq_{k-1} + q_{k-2}).\end{gathered}$$

Then we have

$$\frac{p_n}{q_n} = [a_0, a_1, a_2, \ldots, a_n].$$

Proof. We do this by induction. For $n = 0$, this can be checked directly because $[a_0] = a_0$ and $p_0 = a_0$ and $q_0 = 1$. If this is true for $k-1$, then we can write

$$\begin{aligned}[a_0, a_1, \ldots, a_k] = [a_0, a_1, \ldots, a_{k-1} + \tfrac{1}{a_k}] &= \frac{p'_{k-1}}{q'_{k-1}} = \frac{(a_{k-1} + \frac{1}{a_k})p_{k-2} + p_{k-3}}{(a_{k-1} + \frac{1}{a_k})q_{k-2} + q_{k-3}}\\ &= \frac{\frac{1}{a_k}p_{k-2} + p_{k-1}}{\frac{1}{a_k}q_{k-2} + q_{k-1}} = \frac{p_{k-2} + a_kp_{k-1}}{q_{k-2} + a_kq_{k-1}} = \frac{p_k}{q_k}.\end{aligned}$$

So we get the formula for k. ☐

From this we get some corollaries.

Corollary 20.3. *We have*

$$p_nq_{n-1} - q_np_{n-1} = (-1)^{n-1}, \quad p_nq_{n-2} - q_np_{n-2} = (-1)^na_n.$$

Therefore

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_nq_{n-1}}, \quad \frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^na_n}{q_nq_{n-2}}.$$

Proof. If you expand $p_k q_{k-1} - q_k p_{k-1}$, we get

$$p_k q_{k-1} - q_k p_{k-1} = -(p_{k-1} q_{k-2} - q_{k-1} p_{k-2})$$

and so you get a (-1) each time you go down. For the second formula, we use this to get

$$\begin{aligned} p_k q_{k-2} - q_k p_{k-2} &= (a_k p_{k-1} + p_{k-2}) q_{k-2} - (a_k q_{k-1} + q_{k-2}) p_{k-2} \\ &= a_k (p_{k-1} q_{k-2} - q_{k-1} p_{k-2}) = a_k (-1)^{k-1}. \end{aligned}$$

Then we can use these to get the formulas about the fractions. $\square$

This is what will give us convergence. Define

$$c_n = [a_0, a_1, a_2, \ldots, a_n] = \frac{p_n}{q_n}.$$

Then these formulas tell us that the even c are increasing, and the ood c are decreasing, and the odd ones are greater than the even ones:

$$c_0 < c_2 < c_4 < c_6 < \cdots\cdots < c_7 < c_5 < c_3 < c_1.$$

Then because q_k only increases by at least 1, we see that $q_k \to \infty$ as $k \to \infty$, so the gap between these numbers should decrease and the sequence has to converge.

21 November 19, 2018

So we were looking at continued fractions. Simple continued fractions are like

$$[a_0, a_1, a_2, \ldots] = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\cdots}}}$$

for positive integers $a_0, a_1, \ldots$, and we require that if it is finite, the last one is greater than 1. There is this iterative nature

$$[a_0, a_1, \ldots, a_n, r] = [a_0, a_1, \ldots, a_n + \tfrac{1}{r}].$$

To show that infinite fractions are well-defined we looked at these partial convergents and showed that we can write

$$c_n = [a_0, \ldots, a_n] = \frac{p_n}{q_n}$$

if we define

$$\begin{pmatrix} p_{-1} \\ q_{-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} p_0 \\ q_0 \end{pmatrix} = \begin{pmatrix} a_0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} p_k \\ q_k \end{pmatrix} = a_k \begin{pmatrix} p_{k-1} \\ q_{k-1} \end{pmatrix} + \begin{pmatrix} p_{k-2} \\ q_{k-2} \end{pmatrix}.$$

We showed from these that

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{(-1)^{n-1}}{q_n q_{n-1}}, \quad \frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = \frac{(-1)^n a_n}{q_n q_{n-2}},$$

and so we have

$$c_{2k} < c_{2k+2} < c_{2k+4} < \cdots < c_{2k+5} < c_{2k+3} < c_{2k+1}.$$

Here, we defined q_n by the recursion $q_n = a_n q_{n-1} + q_{n-2}$, and because $a_n \geq 1$, we have $q_n \geq q_{n-1}$. So

$$q_n = a_n q_{n-1} + q_{n-2} \geq q_{n-1} + q_{n-2} \geq q_{n-2} + q_{n-2} \geq 2q_{n-2}$$

shows that q_n grows at least like $2^{n/2}$. Using the formula we had above, we see that

$$\left| \frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}} \right| \leq \frac{1}{2^{n-1}}.$$

This shows that the numbers $c_n = p_n/q_n$ indeed converges to some value.

