M3P14 Number Theory

Lectured by Prof Toby Gee
Typeset by David Kurniadi Angdinata

Autumn 2018

Contents

Lecture 1
Friday
05/10/18

0 Introduction

Roughly speaking number theory is the study of the integers. More specifically, problems in number theory often have a lot to do with primes and divisibility, congruences, and include problems about the rational numbers. For example, solving equations in integers or in the rationals, such as $x^2 - 2y^2 = 1$, etc. We will be looking at problems that can be tackled by elementary means, but this does not mean easy. Also the statements of problems can be elementary without the solution being elementary, such as Fermat's Last Theorem, or even known, such as the twin prime conjecture. Sometimes we will state interesting things, like the prime number theorem, without proving them. Typically these will be things that we could prove if the course was much longer. We will start the course with a look at prime numbers and factorisation, a review of Euclid's algorithm and consequences, congruences, the structure of $(\mathbb{Z}/n\mathbb{Z})^\times$, RSA algorithm, and quadratic reciprocity. We will return to primes at the end, too. Typical questions here include the following.

1. How do you tell if a number is prime?
2. How many primes are there congruent to a modulo b for given a, b?
3. How many primes are there less than n?

A warning is that we will be using plenty of things from previous algebra courses, about groups, rings, ideals, fields, Lagrange's theorem, the first isomorphism theorem, and so on. You may want to revise this material if you are not comfortable with it. The course is not based on any particular book, although some material, such as continued fractions, was drawn from the following.

1. A Baker, A concise introduction to the theory of numbers, 1984

Not everything we will do is in that book, though.

1 Euclid's algorithm and unique factorisation

1.1 Divisibility

Definition 1. Let $a, b \in \mathbb{Z}$. We say that a **divides** b, written $a \mid b$, if there exists $c \in \mathbb{Z}$ such that $b = ac$. If a does not divide b, write $a \nmid b$.

Note. If $a, b, c \in \mathbb{Z}$ such that $a \mid b$ and $a \mid c$, then $a \mid rb + sc$ for any $r, s \in \mathbb{Z}$.

Definition 2. Let $a, b \in \mathbb{Z}$, not both zero. The **greatest common divisor** (GCD) or **highest common factor** (HCF) of a and b, written (a, b), is the largest positive integer dividing both a and b.

Such an integer always exists since if $a \neq 0$ and $c \mid a$, then $-a \leq c \leq a$.

Example. $(-10, 15) = 5$.

Note. This notation is consistent with notation from ring theory. The ring $\mathbb{Z}$ is a principal ideal domain (PID), that is it is an integral domain, and every ideal can be generated by one element. The ideal generated by $f_1, \dots, f_n \in R$ for some ring R is usually written $(f_1, \dots, f_n)$, and indeed the ideal (a, b) is generated by the highest common factor of a and b, by Theorem 6 below.

Definition 3. $n \in \mathbb{Z}$ is **prime** if n has exactly two positive divisors, namely 1 and n.

Note. By definition, primes can be both positive and negative. In spite of this, frequently when people talk about prime numbers they restrict to the positive case. In this course when we say let p be a prime number we will generally mean $p > 0$. Also 1 is not prime.

1.2 Euclid's algorithm

Proposition 4. Let $a, b \in \mathbb{Z}$, not both zero. Then for any $n \in \mathbb{Z}$, we have $(a, b) = (a, b - na)$.

Proof. By definition of (a, b), it suffices to show that any $r \in \mathbb{Z}$ divides both a and b if and only if it divides both a and $b - na$. But if r divides a and b, it clearly divides $b - na$, and if it divides a and $b - na$, it clearly divides b. $\square$

This suggests an approach to computing (a, b) by replacing (a, b) by a pair $(a, b - na)$, and repeat until the numbers involved are small enough that it is easy to compute the greatest common divisor. The key to being able to do this is the following innocuous looking result.

Theorem 5. Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist unique $q, r \in \mathbb{Z}$ such that $a = qb + r$ and $0 \leq r < b$.

Proof. Let $q = \lfloor a/b \rfloor$ be the largest integer less than a/b. Then by definition $0 \leq a/b - q < 1$. Thus $0 \leq a - qb < b$, so we can take $r = a - bq$. Uniqueness is easy. $\square$

This gives us **Euclid's algorithm** for finding (a, b) for any $a, b \in \mathbb{Z}$ not both zero. Without loss of generality, assume $0 \leq b \leq a$ and $a > 0$.

1. Check if $b = 0$. If so then $(a, b) = a$.
2. Otherwise, replace (a, b) with (b, r) as in Theorem 5. Then return to step 1.

Since at every stage $|a| + |b|$ is decreasing, this algorithm terminates. We have shown that $(a, b) = (b, r)$ so the output is always equal to (a, b).

Example. Let us make this explicit.

$$\begin{aligned} (120, 87) &= (87, 33) & 120 &= 87 + 33 \\ &= (33, 21) & 87 &= 2(33) + 21 \\ &= (21, 12) & 33 &= 21 + 12 \\ &= (12, 9) & 21 &= 12 + 9 \\ &= (9, 3) & 12 &= 9 + 3 \\ &= (3, 0) & 9 &= 3(3) + 10 \end{aligned}$$

Now run this backwards, writing out the equations.

$$\begin{aligned} 3 &= 12 - 9 \\ &= 12 - (21 - 12) \\ &= 2(12) - 21 \\ &= 2(33 - 21) - 21 \\ &= 2(33) - 3(21) \\ &= 2(33) - 3(87 - 2(33)) \\ &= 8(33) - 3(87) \\ &= 8(120 - 87) - 3(87) \\ &= 8(120) - 11(87). \end{aligned}$$

The same works in general, that is the algorithm gives us more than just a way to compute (a, b). It also allows us to express (a, b) in terms of a and b.

Theorem 6. Let $a, b \in \mathbb{Z}$, not both zero. Then there exist $r, s \in \mathbb{Z}$ such that $(a, b) = ra + sb$.

Proof. Let $a_0 = a$ and $b_0 = b$, and for each i let (a_i, b_i) be the result after running i steps of Euclid's algorithm on the pair (a, b). For some r we have $a_r = (a, b)$ and $b_r = 0$. We will show, by downwards induction on i, that there exist $n_i, m_i \in \mathbb{Z}$ such that $(a, b) = n_i a_i + m_i b_i$. For $i = r$ this is clear. On the other hand, for any i we have $a_i = b_{i-1}$ and $b_i = a_{i-1} - q_i b_{i-1}$ for some $q_i \in \mathbb{Z}$. Thus if $(a, b) = n_i a_i + m_i b_i$, we have

$$(a, b) = n_i b_{i-1} + m_i (a_{i-1} - q_i b_{i-1}) = (n_i - m_i q_i) b_{i-1} + m_i a_{i-1},$$

and the claim follows. $\square$

1.3 Unique factorisation

The fact that (a, b) is an integer linear combination of a and b has strong consequences for factorisation and divisibility. First note the following.

Proposition 7. Let $n, a, b \in \mathbb{Z}$, and suppose that $n \mid ab$ and $(n, a) = 1$. Then $n \mid b$.

Proof. Since $(n, a) = 1$, there exist $r, s \in \mathbb{Z}$ such that $rn + sa = 1$. Thus $rnb + sab = b$. But n clearly divides rnb and sab, so $n \mid b$. □

By definition, if n is prime, then either $n \mid a$ or $(n, a) = 1$. If $(n, a) = 1$, we say that n, a are **coprime**. Lecture 2 Tuesday 09/10/18

Corollary 8. If p is prime, and $a, b \in \mathbb{Z}$ are such that $p \mid ab$, then either $p \mid a$ or $p \mid b$.

Proof. If $p \nmid a$ then $(p, a) = 1$, so 7 implies $p \mid b$. □

Proposition 9. If $(a, b) = 1$, and $a \mid n$ and $b \mid n$, then $ab \mid n$.

Proof. By 6, we can write $n = n\,(a, b) = nra + nsb$ with $r, s \in \mathbb{Z}$. Each term is divisible by ab, so $ab \mid n$. □

We say that $m_1, \ldots, m_n \in \mathbb{Z}$ are **pairwise coprime** if $(m_i, m_j) = 1$ for all $i \neq j$.

Corollary 10. Suppose that $m_1, \ldots, m_n$ are pairwise coprime. If $m_i \mid N$ for all i, then $m_1 \ldots m_n \mid N$.

Proof. Induction on n. $n = 2$ is Proposition 9. (TODO Exercise) □

We can now prove the existence and uniqueness of prime factorisations.

Proposition 11. Every $n \in \mathbb{Z}^{\times}$ can be written as $\pm p_1 \ldots p_r$ for some $r \geq 0$ and some primes $p_1, \ldots, p_r$.

Proof. Use induction on $|n|$. The case $|n|$ is trivial, so suppose $|n| > 1$. Then either $|n|$ is prime, or $|n| = ab$ with $1 < a, b < |n|$, and by induction each of a, b is a product of primes. □

Theorem 12. Let $n \in \mathbb{Z}_{>0}$. Then n can be written as $p_1 \ldots p_r$ where the p_i are prime, and are uniquely determined up to reordering.

Proof. Existence is Proposition 11. For uniqueness, suppose that

$$n = p_1 \ldots p_r = q_1 \ldots q_s,$$

with p_i, q_i prime. Then without loss of generality suppose $r, s \geq 1$. Then $p_1 \mid p_1 \ldots p_r$, so $p_1 \mid q_1 \ldots q_s$. By Corollary 8, either $p_1 \mid q_1$ or $p_1 \mid q_2 \ldots q_s$. Proceeding inductively, eventually $p_1 \mid q_i$ for some i. Since q_i is prime this means $p_1 = q_i$. We then have

$$p_2 \ldots p_r = q_1 \ldots q_i \ldots q_s.$$

Since this product is smaller than n, by the inductive hypothesis we must have $r - 1 = s - 1$ and the p_i except p_1 are a rearrangement of the q_i except q_i. □

Put together, these are the fundamental theorem of arithmetic.

1.4 Linear diophantine equations

Suppose now that we are given $a, b, c \in \mathbb{Z}^{\times}$ and we want to solve $ax + by = c$ for $x, y \in \mathbb{Z}$.

Note. (a, b) divides both a and b, so for there to be any solutions, we must have $(a, b) \mid c$.

Example. $2x + 6y = 3$ has no solutions.