21.1 Representing as a continued fraction

Theorem 21.1. *A rational number (greater than or equal to* 1*) has a unique finite simple continued fraction representation. An irrational number (greater than* 1*) has a unique infinite one.*

How can we prove this? Let's first look at the rational case. We see that given any finite simple continued fraction

$$[a_0, a_1, \dots, a_n] = \frac{p_n}{q_n},$$

this is rational. Moreover, the two integers p_n and q_n are necessarily relatively prime, because the formula for $p_n/q_n - p_{n-1}/q_{n-1}$ gives

$$p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1},$$

so they can't share a common prime factor. Now given a number m, how do we write this as a continued fraction? We first write

$$m = a_0 + r_0, \quad a_0 \in \mathbb{Z}, \quad r_0 \in [0, 1).$$

Then unless $r_0 = 0$, in which case we just have $m = [m]$, we can take its reciprocal and $1/r_0$ is greater than 1. So we should be able to write

$$\frac{1}{r_0} = a_1 + r_1, \quad a_1 \in \mathbb{Z}_{\geq 1}, \quad r_1 \in [0, 1), \quad m = a_0 + \frac{1}{a_1 + r_1}.$$

Then we just repeat this by writing $1/r_1 = a_2 + r_2$ and so on, and at the end we will get

$$m = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{\cdots}}}.$$

Note that this is the only thing we can do, because a_0 has to be the integer part of m, and then if a_0 is determined, then a_1 has to be the integer part of $1/(m - a_0)$, and so on. This shows that the representation of a real number as a continued fraction is unique.

Let us try to do this with a rational number, say $m = \frac{b}{a}$ where $a < b$. The claim is that this will be just the Euclidean algorithm. If we write

$$b = aq_1 + r_1, \quad 0 \leq r_1 < a,$$

then we can write

$$\frac{b}{a} = q_1 + \frac{r_1}{a}, \quad 0 \leq \frac{r_1}{a} < 1, \quad m = \frac{b}{a} = q_1 + \frac{r_1}{a}.$$

Then we find

$$a = r_1 q_2 + r_2, \quad 0 \leq r_2 < r_1,$$

and this will give

$$\frac{a}{r_1} = q_2 + \frac{r_2}{r_1}, \quad 0 \leq \frac{r_2}{r_1} < 1, \quad m = \frac{b}{a} = q_1 + \frac{1}{\frac{a}{r_1}} = q_1 + \frac{1}{q_2 + \frac{r_2}{r_1}}.$$

So trying to write $m = \frac{b}{a}$ has a continued fractions is equivalent to doing the Euclidean algorithm, and this shows that the process must end at some point.

Proposition 21.2. *If $m \geq 1$ is a rational number, then m can be represented as a finite simple continued fraction.*

If m is irrational, we can still do the process, but this cannot end after a finite step, since if it does then m has to be rational. But we still have that the continued fraction actually converges to m. This is because if we have

$$m = [a_0, a_1, \dots, a_{n+1} + r_{n+1}]$$

at any middle step, then we see that m is between the two numbers

$$[a_0, a_1, \dots, a_{n+1}], \quad [a_0, a_1, \dots, a_n].$$

(These are the two things we get when we see $r_{n+1} \to 0$ and $r_{n+1} \to \infty$.) Moreover, the difference between these two numbers is

$$\left| x - \frac{p_n}{q_n} \right| \leq \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}}.$$

Since $1/(q_n q_{n+1}) \to 0$ as $n \to \infty$, we see that $p_n/q_n \to x$ as $n \to \infty$.

This error bound can be used to show that that for a rational number, the expansion should stop at some point. Also, continued fractions provide the "best rational approximation" given the size of the denominator.

21.2 Repeating continued fractions

If we do base 10 (or any base) expansion of a rational number, we ultimately get a periodic sequence. We get periodic sequences for numbers involving one square root. For instance, if we try to expand

$$x = \frac{1 + \sqrt{5}}{2} \approx 1.6,$$

then we can write this as

$$x = 1 + \frac{1}{\frac{2}{\sqrt{5}-1}} = 1 + \frac{1}{x}.$$

So we can repeat this process and get

$$x = \frac{1 + \sqrt{5}}{2} = 1 + \frac{1}{1 + \frac{1}{1 + \frac{1}{\cdots}}} = [1, 1, 1, 1, \dots].$$

What about $y = \sqrt{2}$? We can get

$$y = 1 + (\sqrt{2} - 1) = 1 + \frac{1}{\sqrt{2} + 1} = 1 + \frac{1}{2 + (\sqrt{2} - 1)} = 1 + \frac{1}{2 + \frac{1}{\sqrt{2}+1}}$$

and so repeating this will give

$$y = \sqrt{2} = [1, 2, 2, 2, \dots].$$

We can do this for $\sqrt{3}$, and this will give

$$\sqrt{3} = [1, 1, 2, 1, 2, 1, 2, \dots].$$

Definition 21.3. A simple continued fraction $[a_0, a_1, a_2, \ldots]$ is said to be **repeating** if it takes the form of

$$[a_0, a_1, \ldots, a_n, a_{n+1}, \ldots, a_{n+k}, a_{n+1}, \ldots, a_{n+k}, a_{n+1}, \ldots].$$

In this case, we also write this as

$$[a_0, a_1, \ldots, a_n, \overline{a_{n+1}, \ldots, a_{n+k}}].$$

In fact, given our computations above, we can make a bold conjecture that square roots have repeating continued fractions.