From now on, suppose this is true. Let $a' = a/(a,b)$, $b' = b/(a,b)$, and $c' = c/(a,b)$. Then $ax + by = c$ if and only if $a'x + b'y = c'$. By Theorem 6, since $(a', b') = 1$, we can find $r, s \in \mathbb{Z}$ with $a'r + b's = 1$, so $a'rc' + b'sc' = c'$. So $x = rc'$, $y = sc'$ is a solution. X, Y is another solution if and only if $a'X + b'Y = a'x + b'y$, if and only if $a'(X - x) = b'(y - Y)$. For this to hold, we need $a' \mid y - Y$, $b' \mid X - x$. Putting this all together, we find that if x, y is one solution to $ax + by = c$, then the other solutions are exactly of the form

$$X = x + n\frac{b}{(a,b)}, \qquad Y = y - n\frac{a}{(a,b)}$$

for all $n \in \mathbb{Z}$.

Example. Using the example above where we have $8(120) - 11(87) = 3$, we can solve $120x + 87y = 9$. One solution is $x = 24$ and $y = -33$. The general solution is $x = 24 + 29n$ and $y = -33 - 40n$. Taking $n = -1$, we have for example, $x = -5$ and $y = 7$.

2 Congruences and modular arithmetic

2.1 Congruences

Definition 13. Let $n \in \mathbb{Z}^\times$, and let $a, b \in \mathbb{Z}$. We say a is **congruent to** b **modulo** n, written $a \equiv b \mod n$, if $n \mid a - b$.

For n fixed, it is easy to verify that congruence modulo n is an equivalence relation, and therefore partitions $\mathbb{Z}$ into equivalence classes. The set of equivalence classes modulo n is denoted $\mathbb{Z}/n\mathbb{Z}$.

Example. If $a \equiv b \mod n$, $c \equiv d \mod n$, then $a + c \equiv b + d \mod n$ and $ac \equiv bd \mod n$.

In fact $\mathbb{Z}/n\mathbb{Z}$ is a ring, with the obvious addition and multiplication. Indeed $n\mathbb{Z} = \{nr \mid r \in \mathbb{Z}\}$ is an ideal in $\mathbb{Z}$, and $\mathbb{Z}/n\mathbb{Z}$ is just the quotient ring. For any $a \in \mathbb{Z}$, we sometimes write $\overline{a}$ for the image of a in $\mathbb{Z}/n\mathbb{Z}$. We can write $a = qn + r$ with $0 \le r < n$. Then $a \equiv r \mod n$, so $\overline{a} = \overline{r}$.

Example. If $n = 12$, then $\overline{25} = \overline{1}$.

It follows that $0, \dots, n-1$ are representatives for the elements of $\mathbb{Z}/n\mathbb{Z}$, so every element of $\mathbb{Z}/n\mathbb{Z}$ is equal to $\overline{r}$ for some unique $r \in \{0, \dots, n-1\}$. It will also be convenient to write $\mathbb{Z}/n\mathbb{Z} = \{0, \dots, n-1\}$.

Example. If $n = 6$, we could write $3 + 4 = 1$ and $3 \times 4 = 0$.

Recall that if R is a commutative ring, a **unit** of R is an element with a multiplicative inverse, that is x such that there exists $y \in R$ with $xy = 1$. Write $R^\times$ for the set of units in R. This is a group under multiplication.

Example. $\mathbb{Z}^\times = \{\pm 1\}$ and $\mathbb{Q}^\times = \mathbb{Q} \setminus \{0\} = \{x \in \mathbb{Q} \mid x \ne 0\}$.

We want to understand $(\mathbb{Z}/n\mathbb{Z})^\times$. Which elements of $\{0, \dots, n-1\}$ are in $(\mathbb{Z}/n\mathbb{Z})^\times$? If $r \in \mathbb{Z}$ and $\overline{r} \in (\mathbb{Z}/n\mathbb{Z})^\times$ then there exists $s \in \mathbb{Z}$ such that $rs \equiv 1 \mod n$. This implies that $(r, n) = 1$. Conversely, if $(r, n) = 1$, then there exist $x, y \in \mathbb{Z}$ such that $rx + ny = 1$, so $\overline{rx} = 1$, so $\overline{r}$ is a unit. Thus we have $(\mathbb{Z}/n\mathbb{Z})^\times = \{\bar{i} \mid (i, n) = 1\}$.

Note. If p is a prime, then either $a \equiv 0 \mod p$ or $(a, p) = 1$, so $(\mathbb{Z}/p\mathbb{Z})^\times = \{1, \dots, p-1\}$. Thus every non-zero congruence class modulo p is a unit, that is $\mathbb{Z}/p\mathbb{Z}$ is a ring with the property that every non-zero element has a multiplicative inverse, so it is a field. Another equivalent way to see this is to check that $p\mathbb{Z}$ is a maximal ideal of $\mathbb{Z}$.

Lecture 3
Wednesday
10/10/18

2.2 Linear congruence equations

Fix $a, b \in \mathbb{Z}$ and $c \in \mathbb{Z}^\times$. Suppose we want to solve $ax \equiv b \mod c$. This is equivalent to finding x, y such that $ax + cy = b$. In particular, by our analysis of linear diophantine equations, there is a solution precisely when $(a, c) \mid b$. Furthermore, there is a unique solution modulo $c' = c/(a, c)$, because all the solutions are obtained by adding multiples of c' to our given x, and subtracting the corresponding multiple of $a/(a, c)$ from y. This implies that there are a total of (a, c) solutions to the original congruence modulo c. If x is a solution, the other solutions are of the form $X = x + c'j$ for $0 \leq j < (a, c)$. In particular, if $(a, c) = 1$, then there is a unique solution to $ax \equiv b \mod c$. Indeed $a \in (\mathbb{Z}/c\mathbb{Z})^\times$, so it has an inverse a^{-1}, and $x \equiv a^{-1}b \mod c$ is the unique solution.

Example.

1. $2x \equiv 3 \mod 6$ has no solutions as $(2, 6) = 2 \nmid 3$.
2. $2x \equiv 4 \mod 6$, which is equivalent to $x \equiv 2 \mod 3$, has solutions $x \equiv 2 \mod 6$ and $x \equiv 5 \mod 6$.

2.3 Chinese remainder theorem

Theorem 14 (Chinese remainder theorem)**.** Let $m_1, \dots, m_n \in \mathbb{Z}_{\geq 0}$ be pairwise coprime. Then the natural map

$$\frac{\mathbb{Z}}{m_1 \dots m_n \mathbb{Z}} \to \frac{\mathbb{Z}}{m_1 \mathbb{Z}} \times \dots \times \frac{\mathbb{Z}}{m_n \mathbb{Z}}$$

is an isomorphism of rings, and the induced map

$$\left(\frac{\mathbb{Z}}{m_1 \dots m_n \mathbb{Z}}\right)^\times \to \left(\frac{\mathbb{Z}}{m_1 \mathbb{Z}}\right)^\times \times \dots \times \left(\frac{\mathbb{Z}}{m_n \mathbb{Z}}\right)^\times$$

is an isomorphism of abelian groups.

Note. This is false without the assumption that m_i pairwise coprime, for example $m_1 = m_2 = 2$.

Proof. Note firstly that the map exists and is a ring homomorphism. This follows from the fact that if $x \equiv y \mod m_1 \dots m_n$ then certainly $x \equiv y \mod m_i$ for each i. The source and target of the ring homomorphism both have order $m_1 \dots m_n$, so it suffices to show that the map is injective to show that it is an isomorphism. So we only need to check that the kernel is zero. So we need to know that if $m_i \mid N$ for all i, that is $\overline{N} = 0$ in $\mathbb{Z}/m_i\mathbb{Z}$, then $m_1 \dots m_n \mid N$, that is $\overline{N} = 0$ in $\mathbb{Z}/m_1 \dots m_n\mathbb{Z}$. This is just Corollary 10. The statement about unit groups follows by noting that if R, S are rings, then $(R \times S)^\times = R^\times \times S^\times$. $\square$

Note. This can be reformulated more concretely as a statement about congruences. It says that for any a_i, there is a unique $x \mod m_1 \dots m_n$ such that $x \equiv a_i \mod m_i$. The proof does not tell us how to find x, but it is actually quite easy in practice. Here is one way to do it. Write $M = m_1 \dots m_n$ and $M_i = M/m_i$. Choose q_i such that $q_i M_i \equiv 1 \mod m_i$, using Euclid's algorithm and $(M_i, m_i) = 1$ because $(m_j, m_i) = 1$ for all $j \neq i$. Then set

$$x = a_1 q_1 M_1 + \dots + a_n q_n M_n.$$

For each i we have $M_j \equiv 0 \mod m_i$ if $i \neq j$, so $x \equiv a_i q_i M_i \equiv a_i \mod m_i$ for each i.

3 The structure of $(\mathbb{Z}/n\mathbb{Z})^\times$

For the next few lecture we will study the abelian group $(\mathbb{Z}/n\mathbb{Z})^\times$.

3.1 The Euler Φ function

We define a function $\Phi(n)$ on $\mathbb{Z}_{>0}$ by letting $\Phi(n)$ denote the order of $(\mathbb{Z}/n\mathbb{Z})^{\times}$. Explicitly we have $\Phi(n) = \#\{1 \leq i < n \mid (i,n) = 1\}$, that is $\Phi(n)$ is the number of integers between 0 and $n-1$ coprime to n.

Example. If p is prime, $\Phi(p) = p-1$.

Φ is called **Euler's Φ function**.

Definition 15. A function f on $\mathbb{Z}_{>0}$ is **multiplicative** if for all $m,n \in \mathbb{Z}$ such that $(m,n) = 1$, we have $f(mn) = f(m)f(n)$. We say f is **strongly multiplicative** if for any pair of $m,n \in \mathbb{Z}_{>0}$ we have $f(mn) = f(m)f(n)$.

Note. By the Chinese remainder theorem, Φ is multiplicative, because if $(m,n) = 1$ then $(\mathbb{Z}/mn\mathbb{Z})^{\times} \cong (\mathbb{Z}/m\mathbb{Z})^{\times} \times (\mathbb{Z}/n\mathbb{Z})^{\times}$, but not strongly multiplicative, since $\Phi(4) = 2 \neq 1 = \Phi(2)\Phi(2)$.

It is clear that a multiplicative function is determined by its values on prime powers. For p prime we have $(i,p^a) = 1$ if and only if p does not divide i, so $\Phi(p^a)$ is the number of integers between 0 and $p^a - 1$ that are not divisible by p. There are p^{a-1} numbers in this range divisible by p, so we have

$$\Phi(p^a) = \#\{1 \leq i < p^a \mid (i,p^a) = 1\} = \#\{1 \leq i < p^a \mid p \nmid i\} = p^a - p^{a-1} = p^a\left(1 - \frac{1}{p}\right).$$

Write $n = \prod_i p_i^{a_i}$ where p_i are distinct primes. From this and multiplicativity of Φ one has that

$$\Phi(n) = \prod_i \Phi\left(p_i^{a_i}\right) = \prod_i p_i^{a_i}\left(1 - \frac{1}{p_i}\right) = n\prod_i\left(1 - \frac{1}{p_i}\right) = n\prod_{p \mid n}\left(1 - \frac{1}{p}\right),$$

where p runs over the primes dividing n.

3.2 Euler's theorem

The units $(\mathbb{Z}/n\mathbb{Z})^{\times}$ form a group under multiplication. By definition, $\phi(n)$ is the order of this group. Recall that for any group G of finite order d, Lagrange's theorem states that for all $g \in G$, g^d is the identity in G. For the group $(\mathbb{Z}/n\mathbb{Z})^{\times}$, this means the following.

Theorem 16 (Euler's theorem)**.** Let $a \in \mathbb{Z}$ with $(a,n) = 1$. Then $a^{\Phi(n)} \equiv 1 \mod n$.

Proof. This is equivalent to saying that $\overline{a}^{\Phi(n)} = 1$ in $(\mathbb{Z}/n\mathbb{Z})^{\times}$. This is a group of order $\Phi(n)$, so this is immediate from Lagrange's theorem. □

Corollary 17 (Fermat's little theorem)**.** If p is a prime and $p \nmid a$ then $a^{p-1} \equiv 1 \mod p$.

Proof. Theorem 16 with $n = p$, so $\Phi(n) = p - 1$. □

Of course knowing the order of an abelian group does not tell you its structure.

Example. Let $n = 5$. $(\mathbb{Z}/5\mathbb{Z})^{\times} = \{1,2,3,4\}$. This has order 4. There are two isomorphism classes of abelian groups of order 4, namely $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$. So it is either cyclic of order 4 or a product of two cyclic groups of order 2. $2^2 = 4$, $2^3 = 3$, $2^4 = 1$ in $(\mathbb{Z}/5\mathbb{Z})^{\times}$. So $(\mathbb{Z}/5\mathbb{Z})^{\times}$ is cyclic of order 4.

By the Chinese remainder theorem, to understand the structure of $(\mathbb{Z}/n\mathbb{Z})^{\times}$, it is enough to understand the structure of $\mathbb{Z}/p^m\mathbb{Z}$ where p is prime and $m \geq 1$. We will do this next, beginning with the case $m = 1$.

Lecture 4 Friday 12/10/18

Definition 18. If G is a group and $g \in G$ is an element, the **order** of g is the least $a \geq 1$ such that $g^a = 1$. In particular, if $(g,n) = 1$, then we write $ord_n(g)$ for the order of g in $(\mathbb{Z}/n\mathbb{Z})^{\times}$, or the order of g modulo n.

Equivalently, let $n \in \mathbb{Z}_{>0}$ and $g \in \mathbb{Z}$ with $(g,n) = 1$, then the order of g modulo n is the smallest $a \in \mathbb{Z}_{\geq 0}$ such that $g^a \equiv 1 \mod n$.

Proposition 19. If G is a group and g is an element of order a, then $g^n = 1$ if and only if $a \mid n$.

Equivalently, let $g \in \mathbb{Z}$ with $(g, n) = 1$, then if $g^n \equiv \mod n$ then $ord_n(g) \mid n$.

Proof. If $n = ab$ then $g^n = (g^a)^b = 1^b = 1$. Conversely write $n = ab + r$ with $b, r \in \mathbb{Z}$ and $0 \leq r < a$. Then since $g^a = 1$ it follows that $g^r = 1$. Since $r < a$, r cannot be positive by the definition by order, so $r = 0$ and $n = ab$. ☐

In particular, if $(g, n) = 1$, then $g^{\Phi(n)} = 1$ by Euler's theorem, so Proposition 19 gives the order of g modulo n divides $\Phi(n)$. We are going to prove that if p is prime, then $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic. Equivalently, we need to show that there exists g such that $ord_p(g) = \Phi(p) = p - 1$. We will do this by counting the number of elements of each order. Key point is that $\mathbb{Z}/p\mathbb{Z}$ is a field. For any $d \geq 1$, the elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ of order dividing d are exactly the roots of the equation $X^d - 1$ in $\mathbb{Z}/p\mathbb{Z}$ by Proposition 19.

Example. The equation $X^2 = 1$ has exactly two solutions modulo p for any prime p, namely ± 1, but it can have more modulo n if n is composite. If $n = 15$, then $4, 11$ are also solutions. $X^2 - 1 \equiv 0 \mod n$ if and only if $n \mid (X + 1)(X - 1)$, so $15 \mid (4 + 1)(4 - 1)$.

Definition 20. $g \in \mathbb{Z}$ with $(g, p) = 1$ is a **primitive root modulo** p if the order of g modulo p is exactly $p - 1$, equivalently, if $\overline{g}$ is a generator of $(\mathbb{Z}/p\mathbb{Z})^\times$.

To prove that primitive roots exist, we require some results about roots of polynomials modulo p. Over the rational numbers we all know that a polynomial of degree d has at most d roots. This can fail over other rings.

Example. The polynomial $x^2 - x$ has the roots $0, 1, 3, 4$ modulo 6. The issue here is that $\mathbb{Z}/6\mathbb{Z}$ is not a field.

Lemma 21. Let R be a commutative ring, and let $P(X)$ be a polynomial in X with coefficients in R. If $\alpha \in R$ has $P(\alpha) = 0$, then there exists a polynomial $Q(X)$ with coefficients in R such that $P(X) = (X - \alpha) Q(X)$.

Example. If $R = \mathbb{Z}/15\mathbb{Z}$, $X^2 - 1 = (X + 1)(X - 1) = (X + 4)(X - 4)$.

Proof. We proceed by induction on the degree of P, the degree zero case being clear. Suppose the result is true for polynomials of degree less than $d - 1$, and let $P(X)$ have degree d. If the leading term of $P(X)$ is cX^d, so $P(X) = cX^d + \dots$, let $S(X) = P(X) - cX^{d-1}(X - \alpha)$. We have $S(\alpha) = 0$, and $S(X)$ has degree less than $d - 1$. By induction, there exists $R(X)$ with coefficients in R such that we can write $S(X) = (X - \alpha) R(X)$. Set $Q(X) = cX^{d-1} + R(X)$. Then

$$(X - \alpha) Q(X) = (X - \alpha)\left(cX^{d-1} + R(X)\right) = cX^{d-1}(X - \alpha) + S(X) = P(X).$$

☐

Theorem 22. Let F be a field, and $P(X)$ a polynomial of degree d with coefficients in F. Then $P(X)$ has at most d distinct roots in F.

Proof. We again proceed by induction on $d = \deg(P)$. The case $d = 0$ is clear. If P has no roots, then we are done. Otherwise, $P(X)$ has degree d and let α be a root. By Lemma 21, we can write $P(X) = (X - \alpha) Q(X)$. Now if $P(\beta) = 0$, then $(\beta - \alpha) Q(\beta) = 0$, so since F is a field either $\beta = \alpha$ or β is a root of $Q(X)$. By the inductive hypothesis $Q(X)$ has degree $d - 1$, so P has at most d roots and we are done by induction. ☐

As a corollary, we deduce the following.

Corollary 23. Let p be a prime, and let d be any divisor of $p - 1$. Then there are exactly d elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ of order dividing d.

Equivalently, we have to show that the polynomial $X^d - 1$ has exactly d roots modulo p.

Proof. Note that by Fermat's little theorem, $1, \dots, p-1$ are all roots of $x^{p-1}-1$ modulo p. Thus $X^{p-1}-1$ has exactly $p-1$ roots. Now fix d dividing $p-1$ and write

$$X^{p-1}-1 = \left(X^d-1\right)\left(\left(X^d\right)^{\frac{p-1}{d}-1}+\dots+1\right) = \left(X^d-1\right)Q\left(X\right), \qquad \deg\left(Q\right) = p-1-d,$$

for a polynomial $Q(X)$. $Q(X)$ has integer coefficients so we can view it as a polynomial modulo p. Now $X^{p-1}-1$ has exactly $p-1$ roots, X^d-1 has at most d roots, and $Q(X)$ has at most $p-1-d$ roots by Theorem 22. We must therefore have equality in these inequalities, that is X^d-1 has exactly d roots modulo p. ☐

Another way of stating the corollary is to say that for any d dividing $p-1$, there are exactly d elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ whose order divides d.

Example. Let $p=7$. Then $(\mathbb{Z}/p\mathbb{Z})^\times$ has

1. 1 element of order 1,
2. 2 elements of order dividing 2, so 1 element of order 2,
3. 3 elements of order dividing 3, so 2 elements of order 3, and
4. 6 elements of order dividing 6, so 2 elements of order 6.

Lemma 24. For any $n \ge 1$, we have $\sum_{d \mid n,\ d>0} \Phi(d) = n$.

Proof. For each $d \mid n$, the elements $i \in \{1, \dots, n\}$ with $(i,n) = n/d$ are precisely those of the form $i = (n/d)\,j$ with $1 \le j \le d$ and $(j,d) = 1$. There are exactly $\Phi(d)$ possibilities for j, so there are exactly $\Phi(d)$ such elements. Summing over all d, since the n/d run over all the divisors of n, we are done with the result. ☐

Theorem 25. Let p be a prime. Then for any d dividing $p-1$, there are exactly $\Phi(d)$ elements of order d in $(\mathbb{Z}/p\mathbb{Z})^\times$. In particular there are $\Phi(p-1)$ primitive roots modulo p, and $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic.

Proof. We prove this by strong induction on d. The case $d=1$ is clear. Fix d. The inductive hypothesis tells us that for any d' dividing d and strictly less than d there are $\Phi(d')$ elements of exact order d'. On the other hand by Corollary 23 there are a total of d elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ of order dividing d. Thus the number of elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ of order exactly d is

$$\Phi(d) = d - \sum_{d' \mid d,\ d' \ne d} \Phi(d').$$

precisely by Lemma 24. Now use inductive hypothesis. ☐

Lecture 5 Tuesday 16/10/18

We can now go on to the case of $(\mathbb{Z}/p^n\mathbb{Z})^\times$ for $n \ge 2$. Firstly we do the case $p>2$.