Definition 21.4. An irrational number x is said to be a **quadratic irrational** if it obeys an equation that look like

$$ax^2 + bx + c = 0,$$

where a, b, c are all integers.

Theorem 21.5. *Any quadratic irrational has a repeating continued fraction expansion. Conversely, every repeating continued fraction is a quadratic irrational.*

First let us show that any repeating continued fraction should be quadratic. Suppose that we have

$$x = [a_0, a_1, \ldots, a_n, a_0, a_1, \ldots, a_n, \ldots].$$

Then we can also write this as

$$x = [a_0, a_1, \ldots, a_n, x].$$

So according to our theorem, we should be able to write

$$x = \frac{\hat{p}_{n+1}}{\hat{q}_{n+1}} = \frac{xp_n + p_{n-1}}{xq_n + q_{n-1}}$$

and this gives

$$x^2 q_n + x(q_{n-1} - p_n) - p_{n-1} = 0.$$

This is a quadratic equation. The hard part is to show that if it obeys a quadratic equation then it has a continued fraction expansion.

22 November 26, 2018

We were looking at periodic continued fractions. We showed that if some number has a periodic continued fraction expansion, then it is a root of a quadratic equation with integer coefficients. By the way, we are using Stein's notations. Here is an interesting fact.

Proposition 22.1. *If $|x - \frac{a}{b}| < \frac{1}{2b^2}$ then $\frac{a}{b}$ is one of the convergents for x, so that there exists some n such that $p_n = a$ and $q_n = b$.*

Proof. Fix a n such that $q_n \le b \le q_{n+1}$. We have seen that

$$\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}.$$

If $\frac{a}{b}$ lies between x and $\frac{p_n}{q_n}$, then

$$\left| \frac{a}{b} - \frac{p_n}{q_n} \right| < \frac{1}{q_n q_{n+1}}, \quad |aq_n - bp_n| < \frac{b}{q_{n+1}} < 1$$

and so $\frac{p_n}{q_n} = \frac{a}{b}$. There are more cases, and you can do similar things. □

22.1 Quadratic irrationals have periodic continued fractions

We have things like

$$\sqrt{5} = [2, 4, 4, 4, \dots], \quad \sqrt{3} = [1, 2, 2, 2, \dots].$$

Definition 22.2. A continued fraction is called **periodic** or **repeating** if it looks like

$$[a_0, \dots, a_n, \dots, a_{n+k}, a_n, \dots, a_{n+k}, \dots]$$

and it is called **purely periodic** if it looks like

$$[a_0, \dots, a_n, a_0, \dots, a_n, a_0, \dots].$$

Theorem 22.3. *A continued fraction is periodic if and only if x is a quadratic irrational.*

Last time we showed that a purely periodic continued fraction is quadratic irrational. This can be extended to all periodic continued fractions. Given any periodic continued fraction α, we can write

$$\alpha = [a_0, \dots, a_n, x], \quad x = [b_0, \dots, b_k, b].$$

Here, x is purely periodic, so we can find a quadratic equation for x. Then we have

$$\alpha = \frac{p_n x + p_{n-1}}{q_n x + q_{n-1}}, \quad x = \frac{p\alpha + q}{r\alpha + s}$$

for some rational numbers p, q, r, s, and so we can plug this into the quadratic equation for x and get a quadratic equation for α. This proves one direction.

Now let us prove the other direction. Let x solve the quadratic equation

$$ax^2 + bx + c = 0, \quad x = [a_0, r_0] = [a_0, a_1, r_1] = \cdots = [a_0, \dots, a_n, r_n] = \cdots .$$

We want to show that $r_n = r_{n+h}$ for some n and $h > 0$. We note that

$$x = \frac{r_n p_n + p_{n-1}}{r_n q_n + q_{n-1}}, \quad a(r_n p_n + p_{n-1})^2 + b(r_n p_n + p_{n-1})(r_n q_n + q_{n-1}) + c(r_n q_n + q_{n-1})^2 = 0.$$

This is like the quadratic form $aX^2 + bXY + cY^2$, with $Y = 1$. Then we are making a change of basis

$$x = Xp_n + Yp_{n-1}, \quad y = Xq_n + Yq_{n-1}.$$

This really is a change of basis because

$$\det \begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix} = p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1} = \pm 1.$$

So instead of looking at the numbers and showing that there is a repetition there, we can instead show that the equations (which corresponds to a quadratic form) repeats itself over.