Proposition 26. Let p be an odd prime and let $n \ge 1$. Then $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic.

Proof. Consider three cases.

$n=1$ Theorem 25.

$n=2$ Let g be a primitive root modulo p. Claim that either $g^{p-1} \not\equiv 1 \mod p^2$, and g is a generator for $\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times$, or $g^{p-1} \equiv 1 \mod p^2$, and $g+p$ is a generator for $\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times$. Either way, $\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times$ is cyclic. Suppose firstly that $g^{p-1} \not\equiv 1 \mod p^2$. $g^{ord_{p^2}(g)} \equiv 1 \mod p^2$ gives $g^{ord_{p^2}(g)} \equiv 1 \mod p$, so we have by assumption

$$p-1 = ord_p(g) \mid ord_{p^2}(g) \mid \#\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times = \Phi\left(p^2\right) = p\left(p-1\right).$$

But $ord_{p^2}(g) \ne p-1$, as $g^{p-1} \not\equiv 1 \mod p^2$. So $ord_{p^2}(g) = p(p-1)$ as required. Now suppose that $g^{p-1} \equiv 1 \mod p^2$, and set $h = g+p$. It suffices to show that $h^{p-1} \not\equiv 1 \mod p^2$, as we can then apply the analysis above with h in place of g to show that $ord_{p^2}(h) = p(p-1)$ and $\left(\mathbb{Z}/p^2\mathbb{Z}\right)^\times$ is cyclic. To see the claim, observe that if we expand with the binomial theorem, then we get

$$h^{p-1} = (g+p)^{p-1} \equiv g^{p-1} + (p-1) p g^{p-2} \equiv 1 + p(p-1) g^{p-2} \mod p^2,$$

and since $p \nmid (p-1) g^{p-2}$, $(g+p)^{p-1} \not\equiv 1 \mod p^2$, as required.

$n \geq 2$ We claim that if $ord_{p^2}(g) = p(p-1)$ then in fact $ord_{p^n}(g) = p^{n-1}(p-1)$ for all $n \geq 2$, so that in particular $(\mathbb{Z}/p^n\mathbb{Z})^\times$ is cyclic. We do this by induction on n. So assume that $ord_{p^n}(g) = p^{n-1}(p-1)$. Then

$$p^{n-1}(p-1) = ord_{p^n}(g) \mid ord_{p^{n+1}}(g) \mid \Phi\left(p^{n+1}\right) = p^n(p-1).$$

So either $ord_{p^{n+1}}(g) = p^n(p-1)$, or $ord_{p^{n+1}}(g) = p^{n-1}(p-1)$. The statement that $ord_{p^{n+1}}(g) = p^n(p-1)$ is equivalent to showing that $g^{p^{n-1}(p-1)} \not\equiv 1 \mod p^{n+1}$. To do this, consider $g^{p^{n-2}(p-1)} \mod p^{n-1}$ and $g^{p^{n-2}(p-1)} \mod p^n$. Since $\Phi\left(p^{n-1}\right) = p^{n-2}(p-1)$, $g^{p^{n-2}(p-1)} \equiv 1 \mod p^{n-1}$ by Euler's theorem, so we may write

$$g^{p^{n-2}(p-1)} = 1 + p^{n-1}t.$$

Since $ord_{p^n}(g) = p^{n-1}(p-1)$ by assumption, $g^{p^{n-2}(p-1)} \not\equiv 1 \mod p^n$, that is $p \nmid t$. Then the binomial theorem shows that

$$g^{p^{n-1}(p-1)} = \left(g^{p^{n-2}(p-1)}\right)^p = \left(1 + p^{n-1}t\right)^p \equiv 1 + p^n t + \binom{p}{2} p^{2(n-1)} t^2 + \cdots + p^{p(n-1)} t^p \mod p^{n+1},$$

Now $r(n-1) \geq n+1$ if and only if $(r-1)n \geq r+1$. Since $p > 2$,

$$p \;\Big|\; \binom{p}{2} \qquad \implies \qquad p^{n+1} \;\Big|\; p^{2n-1} = p^{2(n-1)+1} \;\Big|\; \binom{p}{2} p^{2(n-1)}.$$

So $g^{p^{n-1}(p-1)} \equiv 1 + p^n t \not\equiv 1 \mod p^{n+1}$, because $p \nmid t$. So the statement holds for $n+1$, and we are done by induction

∎

Note. We used the hypothesis that $p \neq 2$ right at the end here. If $p = 2$ then we cannot ignore the higher order terms.

If $n = 1, 2$ then the proof of Proposition 26 did not use $p > 2$, and indeed

1. $(\mathbb{Z}/2\mathbb{Z})^\times = \{1\}$ is cyclic,
2. $(\mathbb{Z}/4\mathbb{Z})^\times = \{1, 3\}$ is cyclic of order 2, with 3 as a generator, but
3. this fails for higher powers, say $(\mathbb{Z}/8\mathbb{Z})^\times = \{1, 3, 5, 7\}$ is not cyclic since $1^2 \equiv 3^2 \equiv 5^2 \equiv 7^2 \equiv 1 \mod 8$, so every element has order two.

The key is the following lemma.

Lemma 27. For $n \geq 0$ we have $5^{2^n} \equiv 1 + 2^{n+2} \mod 2^{n+3}$.

Proof. Induction on n. The case $n = 0$ follows from $5 = 1 + 4$. Suppose that $5^{2^n} = 1 + 2^{n+2}t$ with t odd. Then

$$5^{2^{n+1}} = \left(1 + 2^{n+1}t\right)^2 = 1 + 2^{n+3}t + 2^{2(n+2)}t^2 = 1 + 2^{n+3}\left(t + 2^{n+1}t^2\right),$$

and since $n+1 \geq 1$ and $t + 2^{n+1}t^2$ is odd we are done by induction. ∎

Proposition 28. If $n \geq 2$ then we have an isomorphism $(\mathbb{Z}/2^n\mathbb{Z})^\times \to (\mathbb{Z}/2\mathbb{Z}) \times \left(\mathbb{Z}/2^{n-2}\mathbb{Z}\right)$, so that in particular if $n \geq 3$ then $(\mathbb{Z}/2^n\mathbb{Z})^\times$ is not cyclic.

Proof. Consider the natural map

$$\langle -1\rangle \times \langle 5\rangle \to \left(\frac{\mathbb{Z}}{2^n\mathbb{Z}}\right)^\times ,$$

where if G is a group and $g \in G$ we write $\langle g\rangle$ for the cyclic subgroup $\left\{1,\dots,g^{ord(g)-1}\right\}$ of G generated by g. We claim that this map is an isomorphism. To see this, note that it is injective, because if $(-1)^r(5)^s \equiv 1 \mod 2^n$ then in particular $(-1)^r(5)^s \equiv 1 \mod 4$ so $(-1)^r \equiv 1 \mod 4$, so we must have $r = 1$ and $5^s \equiv 1 \mod 2^n$, that is $5^s = 1$ in $\langle 5\rangle$. $\langle -1\rangle \cong \mathbb{Z}/2\mathbb{Z}$ has order 2 and $\langle 5\rangle \cong \mathbb{Z}/2^{n-2}\mathbb{Z}$ has order $ord_{2^n}(5) = 2^{n-2}$ by Lemma 27. So $\langle -1\rangle \times \langle 5\rangle$ has order $2\left(2^{n-2}\right) = 2^{n-1} = \Phi(2^n) = \#\left(\mathbb{Z}/2^n\mathbb{Z}\right)^\times$. So the map $\langle -1\rangle \times \langle 5\rangle \to (\mathbb{Z}/2^n\mathbb{Z})^\times$ is an injection of groups of the same order, so it is a bijection. □

Using what we have shown so far, one can conclude the following. See the first example sheet.

Theorem 29. Let $n \in \mathbb{Z}_{>0}$. The group $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic if and only if either

1. $n = 1, 2, 4$,
2. $n = p^r$ where p is an odd prime and $r \geq 1$, or
3. $n = 2p^r$ where p is an odd prime and $r \geq 1$.

Lecture 6 Wednesday 17/10/18

Note that while the existence of primitive roots is very useful both theoretically and computationally, there is no simple procedure for finding them in practice, beyond trial and error by guessing small values of g, and see if g is a generator. If p is prime then there are $\Phi(p-1)$ primitive roots modulo p, so they are plentiful, which means that you have a high probability of success, and so trying $2, 3, 5, 6, \dots$ is a reasonable strategy, but note that trying 4 would not be a good idea. We could work out $1, \dots, g^{p-2}$ and check these are distinct, but this would be inefficient. Better is to check that if q is any prime factor of $p-1$, then $g^{(p-1)/q} \neq 1$. This works, because if g is not a primitive root, then the order of g modulo p is a proper factor of $p-1$, so divides some $(p-1)/q$, so $g^{(p-1)/q} = 1$, while if $g^{(p-1)/q} \neq 1$ then $ord_p(g) \mid p-1$ and $ord_p(g) \nmid (p-1)/q$. If this holds for all $q \mid p-1$, then $ord_p(g) = p-1$, because otherwise it would be a proper divisor, and so would divide $(p-1)/q$ for some prime $q \mid p-1$.

Note. This does rely on being able to factor $p-1$, which is a hard problem in general. See the next section.

The work of computing powers of a^k modulo p can be done efficiently by repeated squaring followed by multiplication according to the binary expansion of k.

Example. Let us find a primitive root modulo $p = 31$. $p - 1 = 30 = (2)(3)(5)$. g is a primitive root if and only if $g^{15} \neq 1$, $g^{10} \neq 1$, $g^6 \neq 1$. It is easy to see that 2 does not work because $2^2 = 4$, $2^4 = 16$, $2^6 = 2$, but $2^10 = 2^15 = 1$ because $2^5 = 32 = 1$, so 2 is not a primitive root. We claim that 3 is a primitive root. We need to show that none of 3^6, 3^{10}, 3^{15} are 1.

$$3^2 = 9, \quad 3^4 = -12, \quad 3^8 = 20 \quad \implies \quad 3^6 = 9(-12) = 16, \quad 3^{10} = 9(20) = 25, \quad 3^{15} = 3(25)(-12) = -1.$$

So 3 is a primitive root modulo 31, as required.