Let us actually expand the equation for r_n and see what it looks like. This is given by

$$\begin{aligned}[ap_n^2 + bp_n q_n + cq_n^2]r_n^2 + [2ap_n p_{n-1} + b(p_n q_{n-1} + q_n p_{n-1}) + 2cq_n q_{n-1}]r_n \\ + [ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2] = 0.\end{aligned}$$

Let us write this as $A_n r_n^2 + B_n r_n + C_n = 0$. Then since we are making only a unimodular transformation, the discriminant doesn't change, that is, we always have

$$B_n^2 - 4A_n C_n = b^2 - 4ac.$$

Moreover, $C_n = A_{n-1}$. So if we can show that there are only finitely many values that can appear as A_n for n large enough, then we show that there are only finitely many choices for B_n, so the numbers must repeat themselves.

Therefore we focus on A_n. Note that we have

$$p_n = q_n x + \delta, \quad |\delta| \leq \frac{1}{q_{n-1}}.$$

Then we can write

$$\begin{aligned} A_n &= qp_n^2 + bp_n q_n + cq_n^2 \\ &= q_n^2(ax^2 + bx + c) + (2ax + b)q_n\delta + a\delta^2 = (2ax + b)q_n\delta + a\delta^2. \end{aligned}$$

Here, $|q_n \delta| < 1$ and so the possible values for A_n are bounded. This finishes the proof that the equations must repeat at some point.

22.2 Pell's equations

Consider the equation

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

where N is a given fixed integer. If $N = d^2$ is a square number, this is not very interesting because this factors as $(x+dy)(x-dy) = 1$. But if N is not a square, this becomes interesting.

Why I am bringing this up when we were talking about continued fractions? If $y \neq 0$, we can write this as

$$\frac{x^2}{y^2} = N + \frac{1}{y^2}, \quad \frac{x}{y} = \sqrt{N}(1 + \tfrac{1}{Ny^2})^{1/2}.$$

So this is a really good rational approximation to $\sqrt{N}$. For ϵ really small, $(1+\epsilon)^{1/2}$ behaves like $1+\epsilon/2$. So the approximation is better than $1/\sqrt{y}^2$. This shows that once we have a nontrivial solution to Pell's equation, it is going to be one of the convergents.

Proposition 22.4. *If (x, y) with $x, y > 0$ obeys $x^2 - Ny^2 = 1$, then $\frac{x}{y}$ is $\frac{p_n}{q_n}$ for $\sqrt{N} = [a_0, a_1, \dots]$ and some n.*

Let's check this with an example. If we look at the expansion of $\sqrt{2}$, we get

$$\sqrt{2} = [1, 2, 2, 2, \dots]$$

then we can the convergents

$$\frac{3}{2}, \quad \frac{7}{5}, \quad \frac{17}{12}, \dots.$$

Then we see that

$$3^2 - 2 \cdot 2^2 = 1, \quad 7^2 - 2 \cdot 5^2 = -1, \quad 17^2 - 2 \cdot 12^2 = 1, \dots.$$

In fact, the odd convergents give 1 and the even convergents give -1.

So in general, we want to show that Pell's equations have a solution. If we write

$$\sqrt{N} = [a_0, a_1, \dots, a_n, r_n],$$

then we can write

$$\sqrt{N} = \frac{r_n p_n + p_{n-1}}{r_n q_n + q_{n-1}}.$$

Then the distance from $\frac{p_n}{q_n}$ is given by

$$\sqrt{N} - \frac{p_n}{q_n} = \frac{(-1)^n}{(r_n q_n + q_{n-1}) q_n}.$$

Because we want $p_n = x$ and $q_n = y$, we need n to be odd. So let us multiply by q_n and square both sides. Then we get

$$Nq_n^2 + \frac{2\sqrt{N} q_n}{r_n q_n + q_{n-1}} + \frac{1}{(r_n q_n + q_{n-1})^2} = p_n^2.$$

So

$$p_n^2 - Nq_n^2 = \frac{2\sqrt{N}q_n}{r_nq_n + q_{n-1}} + \frac{1}{(r_nq_n + q_{n-1})^2}.$$

Because the left hand side is an integer and the right hand side is positive, we only need to show that r_n sometimes get big enough, so that $2\sqrt{N}q_n/(r_nq_n + q_{n-1})$ has to be exactly 1.

23 November 28, 2018

We were studying the equation $x^2 - Ny^2 = 1$, where N is square-free. Then we could show that
$$\left|\frac{x}{y} - \sqrt{N}\right| < \frac{1}{\sqrt{N}y^2}.$$
So if there is a solution, it has to be one of the convergents for the continued fraction. In fact, if we look at the n-th convergent p_n/q_n, then
$$p_n^2 - Nq_n^2 = \pm\frac{2\sqrt{N}q_n}{r_nq_n + q_{n-1}} + \frac{1}{(r_nq_n + q_{n-1})^2}, \quad \sqrt{N} = [a_0, \dots, a_n, r_n].$$
The term $1/(r_nq_n + q_{n-1})^2$ gets really small, so it is enough to show that the other term is smaller than 2 at some point.

23.1 Existence of nontrivial solutions for Pell's equation

We have
$$\frac{2\sqrt{N}q_n}{r_nq_n + q_{n-1}} < \frac{2\sqrt{N}q_n}{a_{n+1}q_n + q_{n-1}} < \frac{2\sqrt{N}}{a_{n+1}}.$$
So it is enough to show that there is a a_n that is bigger than $\sqrt{N}$. In fact, here is what we can prove.

Theorem 23.1. *Let the continued fraction of $\sqrt{N}$ be*
$$\sqrt{N} = [a_0, \dots, a_n, \overline{a_{n+1}, \dots, a_{n+h}}].$$
Then $a_{n+h} = 2\lfloor\sqrt{N}\rfloor$, where $\lfloor\sqrt{N}\rfloor$ is the greatest integer less than $\sqrt{N}$.

Proof. Note that $a_0 = \lfloor\sqrt{N}\rfloor$. We can write
$$\sqrt{N} = a_0 + (\sqrt{N} - a_0) = a_0 + \frac{1}{\frac{\sqrt{N}+a_0}{N-a_0^2}}.$$
So if we consider $x = \sqrt{N} + a_0$, we get a continued fraction
$$x = 2a_0 + \frac{1}{(\frac{2a_0}{N-a_0^2}) + \frac{1}{x}} = [2a_0, \gamma_0, 2a_0, \gamma_0, 2a_0, \gamma_0, \dots], \quad \gamma_0 = \frac{2a_0}{N - a_0^2}.$$
If γ_0 is an integer, this really is going to be a simple continued fraction expansion of x. Then we would also have
$$\sqrt{N} = [a_0, \gamma_0, 2a_0, \gamma_0, 2a_0, \gamma_0, \dots].$$
For instance, for $N = 2$ or 3 or 5 or 6 or 10, this does give the right expansion. But for $N = 7$, it does not work. $\square$

Well, for those N that does not work, you need to do more complicated things.

23.2 The Riemann zeta function

This is given by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \cdots .$$

The infinite series converges for $s > 1$. Why does it converge? We can show that

$$\sum_{n=1}^{T} \frac{1}{n^s} \leq 1 + \int_1^T \frac{dx}{x^2} = 1 + \frac{1}{1-s}(T^{1-s} - 1) \leq 1 + \frac{1}{s-1}$$

is $s \geq 1$. This shows that the series indeed converges, and we can show the other side inequality

$$1 \leq (s-1)\zeta(s) \leq 1 + (s-1)$$

for $s > 1$. So $\zeta(s)$ indeed grows as fast as $1/(s-1)$. This also implies that

$$\lim_{s \to 1} \frac{\log \zeta(s)}{|\log(s-1)|} = 1.$$

Let's now see why this says something about the distribution of primes.

Proposition 23.2. *For $s > 1$, we have*

$$\zeta(s) = \prod_{p \text{ primes}} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Does this make sense at all? What are infinite products? Note that all the terms we are multiplying are bigger than 1, but they become really close to 1 as $p \to \infty$. So we can define this product as

$$\prod_{p \text{ primes}} \left(1 - \frac{1}{p^s}\right)^{-1} = \lim_{N \to \infty} \prod_{p<N} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Because it is getting bigger and bigger, this converges if and only if the products don't go to infinity.

Let me first explain formally why this infinite product should be the zeta function. We first note that

$$\frac{1}{1 - \frac{1}{p^s}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots .$$

So if we multiply all the powers of $1/2^{ks}$ and $1/3^{ls}$ together, we get

$$\frac{1}{1 - \frac{1}{2^s}} \frac{1}{1 - \frac{1}{3^s}} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{6^s} + \frac{1}{8^s} + \frac{1}{9^s} + \frac{1}{12^s} + \cdots ,$$

where the right hand side has only number that can be written as $2^k 3^l$, for $k, l \geq 0$. If we multiply by $1/(1 - 1/5^s)$, then we further get numbers like

$2^k 3^l 5^m$ after distributing out. By the fundamental theorem of arithmetic, every number has a unique prime factorization, so once we multiply all prime factors, we get every integer exactly once.

Now we can use take this formula and take the log of both sides. Then

$$\log \zeta(s) = -\sum_p \log\left(1 - \frac{1}{p^s}\right).$$

We know that $-\log(1-x)$ looks like

$$-\log(1-x) = x + (\text{error term of the order } x^2)$$

by Taylor's theorem. Then we can add them to obtain a bound like

$$\log \zeta(s) = \sum_p \frac{1}{p^s} + (\text{error term of the order } \zeta(2)),$$

and $\zeta(2) = \pi^2/6$. This means that the error term is of constant size, at most.

Here is what we can say immediately. If the nth prime has size $p_n > n^{1.1}$ for all n, for instance, we would have that the right hand side converges to $s = 1.0001$. But since the left hand side diverges, it couldn't be true. So the primes are pretty common. We can also look at

$$\frac{\sum p^{-s}}{|\log(s-1)|} \to 1$$

as $s \to 0$ and use this to do finer analysis. Dirichlet also had this idea of summing only over primes of a certain form.