4 Primality testing and factorisation

The basic idea of this section, which will be exploited in the next brief section on cryptography, is that checking whether $n \in \mathbb{Z}$ is prime or not is easy, but factorising n is expected to be hard, even if we know that it is not prime. Actually, there is no proof that factorisation is hard, merely an expectation. We will not make the notions of easy and hard precise, but the difficulty should be measured in terms of $\log n$, that is in terms of the number of digits of n in some base. Easy here means that there is an algorithm to check whether n is prime or not which runs in time polynomial in $\log n$. Then it is known that there exists a deterministic algorithm, the Agrawal-Kayal-Saxena (AKS) algorithm in 2005, to check whether or not n is prime which runs in time which is polynomial in $\log n$. We will not describe this algorithm, although it is fairly elementary. We will discuss another algorithm which is more effective than this in practice. In contrast it is unknown whether or not there is an algorithm for factorising n that runs in time polynomial in $\log n$, but it is suspected that no such algorithm should exist. There are algorithms better than exponential in $\log n$, but nothing close to polynomial time.

4.1 Factorisation

One way to check if n is prime or not is to try dividing by all primes up to $\sqrt{n}$, because if d is a proper divisor of n, then either $d \leq \sqrt{n}$ or $n/d \leq \sqrt{n}$. This is fine if you are factoring a small integer in your head, but hopeless if you want to factor numbers which are hundreds of digits long on a computer.

Note. If you want to check primality or factorise relatively small numbers, there are tricks. In particular, if you want to check divisibility by 2, 3, 5, 7, or 11 there are the following tests.

1. Checking by divisibility by 2 or 5 is easy, just a matter of looking at the last decimal digit.
2. For 3 and 11, we have $10 \equiv 1 \mod 3$ and $10 \equiv -1 \mod 11$, so
$$\sum_{i=0}^{\log n} a_i 10^i \equiv \sum_{i=0}^{\log n} a_i \mod 3, \qquad \sum_{i=0}^{\log n} a_i 10^i \equiv \sum_{i=0}^{\log n} a_i \left(-1\right)^i \mod 11,$$
so we can check divisibility by taking the sum of the decimal digits for 3 or 9, and taking the alternating sum for 11.
3. For 7 things are slightly more awkward, but there is the following observation. $10x + y \equiv 0 \mod 7$ if and only if $-2\left(10x + y\right) \equiv 0 \mod 7$ if and only if $x - 2y \equiv 0 \mod 7$. So we can repeatedly subtract off twice the last digit from the number formed by removing the last digit.

If we wanted to factor three digit numbers, or small four digit numbers, say $n \leq 400$ is composite, with paper or calculator, then n has a prime factor $d \leq \sqrt{400} = 20$. Then we only have to worry about checking divisibility by primes up to 19. If n is not divisible by $2, 3, 5, 7, 11$, then the smallest prime factor of n is at least 13. So with these tests we only have difficulties if the only prime factors are 13, 17, 19, where there are no good tests. Since $13^3 > 400$, it can have at most 2 prime factors. If you can recognise the squares 169, 289, 361, then you only have to remember a short list $13 \times 17 = 221$, $13 \times 19 = 247$, $13 \times 23 = 299$, $13 \times 29 = 377$, $17 \times 19 = 323$, $17 \times 23 = 391$.

Example.

1. $143 \equiv 1 - 4 + 3 \equiv 0 \mod 11$.
2. $144 \equiv 1 + 4 + 4 \equiv 0 \mod 9$.
3. $154 \equiv 15 - 2\left(4\right) = 7 \equiv 0 \mod 7$.

Lecture 7 Friday 19/10/18

In fact, there is a method due to Fermat which allows you to factor even four digit numbers by hand, if you really have to. Idea is to first eliminate small prime factors by hand, up to say $p = 2, \dots, 19$. If n is composite and does not have any small factors, the remaining possibility is that if n has prime factors, they are close together to $\sqrt{n}$, as in the exceptional cases we considered in the last paragraph. Now, if $n = ab$ with $a \leq b$ both odd, then we can write
$$n = ab = \left(\frac{a+b}{2}\right)^2 - \left(\frac{b-a}{2}\right)^2, \qquad \implies \qquad \left(\frac{a+b}{2}\right)^2 - n = \left(\frac{b-a}{2}\right)^2.$$
If you know $\left(a+b\right)/2$ and $\left(b-a\right)/2$, you can recover a, b. This suggests the following procedure. If n is a square, we are done. If not, let m be the least integer with $m^2 \leq n < \left(m+1\right)^2$, and check if $\left(m+1\right)^2 - n$ is a square. If it is not, try $\left(m+2\right)^2 - n$, and so on. Once you can write $y^2 - n = x^2$ then $n = y^2 - x^2 = \left(y+x\right)\left(y-x\right)$ and we have a factorisation.

Example. If $n = 6077$, then you find that $77^2 < 6077 < 78^2$, and
$$\begin{aligned} 78^2 - 6077 &= 7, \\ 79^2 - 6077 &= 164, \\ 80^2 - 6077 &= 323, \\ 81^2 - 6077 &= 484 = 22^2, \end{aligned}$$
so $6077 = 81^2 - 22^2 = 103 \times 59$.

In the worst case even the combination of trial division and the Fermat method run in time which is exponential in $\log n$. The fastest algorithms known for factoring n run in better than exponential time in $\log n$ are subexponential. They are the quadratic sieve, Lenstra elliptic curve factorisation, and the general number field sieve. They are all significantly more complicated, although at least the **quadratic sieve** could be described in this course if we wanted to. Rather than go through it in detail, we just give a sense of the idea using an example.

Example. Imagine trying to factor $n = 1649$ using the Fermat method. Since $40^2 < 1649 < 41^2$, we compute

$$\begin{aligned} 41^2 - 1649 &= 32 = 2^5, \\ 42^2 - 1649 &= 115 = 5 \times 23, \\ 43^2 - 1649 &= 200 = 2^3 \times 5^2. \end{aligned}$$

We certainly have not factored it yet, as none of these is a square, and indeed we would have to do another fourteen steps to find a factor. However, $32 \times 200 = 2^8 \times 5^2 = 80^2$. This means that we have a congruence

$$\left(2^5 \times 2^3 \times 5^2\right)^2 = (41 \times 43)^2 \equiv 80^2 \mod 1649.$$

Since $41 \times 43 = 1763 \equiv 114 \mod 1649$, this means that $1649 \mid (114 + 80)(114 - 80) = 194 \times 34 = 2^2 \times 17 \times 97$ and indeed $1649 = 17 \times 97$.

This is the basic idea of the quadratic sieve. In order to factor n, rather than trying to find numbers a, b with $a^2 - b^2 = n$, you try to find them with $a^2 \equiv b^2 \mod n$. Then you can hope that one of $a \pm b$ has a common factor with n.

Note. For illustration we factored $a \pm b$ above, but in general a better idea for the last step would be to compute the GCD $(a \pm b, n)$, $(194, 1649) = 97$ and $(34, 1649) = 17$, quickly using Euclid's algorithm.

How do we find congruences like this? To make this into an efficient algorithm, the idea is that it is easy to spot relations like the one we found if the numbers have only small prime factors. In fact, we can turn it into a linear algebra problem over the field with two elements $\mathbb{Z}/2\mathbb{Z}$, in the following way. Suppose that we have a set $x_1, \dots, x_r \in \mathbb{Z}$ and we want to find a product of a subset of them which is a square. If we know the prime factorisation for the x_i, we can write $x_i = p_1^{a_{i1}} \dots p_k^{a_{ik}}$, then we are trying to find $\epsilon_i \in \{0, 1\}$ such that $\prod_{i=1}^r x_i^{\epsilon_i}$ is a square. Equivalently, for each $1 \le j \le k$, want the exponent of p_j to be even, that is we have $\sum_{i=1}^r \epsilon_i a_{ij} = \epsilon_1 a_{1j} + \dots + \epsilon_r a_{rj} \equiv 0 \mod 2$. This is just a linear algebra question.

Example. Let $x_1 = 2^5$, $x_2 = 5 \times 23$, $x_3 = 2^3 \times 5^2$. If we just look at the numbers above which only had small primes less than or equal to 5 in their factorisations, then we are looking at x_1 and x_3. Taking $p_1 = 2$ and $p_2 = 5$, we need to solve

$$\begin{pmatrix} \epsilon_1 & \epsilon_2 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 3 & 2 \end{pmatrix} \equiv \begin{pmatrix} 0 & 0 \end{pmatrix} \mod 2 \qquad \iff \qquad \begin{pmatrix} \epsilon_1 & \epsilon_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

in the field $\mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$, that is $\epsilon_1 + \epsilon_2 = 0$, which has the non-trivial solution $\epsilon_1 = \epsilon_2 = 1$.

This step, solving linear equations in $\mathbb{Z}/2\mathbb{Z}$, can be done efficiently. In order to make this into a practical algorithm to factor n, the remaining difficulty is to find a good supply of $m \in \mathbb{Z}$ such that $m^2 - n$ is smooth, in the sense that it only has small prime factors.

Note. This is something that you could try to do by trial division. Once we had decided above that we only wanted prime factors up to 5, we could just keep dividing $a^2 - n$ by 2, 3, 5, and if we did not get to 1 then we could throw the number away. In practice, there are faster ways to proceed, which is where the sieve in quadratic sieve comes from.

The basic idea is if we fix a list of small primes to start with, we use congruence conditions on m, because for each prime $2 < p \nmid n$, there will be zero or two possible values for m in $m^2 \equiv n \mod p$. It turns out that there is a straightforward algorithm for solving $m^2 \equiv n \mod p$, which is part of the theory of quadratic residues, which we will get to shortly. If you do this for lots of primes p, you get a supply of congruence conditions for m, so you can eliminate ever considering m such that $m^2 - n$ has large prime factors.

Example. It is easy to check that $m^2 \equiv 1649 \mod 3$ has no solutions, so there would have been no point in looking for 3, and if we want 5 to be a factor of $m^2 - 1649$ then we need $x \equiv \pm 2 \mod 5$, and so on.

Lecture 8 Tuesday 23/10/18

4.2 Testing primality

By Euler's theorem if $(a, n) = 1$ then $a^{\Phi(n)} \equiv 1 \mod n$. In particular, if p is prime then $a^{p-1} \equiv 1 \mod p$ for all $1 \le a \le p-1$. Conversely, if we can find an $1 \le a < n$ with $a^{n-1} \not\equiv 1 \mod n$, then n is not prime. Even just taking $a = 2$ in $2^{n-1} \not\equiv 1 \mod n$ is often enough to show that n is not prime, and using repeated squaring this can be checked quickly.

Example. To check if 9 is prime, we can compute $2^8 \equiv 2^{2^{2^2}} \equiv 4^{2^2} \equiv 7^2 \equiv 4 \mod 9$.

However it does not always work. $341 = 11 \times 31$ has $2^{340} \equiv 1 \mod 341$. In general even varying a is not enough. A **Carmichael number** is a composite number such that for all $(a, n) = 1$ then $a^{n-1} \equiv 1 \mod n$. It is known that infinitely many of them exist, a hard theorem, although they are rare. See the example sheet for an example of a few of them. A variant on this idea gives an efficient primality test, the **Miller-Rabin test**, a test for whether $n \in \mathbb{Z}$ is prime or not. We restrict ourselves to considering the case that $n \equiv 3 \mod 4$, as the essential idea is already clear in this case, but the analysis is more complicated in the case $n \equiv 1 \mod 4$, in an example sheet. Of course if n is even we do not need a primality test. The key point is the following.

Lemma 30. If $n \equiv 3 \mod 4$, then n is prime if and only if for every $a \in (\mathbb{Z}/n\mathbb{Z})^\times$, we have $a^{(n-1)/2} \equiv \pm 1 \mod n$.

Proof.

1. Suppose firstly that n is prime. Then $a^{n-1} \equiv 1 \mod n$ by Fermat's little theorem, so $\left(a^{(n-1)/2}\right)^2 \equiv 1 \mod n$, so $a^{(n-1)/2} \equiv \pm 1 \mod n$, because n is prime, the equation $x^2 = 1$ has just the roots ± 1 in $\mathbb{Z}/n\mathbb{Z}$.

2. Suppose next that $n = p^k$ for p prime is a prime power with $k \ge 2$, and take $a = 1 + p$. Then
$$(1+p)^{\frac{n-1}{2}} \equiv 1 + \binom{n-1}{2} p \mod p^2,$$
by the binomial theorem. If $(1+p)^{(n-1)/2} \equiv \pm 1 \mod p^k = n$, then $(1+p)^{(n-1)/2} \equiv \pm 1 \mod p, p^2$ gives
$$\pm 1 \equiv (1+p)^{\frac{n-1}{2}} \equiv 1 + \binom{n-1}{2} p \equiv 1 \mod p \quad \implies \quad 1 \equiv (1+p)^{\frac{n-1}{2}} \equiv 1 + \binom{n-1}{2} p \mod p^2,$$
then $p \mid (n-1)/2$, so $p \mid n-1$. But $p \mid n$, a contradiction.

3. Finally for the remaining case suppose that n is composite but not a power of a prime, and write $n = rs$ for $r, s > 1$ and odd, and $(r, s) = 1$. By the Chinese remainder theorem, $\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}/r\mathbb{Z} \times \mathbb{Z}/s\mathbb{Z}$. We can choose a with $(a, n) = 1$ such that $a \equiv -1 \mod r$ and $a \equiv 1 \mod s$. Then $(a, r) = (a, s) = 1$, so $(a, n) = 1$. Since $n \equiv 3 \mod 4$ by assumption, $(n-1)/2$ is odd, so $a^{(n-1)/2} \equiv -1 \mod r$ and $a^{(n-1)/2} \equiv 1 \mod s$. So $a^{(n-1)/2} \not\equiv \pm 1 \mod n$.

$\square$

So we know that if n is composite, there will exist values of a with $a^{(n-1)/2} \not\equiv \pm 1 \mod n$. To understand how efficient the algorithm is, we need to know how many such a there are.

Lemma 31. Suppose that $n \equiv 3 \mod 4$ and that n is composite. Then the set of $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ satisfying $a^{(n-1)/2} \equiv \pm 1 \mod n$ is a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$.

Proof. That it is a subgroup follows easily from the definition. Just check that it is closed under products and inverses. Certainly $1^{(n-1)/2} \equiv 1 \mod n$. If $a^{(n-1)/2} \equiv \pm 1 \mod n$ and $b^{(n-1)/2} \equiv \pm 1 \mod n$,

$$(ab)^{\frac{n-1}{2}} \equiv a^{\frac{n-1}{2}} b^{\frac{n-1}{2}} \equiv (\pm 1)(\pm 1) \equiv \pm 1, \quad \left(a^{-1}\right)^{\frac{n-1}{2}} \equiv \left(a^{\frac{n-1}{2}}\right)^{-1} \equiv (\pm 1)^{-1} \equiv \pm 1 \mod n.$$

So this set is a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$. That it is a proper subgroup is then immediate from Lemma 30. □

We immediately get the following bound.

Corollary 32. Suppose that $n \equiv 3 \mod 4$ and that n is composite. Then at most half of the values of $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ satisfy $a^{(n-1)/2} \equiv \pm 1 \mod n$.

Proof. The set of such elements is a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$ by Lemma 31, and any proper subgroup has index at least two. □

In fact we can do better than this. It can be shown with some more work that you can improve this to show that at least $3/4$ of the integers $1 \leq a \leq n-1$ satisfy $a^{(n-1)/2} \not\equiv \pm 1 \mod n$. It follows that if $n \equiv 3 \mod 4$ and you choose x integers $1 \leq a \leq n-1$ at random, and n is composite, then the probability that you find such an a with $a^{(n-1)/2} \not\equiv \pm 1 \mod n$ is at least $1 - 1/4^x$, so this gives an efficient polynomial time probabilistic random algorithm for testing if n is prime. If you want it to be deterministic, if you assume the generalised Riemann hypothesis (GRH), then it is known that you can find a counterexample $a \in \mathbb{Z}$ with

$$1 \leq a \leq \left\lceil 2 (\log n)^2 \right\rceil, \qquad a^{\frac{n-1}{2}} \not\equiv \pm 1 \mod n,$$

and again we have a polynomial time algorithm. In practice it is even better than this.

Example. If $n < 341550071728321$, it is enough to test $a = 2, 3, 5, 7, 11, 13, 17$.

5 Public-key cryptography

5.1 Messages as sequences of classes modulo n

How do we turn messages into numbers in $\mathbb{Z}/n\mathbb{Z}$? Since the advent of computers, the idea of representing a message by a string of numbers is a familiar one. In practice, to do this one typically chooses a way of encoding individual characters as binary numbers of a fixed length d, usually eight or sixteen bits, that is binary digits. If we then cut a message up into blocks or strings of at most k characters and concatenate the binary representations of each character in the block together, we obtain a dk bit binary number that represents an k character block as an integer between 0 and 2^{dk}. If we choose some very large modulus $n > 2^{dk}$, then we can alternatively represent a block as a class in $\mathbb{Z}/n\mathbb{Z}$. Thus we will be mainly concerned with the problem of communicating a congruence class c modulo n, for some large n, between a sender A and a recipient B. The goal is to do this in such a way that any eavesdroppers on the communication cannot deduce what c is, but B can.

5.2 The Rivest-Shamir-Adleman (RSA) algorithm

Most traditional forms of cryptography rely on a shared secret known to both A and B. This shared secret is effectively some invertible function $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$. The idea is that rather than sending c to B directly, A applies f to c and computes $f(c)$, sends that to B, and then B applies some other function $g : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ and computes $g(f(c))$, where $g = f^{-1}$, to get back the number A started with. Since eavesdroppers do not know f, they, at least in principle, cannot recover c from $f(c)$. In practice, for A and B to agree on a function f poses problems. In particular, they have to communicate to do so, and if eavesdroppers listen to that communication they can learn f. Want to be able to make f public without making g public. The algorithm we describe today avoids this problem completely. It is what is known as a public-key algorithm. Instead of secrets being shared between A and B, our recipient B creates a secret known only to B, his **private key**, and then releases additional information, his **public key**, to anyone who wants to communicate with him. For anyone to send B a message, only the public key is required, but decoding the message requires the private key. Here is how the algorithm works.

1. B first chooses two large prime numbers p and q and sets $n = pq$. In practice, each of these is around 2^{1024} or so. An integer modulo n thus allows B to represent 2048 bits of information, or 256 eight bit characters.
2. B also chooses a number e such that $(e, \Phi(n)) = 1$, and lets d be a multiplicative inverse of e modulo $\Phi(n)$ such that $de = 1 \mod \Phi(n) = (p-1)(q-1) = n - (p+q) + 1$.
3. The public key, that B publishes and shares with everyone, consists of the numbers n and e.
4. The private key, that B must keep secret, consists of the numbers p, q, $\Phi(n)$, and d.
5. To encode a message x, a sender A computes $f(x) \equiv x^e \mod n$, and sends it to B.
6. Given an encoded message y, B decodes it by computing $g(y) \equiv y^d \mod n$.

The reason this works is that if $y \equiv x^e \mod n$, then one has $(x^e)^d \equiv x^{ed} \equiv x^{1+k\Phi(n)} \mod n$, since, by construction, $de \equiv 1 \mod \Phi(n)$. Thus $y^d \equiv x^{ed} \equiv x^1 \equiv x \mod n$ by Euler's theorem, $x^{\Phi(n)} \equiv 1 \mod n$.

Note. This works provided x is coprime to n, but the probability of this is extremely high. It is still ok even without that, since n is squarefree. (TODO Exercise using Fermat's little theorem plus Chinese remainder theorem)

The prevailing assumption is that with only the information n and e, it is hopeless to discover d. Any eavesdropper who knows x^e and wants to recover x from x^e then has to be able to compute an e^{th} root of x modulo n. As far as we know, this is quite difficult computationally. The best publicly known approaches all involve factoring n. For numbers around 2^{2048}, this is not feasible with today's computing equipment, and might well never be feasible. On the other hand, we have no formal proof that factoring is as computationally difficult as it seems to be. As far as I am aware, we do not even have a formal proof that breaking RSA is as computationally difficult as factoring. In spite of these uncertainties, our intuition and experience suggests that recovering x from x^e without knowing a factorisation of n is computationally infeasible. It is this infeasibility that allows the cryptosystem to work.

Lecture 9 Wednesday 24/10/18

Lecture 9 is a problem class.