24 December 3, 2018

I was talking about the Riemann zeta function,

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

This converges for $s > 1$ and we further have an estimate

$$\frac{1}{s-1} < \zeta(s) < \frac{1}{s-1} + 1.$$

So we have $\log \zeta(s)/|\log(s-1)| \to 1$ as $s \to 1$. There also was this product expansion

$$\alpha(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

Using this, we showed that

$$\log \zeta(s) = \sum_p \frac{1}{p^s} + R, \quad |R| < 4\zeta(2).$$

This implies that

$$\lim_{s\to 1} \frac{\sum \frac{1}{p^s}}{|\log(s-1)|} = 1.$$

24.1 Dirichlet's L-function

We now want to study not all primes but a certain subset of primes.

Definition 24.1. Let $\mathcal{P}$ be some set of primes, not necessarily all primes. Then we define its **Dirichlet density** as

$$d(\mathcal{P}) = \lim_{s\to 1} \frac{\sum_{p\in\mathcal{P}} p^{-s}}{\log|s-1|}.$$

For instance, we have just seen that $d(\{\text{all primes}\}) = 1$. If we take the density of all but finitely many primes, we would still have

$$d(\{\text{all but finitely many primes}\}) = 1.$$

Similarly, any finite set would have density 0.

Theorem 24.2 (Dirichlet)**.** *Let $a, m > 0$ be positive integers, relatively prime. Consider the set*

$$\mathcal{P} = \{\textit{primes of the form } p = mk + a\}.$$

Then its Dirichlet density is $d(\mathcal{P}) = 1/\phi(m)$, where ϕ is Euler's function.

This is really amazing. If we look at $m = 3$, for instance, we have the following list of primes of the form $3k+1$ and $3k+2$:

$$\begin{aligned} &3k+1 : 7, 13, 19, 31, 37, 43, 61, 67, 73, 79, 97, 103, \ldots, \\ &3k+2 : 2, 5, 11, 17, 23, 29, 41, 47, 53, 59, 71, 83, 89, 101, \ldots. \end{aligned}$$

The theorem is saying that they have approximately the same number of primes. Let us first see how to prove this for $m = 4$.

Definition 24.3. A **character** is a map

$$\chi : \mathbb{Z} \to \mathbb{C}$$

that satisfies $\chi(nm) = \chi(n)\chi(m)$, $\chi(0) = 0$, $\chi(1) = 1$ and is periodic.

Here is an example. We can take

$$\chi(\text{even}) = 0, \quad \chi(4k+1) = 1, \quad \chi(4k+3) = -1.$$

Then we can check that this is indeed multiplicative because the product of $4k+1$s is $4k+1$, the product of a $4k+1$ and a $4k+3$ is $4k+3$, and the product of two $4k+3$s is $4k+1$. The good thing about multiplicativity is that if we have a prime factorization $n = p_1^a p_2^b \cdots$ then we get

$$\chi(n) = \chi(p_1)^a \chi(p_2)^b \cdots.$$

Definition 24.4. We define the **Dirichlet L-function** associated to the character χ as

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

In the above case, we have

$$L(s, \chi) = 1 - \frac{1}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + \frac{1}{9^s} - \frac{1}{11^s} + \cdots.$$

Because this is an alternating sum, this implies that

$$1 - \frac{1}{3^s} < L(s, \chi) < 1.$$

By multiplicativity, we can also write

$$L(s, \chi) = \prod_{p \neq 2} \left(1 - \frac{\chi(p)}{p^s}\right).$$

Then the log of the L function is

$$\log L(s, \chi) = -\sum_{p \neq 2} \log\left(1 - \frac{\chi(p)}{p^s}\right) = \sum_{p \neq 2} \frac{\chi(p)}{p^s} + R$$

for some remainder term $|R| < 4\zeta(2)$. If we add this to the log of the zeta function, we get

$$\log L(s,\chi) + \log\zeta(s) = \frac{1}{2^s} + \sum_{p=4k+1} \frac{2}{p^s} + R^*$$

for some bigger error R^*. Similarly, we have

$$-\log L(s,\chi) + \log\zeta(s) = \frac{1}{2^s} + \sum_{p=4k+3} \frac{2}{p^s} + R^{**}.$$

Now we can divide by $|\log(s-1)|$ and see what happens as $s \to 1$. Because $\log L(s,\chi)$, $1/2^s$, R^*, R^{**} are all bounded, this does not contribute anything. Therefore

$$d(\{p = 4k+1\}) = \lim_{s\to 1} \frac{\sum_{p=4k+1} p^{-s}}{|\log(s-1)|} = \lim_{s\to 1} \frac{\frac{1}{2}\log\zeta(s)}{|\log(s-1)|} = \frac{1}{2},$$
$$d(\{p = 4k+3\}) = \lim_{s\to 1} \frac{\sum_{p=4k+3} p^{-s}}{|\log(s-1)|} = \lim_{s\to 1} \frac{\frac{1}{2}\log\zeta(s)}{|\log(s-1)|} = \frac{1}{2}.$$

We can do a similar thing for $m = 3$. Here, the character we use is

$$\chi(3k) = 0, \quad \chi(3k+1) = 1, \quad \chi(3k+2) = -1,$$

and then we have

$$L(s,\chi) = 1 - \frac{1}{2^s} + \frac{1}{4^s} - \frac{1}{5^s} + \frac{1}{7^s} - \frac{1}{8^s} + \cdots,$$

which is again an alternating series, and hence $1/2 < L(s,\chi) < 1$. Now we again take log of this and ζ, add and subtract them, and this will give us exactly the Dirichlet densities of primes of the form $3k+1$ and $3k+2$.