Lecture 10 Friday 26/10/18

5.3 Signing with RSA

Public-key cryptography can also be used as verification of identity. Suppose B wants to make a declaration to the world, and prove beyond all doubt that it was B who made the declaration, and not an impostor. Perhaps this declaration is a will, or acceptance of a contract, for instance. Suppose B has functions $f, g : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ with $f \circ g = g \circ f = id$. Again, make f public, and any time B publishes a message m, B also publishes $g(m)$. Then anyone can apply f to $g(m)$ to recover $m = f(g(m))$, but without g, no one can forge B's signature. With RSA, B first represents the message he wants to sign as a class m modulo n. To sign this class, B computes m^d modulo n using the private part of the key, and sends the world the pair (m, m^d). Suppose A wants to verify that a pair (m, s) was a message signed by B. Then A computes s^e modulo n, which requires only the public part of the key. If $s \equiv m^d \mod n$, then $s^e = m^{de} \equiv m \mod n$. So A just needs to check that $s^e \equiv m \mod n$ and if so the signature is verified. To fake a message signed by B, a forger needs to solve the problem of, given a message m, finding a signature s such that $s^e \equiv m \mod n$. This is precisely the same problem as deciphering a message sent by the algorithm above. Thus forging signatures is just as hard as breaking the encryption.

5.4 Discrete logarithms

If $(\mathbb{Z}/n\mathbb{Z})^\times$ is cyclic, for example, if n is prime, and g is a generator for this group, that is a primitive root, then the map $\mathbb{Z}/\Phi(n)\mathbb{Z} \to (\mathbb{Z}/n\mathbb{Z})^\times$ taking a modulo $\Phi(n)$ to g^a is an isomorphism, from the additive group of $\mathbb{Z}/\Phi(n)\mathbb{Z}$ to $(\mathbb{Z}/n\mathbb{Z})^\times$. It thus has an inverse, which we call the **discrete logarithm to the base** g. Explicitly, if g is a primitive root modulo n, then the discrete logarithm to the base g, denoted $\log_g$, is defined for any $a \in (\mathbb{Z}/n\mathbb{Z})^\times$ by $\log_g(a) \equiv m \mod \Phi(n)$, for some unique $m \in \mathbb{Z}$ such that $0 \le m < \Phi(n)$ and $g^m \equiv a \mod n$.

Example. One use of the discrete logarithm is to solve exponential equations modulo n. By applying $\log_g$ to both sides of the equation $x^r \equiv a \mod n$, write $x = g^y$, we obtain that the congruence becomes equivalent to the linear congruence equation $yr \equiv \log_g(a) \mod \Phi(n)$, and we can solve those with techniques explained earlier.

Unfortunately, or fortunately for cryptography, it is expected that the discrete logarithm is hard to compute, much in the way that it is expected to be hard to crack RSA, but we do not know for sure. In particular, there is no known polynomial time algorithm.

Example. Here is a practical application of this. Imagine that you have a system where you need to safely store passwords for different users, but you do not want to store the actual passwords. One way to do this is to let p be a large prime, big enough so that all passwords can be thought of as residues modulo $(p-1)$, and fix a primitive root g modulo p. Then if someone inputs their password to be x, you can compute and store g^x modulo p. If they later want to login with input y, you compute g^y, and check if it matches what you stored. If it does then $y \equiv x \mod p-1$. Even if someone has access to what you have stored, and to g, they still cannot recover the password a without solving the discrete logarithm problem. Of course, nor can you, so it is not so good if you require people to be able to be reminded of their passwords.

6 Quadratic reciprocity

6.1 Quadratic residues

Definition 33. Let p be a prime number and $a \in \mathbb{Z}$ not divisible by p, that is $(a, p) = 1$. We say that a is a **quadratic residue modulo** p (QR) if and only if there exists a solution $x \in \mathbb{Z}$ to $x^2 \equiv a \mod p$. If no such d exists, so a is not a QR, it is called a **quadratic non-residue modulo** p (QNR).

Note. By this convention, $a \in \mathbb{Z}$ divisible by p are neither QRs nor QNRs modulo p. Other conventions exist, so sometimes zero is a QR.

Example.

1. If $p = 2$, 1 is a QR.
2. If $p = 3$, 1 is a QR, -1 is a QNR, since $1^2 \equiv (-1)^2 \equiv 1 \mod 3$.
3. If $p = 5$, $1, 4$ are QRs, $2, 3$ are QNRs, since $1^2 \equiv (-1)^2 \equiv 1 \mod 5$ and $2^2 \equiv 3^2 \equiv 4 \mod 5$.

Lemma 34. If $p > 2$ then there are exactly $(p-1)/2$ QRs modulo p, and $(p-1)/2$ QNRs modulo p.

Proof. The QRs are exactly the image of the group homomorphism $(\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times$ by $x \mapsto x^2$. This has kernel $x = \pm 1$, so the image has order $(p-1)/2$. $\square$

Proposition 35. Let $a, b \in \mathbb{Z}$ with $(a, p) = (b, p) = 1$. Then

1. if a and b are both QRs modulo p, then so is ab,
2. if a is a QR modulo p and b is a QNR modulo p, then ab is a QNR modulo p, and
3. if a and b are both QNRs modulo p, then ab is a QR modulo p.

Proof. Note that the set H of QRs in $(\mathbb{Z}/p\mathbb{Z})^\times$ is a subgroup, because it is the image of the group homomorphism $(\mathbb{Z}/p\mathbb{Z})^\times \to (\mathbb{Z}/p\mathbb{Z})^\times$ by $x \mapsto x^2$ by Lemma 34, and by the first isomorphism theorem we have $(\mathbb{Z}/p\mathbb{Z})^\times / H \cong \mathbb{Z}/2\mathbb{Z}$. The proposition is a restatement of this, since $(\mathbb{Z}/p\mathbb{Z})^\times = H \sqcup 1 + H$. $\square$

Definition 36. The **Legendre symbol** $\left(\frac{a}{p}\right)$, for p a prime and $a \in \mathbb{Z}$, is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & a \text{ is a QR modulo } p \\ 0 & p \mid a \\ -1 & a \text{ is a QNR modulo } p \end{cases}.$$

Proposition 35 above then amounts to saying that the map $(\mathbb{Z}/p\mathbb{Z})^\times \to \{\pm 1\}$ defined by $a \mapsto \left(\frac{a}{p}\right)$ is a group homomorphism, that is $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{a}{p}\right)$. Even holds if we do not assume that $(a,p) = (b,p) = 1$. In fact, the existence of primitive roots gives us an easy description of the following map. Lecture 11 Tuesday 30/10/18

Theorem 37 (Euler's criterion)**.** Let p be an odd prime, and $a \in \mathbb{Z}$ not divisible by p. Then

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \mod p.$$

Proof. Let g be a primitive root modulo p, and write $a \equiv g^r \mod p$ for $0 \le r < p-1$. Then $\left(g^{(p-1)/2}\right)^2 = g^{p-1} \equiv 1 \mod p$. So $g^{(p-1)/2} \equiv \pm 1 \mod p$. Since g is a primitive root, $g^{(p-1)/2} \not\equiv 1 \mod p$, so $g^{(p-1)/2} \equiv -1 \mod p$. So $a^{(p-1)/2} \equiv (g^r)^{(p-1)/2} \equiv \left(g^{(p-1)/2}\right)^r = (-1)^2 \mod p$. But

$$\begin{aligned} \left(\frac{a}{p}\right) = 1 \quad &\iff \quad (g^s)^2 \equiv a \mod p \quad &&\iff \quad 2s \equiv r \mod p-1 \\ &\iff \quad 2 \mid r \quad &&\iff \quad (-1)^r \equiv 1 \mod p, \end{aligned}$$

and we are done. □

6.2 Computing Legendre symbols

Euler's criterion lets us determine, for fixed p, which a are QRs modulo p. What if we fix a, and ask for which odd primes p is a a QR? When $a = -1$, Euler's criterion gives an easy answer. $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$, so -1 is a QR modulo p if and only if $(p-1)/2$ is even. In other words, the following holds.

Proposition 38. $-1 \in \mathbb{Z}$ is a QR modulo p if and only if $p = 2$ or $p \equiv 1 \mod 4$.

Proof. $p = 2$ is trivial. If $p > 2$, then by Euler's criterion, $\left(\frac{-1}{p}\right) \equiv (-1)^{(p-1)/2} \mod p$, so in fact $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$. Then

$$(-1)^{\frac{p-1}{2}} = \begin{cases} 1 & p \equiv 1 \mod 4 \\ -1 & p \equiv 3 \mod 4 \end{cases}.$$

□

Example. If $p = 5$ then it is a square, and if $p = 7$ it is not, and in each case we can check directly.

When $a = 2$, the situation is more difficult, but still amenable to a direct approach.

Proposition 39 (A special case of Gauss' Lemma)**.**

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

that is $\left(\frac{2}{p}\right) = (-1)^{\left(p^2-1\right)/8}$.

Example. $\left(\frac{2}{7}\right) = 1$ since $2 \equiv 3^3 \mod 7$. $\left(\frac{2}{11}\right) = -1$ since squares modulo 11 are $1, 4, 9, 5, 3$. $\left(\frac{-1}{11}\right) = -1$, so $\left(\frac{-2}{11}\right) = \left(\frac{2}{11}\right)\left(\frac{-1}{11}\right) = (-1)^2 = 1$ and $-2 \equiv 3^2 \mod 11$.

Proof. $\left(\frac{2}{p}\right) \equiv 2^{(p-1)/2} \mod p$ by Euler's criterion. Let $q = (p-1)/2$, and set

$$Q = (2)(4)\dots(p-3)(p-1) = (2 \times 1)(2 \times 2)\dots(2 \times (q-1))(2 \times q) = 2^q q! = 2^{\frac{p-1}{2}} q!.$$

Reduce all the factors in the product defining Q modulo p so that they lie between $-q$ and q, that is subtract p from every factor which is greater than q. Let Q' be the resulting product $(2)(4)\dots(-3)(-1)$. We have $Q' \equiv Q \mod p$. On the other hand, the factors in the product defining Q' are the even integers from 1 to q

and the negatives of the odd integers from 1 to q. Thus $Q' = (-1)^r\, q!$, where r is the number of odd integers between 1 and q. We thus have $2^q q! \equiv (-1)^r\, q! \mod p$, and since $p \nmid q!$, we have $2^{(p-1)/2} \equiv (-1)^r \mod p$. The result follows by noting that

$$(-1)^r = \begin{cases} 1 & p \equiv \pm 1 \mod 8 \\ -1 & p \equiv \pm 3 \mod 8 \end{cases},$$

and invoking Euler's criterion. □

Example. If $p \equiv 1 \mod 8$, say $p = 1 + 8n$, so $q = 4n$. Odd integers in $1, \ldots, 4n$ are $1, \ldots, 4n-1$, so $r = 2n$.