For $m = 5$, we won't exactly be able to do the same thing, because there are four possibilities: $1, 2, 3, 4$ modulo 5. These still are approximately the same number of primes:

$$\begin{aligned}
5k+1 &: 11, 31, 41, 61, 71, 91, 101, \ldots \\
5k+2 &: 2, 7, 17, 37, 47, 67, 97, 107, \ldots \\
5k+3 &: 3, 13, 23, 43, 53, 73, 83, 103, \ldots \\
5k+4 &: 19, 29, 59, 79, 89, 109, \ldots.
\end{aligned}$$

So what do we do? First, we note that $n^4 \equiv 1 \pmod 5$ for n not divisible by 5. So this means that we should have

$$\chi(n)^4 = \chi(n^4) = \chi(1) = 1$$

assuming that χ is periodic with period 5. There are actually four complex numbers whose fourth power is 1. So we can take

$$\begin{aligned}
\chi_1 &: 5k+1 \mapsto 1, \quad 5k+2 \mapsto i, \quad 5k+3 \mapsto -i, \quad 5k+4 \mapsto -1,\\
\chi_2 &: 5k+1 \mapsto 1, \quad 5k+2 \mapsto -1, \quad 5k+3 \mapsto -1, \quad 5k+4 \mapsto 1,\\
\chi_3 &: 5k+1 \mapsto 1, \quad 5k+2 \mapsto -i, \quad 5k+3 \mapsto i, \quad 5k+4 \mapsto -1,\\
\chi_0 &: 5k+1 \mapsto 1, \quad 5k+2 \mapsto 1, \quad 5k+3 \mapsto 1, \quad 5k+4 \mapsto 1.
\end{aligned}$$

These are all characters, and we have exactly the right number of characters so that we would be able to distinguish all the classes apart.

25 December 5, 2018

We introduced the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ for $s > 1$, and then we got $\log \zeta(s)/\log|s-1| \to 1$ as $s \to 1$. Then using the Euler product formula, we were able to estimate

$$\frac{\sum_p \frac{1}{p^s}}{|\log(s-1)|} \to 1$$

as $s \to 1$.

25.1 Dirichlet's theorem on primes on arithmetic progressions

Dirichlet's strategy for selecting out certain primes was to use these characters. For instance, let us try to distinguish prime modulo 5. We want to look at multiplicative maps

$$\chi : \mathbb{Z} \to \mathbb{C}, \quad \chi(n + 5k) = \chi(n).$$

Then we need $1 = \chi(n^4) = \chi(n)^4$ and so all $\chi(n)$ should look like 0 or $1, i, -1, -i$.

In general, here is how you do this for $\mathbb{Z}/p\mathbb{Z}$, for p a prime. We take a primitive root g for p, and then we see that

$$1 = \chi(1) = \chi(g^{p-1}) = \chi(g)^{p-1}.$$

So $\chi(g)$ should be a $(p-1)$-th root of unity. Once we know $\chi(g)$, we know all $\chi(g^2), \dots, \chi(g^{p-2})$ as well, and this forms a complete set of residues for $\mathbb{Z}/p\mathbb{Z}$. So we completely determine the character. In particular, for $p = 5$, we get the 4 characters from choosing $\chi(2) = 1, i, -1, -i$. Let us call

$$\chi_1(2) = i, \quad \chi_2(2) = -1, \quad \chi_3(2) = -1, \quad \chi_4(2) = 1.$$

Once we have a character, we can look at the Dirichlet L-function

$$L(s, \chi) = \sum_{5 \nmid n} \frac{\chi(n)}{n^s} = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

For instance, we have

$$L(s, \chi_1) = 1 + \frac{i}{2^s} - \frac{i}{3^s} - \frac{1}{4^s} + \frac{1}{6^s} + \frac{i}{7^s} - \frac{i}{8^s} - \frac{1}{9^s} + \frac{1}{11^s} + \cdots$$

Here, we note that the real parts and imaginary parts alternate in sign, as long as $s > 0$. So this converges on $s > 0$ and moreover it satisfies $L(1, \chi_1) \neq 0$. For the second character, we see that

$$L(s, \chi_2) = 1 - \frac{1}{2^s} - \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{6^s} - \frac{1}{7^s} - \frac{1}{8^s} + \frac{1}{9^s} + \frac{1}{11^s} - \cdots$$

and here, we can group the things with the same sign. So technically this is not an alternating series but you can group the two terms and then this is an alternating series. So we get the same conclusion that it converges as along as $s > 0$ and also $L(1, \chi_2) \neq 0$. The third character is just the complex conjugate of the first character, and the fourth one is our usual Riemann zeta function with the prime 5 thrown away.

Now we want to take its logarithm. But these are complex numbers, so we need to think about what it means to take the log. If we write any complex number $a + ib$, then we can write

$$a + ib = r(\cos\theta + i\sin\theta) = re^{i\theta},$$

then we can define $\log(a + ib) = \log r + i\theta$. Here, we should be a bit careful because θ is defined only up to adding or subtracting 2π. We see that for r and θ very small, so that $(1 + r)e^{i\theta}$ is close to 1, we have

$$\log((1 + r)e^{i\theta}) = \log(1 + r) + i\theta \sim r + i\theta$$

while we have

$$(1 + r)e^{i\theta} = (1 + r)(\cos\theta + i\sin\theta) \sim 1 + r + i\theta.$$

This means that

$$\log 1 + u \sim u$$

even for u small complex numbers.

But anyways, let us take the log. We see that

$$\log \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = -\sum_p \log\left(1 - \frac{\chi(p)}{p^s}\right) = -\sum_p \left(-\frac{\chi(p)}{p^s}\right) + R(p)$$

where $|R(p)| \leq O(p^{-2s})$. This shows that when we take logarithms of the Dirichlet L-functions, we obtain

$$\begin{aligned}
\log L(s, \chi_4) &= \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{7^2} + \frac{1}{11^s} + \frac{1}{13^s} + \frac{1}{17^s} + \frac{1}{19^s} + \cdots, \\
\log L(s, \chi_1) &= \frac{i}{2^s} - \frac{i}{3^s} + \frac{i}{7^2} + \frac{1}{11^s} - \frac{i}{13^s} + \frac{i}{17^s} - \frac{1}{19^s} + \cdots, \\
\log L(s, \chi_3) &= -\frac{i}{2^s} + \frac{i}{3^s} - \frac{i}{7^2} + \frac{1}{11^s} + \frac{i}{13^s} - \frac{i}{17^s} - \frac{1}{19^s} + \cdots, \\
\log L(s, \chi_2) &= -\frac{1}{2^s} - \frac{1}{3^s} - \frac{1}{7^2} + \frac{1}{11^s} - \frac{1}{13^s} - \frac{1}{17^s} + \frac{1}{19^s} + \cdots,
\end{aligned}$$

up to a finite error term.

If we add the four up, all the primes not of the form $5k+1$ cancel out. Then we obtain

$$\log L(s, \chi_1) + \log L(s, \chi_2) + \log L(s, \chi_3) + \log L(s, \chi_4) = 4 \sum_{p=5k+1} \frac{1}{p^s} + (\text{error}).$$

So once we divide both sides by $|\log(s-1)|$ and take the limit $s \to 1$, we get

$$0 + 0 + 0 + 1 = 4 \cdot \frac{\sum_{p=5k+1} \frac{1}{p^s}}{|\log(s-1)|} + 0.$$

So we get that the Dirichlet density of the primes $\{p = 5k+1\}$ is exactly $\frac{1}{4}$.

Now we can play the same trick. This time, we add χ_2 and χ_4 and then from this subtract χ_1 and χ_3. Then this time, we are left with

$$-\log L(s,\chi_1) + \log L(s,\chi_2) - \log L(s,\chi_3) + \log L(s,\chi_4) = 4 \sum_{p=5k+4} \frac{1}{p^s} + (\text{error})$$

and then again dividing by $|\log(s-1)|$ and letting $s \to 1$ gives that the primes $\{p = 5k+4\}$ has Dirichlet density $\frac{1}{4}$. To distinguish the $5k+2$ and $5k+3$ primes, we can look at

$$\log L(s,\chi_4) - i \log L(s,\chi_1) - \log L(s,\chi_2) + i \log L(s,\chi_3) = 4 \sum_{p=5k+2} \frac{1}{p^s} + (\text{error}),$$

$$\log L(s,\chi_4) + i \log L(s,\chi_1) - \log L(s,\chi_2) - i \log L(s,\chi_3) = 4 \sum_{p=5k+3} \frac{1}{p^s} + (\text{error}).$$

So this proves that all primes have Dirichlet density $\frac{1}{4}$.

Proposition 25.1. *For each $b = 1, 2, 3, 4$, we have*

$$d(\{\textit{primes of the form } p = 5k+b\}) = \frac{1}{4}.$$

For general q, we can look at characters χ modulo q. Ghese are given by

$$\chi_c(g) = e^{2\pi i c/(p-1)}, \quad \chi_c(g^k) = e^{2\pi i ck/(p-1)}.$$

Now we can look at the L-functions $L(s,\chi_c)$ and take right linear combinations of these L-functions. For instance, you will get

$$\sum_{c=0}^{q-2} \log L(s,\chi_c) = (q-1) \sum_{p=qk+1} \frac{1}{p^s} + (\text{error}).$$

There is a slick algebraic way to do these things, but I won't go into the details. It is not too much different from the $q = 5$ case.

Index