Since we have $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$, to answer this question in full generality it suffices to answer it for $a = -1$, for $a = 2$, and for a an odd prime. In the latter case we have the following.

Theorem 40 (Law of quadratic reciprocity)**.** Let p and q be odd primes. Then

$$\left(\frac{p}{q}\right) = \begin{cases} -\left(\frac{q}{p}\right) & p \equiv q \equiv 3 \mod 4 \\ \left(\frac{q}{p}\right) & \text{otherwise} \end{cases}.$$

One can rephrase this a bit more tersely as the equivalent statement

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

Note that this implies that for each odd prime q, the question of whether q is a QR modulo p has an answer in terms of congruence conditions modulo q and modulo 4. From this and the Chinese remainder theorem, we can deduce that the question, for which primes p is a a QR modulo p, has an answer in terms of congruence conditions on p.

Example. If $p \neq 5$ is an odd prime, then we see that 5 is QR modulo p if and only if p is a QR modulo 5, so $\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right)$, that is if and only if $p \equiv \pm 1 \mod 5$. So

$$\left(\frac{5}{p}\right) = \begin{cases} 1 & p \equiv \pm 1 \mod 5 \\ -1 & p \equiv \pm 2 \mod 5 \end{cases}.$$

Example. Slightly more complicated example is the following.

1. Let $p \neq 3$ be an odd prime. When is 3 a QR modulo p, that is what is $\left(\frac{3}{p}\right)$? Well, if $p \equiv 1 \mod 4$ then this is if and only if p is a QR modulo 3, so if and only if $p \equiv 1 \mod 3$, so

$$\left(\frac{3}{p}\right) = \left(\frac{p}{3}\right) = \begin{cases} 1 & p \equiv 1 \mod 3 \\ -1 & p \equiv -1 \mod 3 \end{cases}.$$

If $p \equiv -1 \mod 4$ then if and only if $p \equiv -1 \mod 3$, so

$$\left(\frac{3}{p}\right) = -\left(\frac{p}{3}\right) = \begin{cases} 1 & p \equiv -1 \mod 3 \\ -1 & p \equiv 1 \mod 3 \end{cases}.$$

Putting this together with the Chinese remainder theorem,

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & p \equiv \pm 1 \mod 12 \\ -1 & p \equiv \pm 5 \mod 12 \end{cases}.$$

2. If $p = 7$, QRs are $1, 2, 4$, so $\left(\frac{3}{p}\right) = -1$. If $p = 11$, $5^2 \equiv 3 \mod 11$, so $\left(\frac{3}{p}\right) = 1$.

In general to compute $\left(\frac{a}{p}\right)$, we could do the following. Use that if $a \equiv b \mod p$ then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$. So without loss of generality $|a| < p$. Then write $a = \pm \prod_i q_i^{s_i}$ for q_i prime. Then $\left(\frac{a}{p}\right) = \left(\frac{\pm 1}{p}\right) \prod_i \left(\frac{q_i}{p}\right)^{s_i}$. If s_i is even, then $\left(\frac{q_i}{p}\right)^{s_i} = 1$. If s_i is odd, then $\left(\frac{q_i}{p}\right)^{s_i} = \left(\frac{q_i}{p}\right)$. We have formulas for $\left(\frac{-1}{p}\right)$ and $\left(\frac{2}{p}\right)$. If q is an odd prime, $q < p$, then use quadratic reciprocity to relate $\left(\frac{q}{p}\right)$ and $\left(\frac{p}{q}\right)$. Then repeat modulo q.

Example. $\left(\frac{6}{19}\right) = \left(\frac{2}{19}\right)\left(\frac{3}{19}\right) = (-1)(-1) = 1$. $\left(\frac{2}{19}\right) = -1$ because $19 \equiv 3 \mod 8$. $\left(\frac{3}{19}\right) \equiv -1 \mod 12$ by the above.

More generally, one can ask, given a monic polynomial f with integer coefficients, for which primes p does f have a root? The above case is the case of the polynomial $X^2 - a$. This is a very deep question in number theory. Indeed, we are still extremely far from having a complete answer. One question it is natural to ask is, for which f does the above question have an answer given in terms of congruence conditions on p. A deep branch of algebraic number theory called class field theory tells us that this will happen precisely when the field extension determined by f has abelian Galois group. Beyond this we know very little, but there are connections to the theory of modular forms. Lecture 12 Wednesday 31/10/18

6.3 Proof of quadratic reciprocity

Quadratic reciprocity was one of the deepest results of the 18^{th} century, and there are many approaches to proving it, none of which are particularly simple. The more motivated ones require algebraic number theory, and even then the motivation is really coming from class field theory, which is a long way beyond the boundaries of this course. The proof that we give is due to Rousseau, from 1991. It has the merits of being elementary and relatively easy to remember, and of resembling the proof of Gauss' Lemma that we gave above, that $\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8}$. Let p, q be distinct odd primes, and consider the group

$$\left(\frac{\mathbb{Z}}{pq\mathbb{Z}}\right)^\times \cong \left(\frac{\mathbb{Z}}{p\mathbb{Z}}\right)^\times \times \left(\frac{\mathbb{Z}}{q\mathbb{Z}}\right)^\times.$$

What are we going to do is to compare the products of different sets of coset representatives for the subgroup $\{\pm 1\}$. That is, we will look at different ways of choosing exactly one element of each pair $\{x, -x\}$ for each $x \in (\mathbb{Z}/pq\mathbb{Z})^\times$. We will always write everything as a pair $(\alpha, \beta) \in (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$. Firstly we recall from the first example sheet.

Theorem 41 (Wilson's theorem)**.** If p is prime then $(p-1)! \equiv -1 \mod p$.

Write $P = (p-1)/2$, $Q = (q-1)/2$, and $R = (pq-1)/2 = pQ + P$.

1. As our first set of coset representatives, consider the product of all the pairs

$$\left\{ (x, y) \in \left(\frac{\mathbb{Z}}{p\mathbb{Z}}\right)^\times \times \left(\frac{\mathbb{Z}}{q\mathbb{Z}}\right)^\times \;\middle|\; 1 \le x \le P,\ 1 \le y \le q-1 \right\}.$$

   Let A be the product of these coset representatives. Then the product of the y-coordinates is $(q-1)!^P \equiv (-1)^P \mod q$. The product of the x-coordinates is $P!^{q-1}$ in the same way, so

$$A = \prod_{1 \le x \le P,\ 1 \le y \le q-1} (x, y) = \left(P!^{q-1}, (-1)^P\right).$$

2. Similarly we let the second set of representatives be all the pairs

$$\left\{ (x, y) \in \left(\frac{\mathbb{Z}}{p\mathbb{Z}}\right)^\times \times \left(\frac{\mathbb{Z}}{q\mathbb{Z}}\right)^\times \;\middle|\; 1 \le x \le p-1,\ 1 \le y \le Q \right\}.$$

   Let B be the product of these representatives. In the same way by symmetry, we get

$$B = \prod_{1 \le x \le p-1,\ 1 \le y \le Q} (x, y) = \left((-1)^Q, Q!^{p-1}\right).$$

3. For the third set of representatives, select the pairs in $\mathbb{Z}/pq\mathbb{Z}$ which correspond via the Chinese remainder theorem to the set
$$\{1 \leq i \leq R \mid (i, pq) = 1\}.$$
Let C be the product of these coset representatives. Let us figure out the product of the x-coordinates. It is
$$\prod_{i=1,\ (i,pq)=1}^{R} i.$$
Since
$$\prod_{i=1,\ (i,pq)=1}^{R} i = \left(\prod_{i=1,\ (i,p)=1}^{R} i\right) \Bigg/ \left(\prod_{i=1,\ (i,p)=1,\ q\mid i}^{R} i\right), \tag{1}$$
$$\prod_{i=1,\ (i,p)=1}^{R} i = \left(\prod_{i=1,\ (i,p)=1}^{pQ} i\right)\left(\prod_{i=pQ+1,\ (i,p)=1}^{pQ+P} i\right), \tag{2}$$
$$\prod_{i=1,\ (i,p)=1,\ q\mid i}^{R} i = \prod_{j=1,\ (j,p)=1}^{P} qj = q^P P!, \tag{3}$$
combining (1), (2), (3), get that the x-coordinate of the product is
$$\prod_{i=1,\ (i,pq)=1}^{R} i = \frac{(p-1)!^Q P!}{q^P P!} = \frac{(-1)^Q}{q^P}.$$
So by symmetry we have the product of these representatives
$$C = \left(\frac{(-1)^Q}{q^P}, \frac{(-1)^P}{p^Q}\right) = \left((-1)^Q \left(\frac{q}{p}\right), (-1)^P \left(\frac{p}{q}\right)\right),$$
by Euler's criterion.

Now, we can compare A, B, C. We know that they all agree up to sign, that is up to possibly multiplication by ± 1, that is up to multiplication by the pair $(-1,-1) \in (\mathbb{Z}/p\mathbb{Z})^\times \times (\mathbb{Z}/q\mathbb{Z})^\times$. Comparing the y-coordinates, we have $C = \left(\frac{p}{q}\right) A$ and comparing the x-coordinates, similarly $C = \left(\frac{q}{p}\right) B$. So
$$B = \left(\frac{q}{p}\right)\left(\frac{p}{q}\right) A.$$
But we can compare A and B more directly. To go between A and B we need to swap the signs of all the PQ elements of the form (x, y) with $1 \leq x \leq P$ and $Q \leq y \leq q-1$. So $B = (-1)^{PQ} A$. So $\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{PQ}$, that is $\left(\frac{q}{p}\right) = (-1)^{PQ}\left(\frac{p}{q}\right)$.

6.4 Jacobi symbols

While it is often quite straightforward to compute Legendre symbols using quadratic reciprocity, there is one serious difficulty, which is that in order to compute $\left(\frac{a}{p}\right)$, we need to factor a, and we do not know a polynomial time algorithm for factorisation. However it is possible to make computations for Legendre symbols in polynomial time, with the key being the following generalisation due to Jacobi.

Definition 42. Let $b \in \mathbb{Z}_{>0}$ be odd, and $a \in \mathbb{Z}$. Then the **Jacobi symbol** $\left(\frac{a}{b}\right)$ is defined to be $\prod_{i=1}^{s}\left(\frac{a}{p_i}\right)^{r_i}$, where $b = \prod_{i=1}^{s} p_i^{r_i}$ for p_i distinct primes is the prime factorisation of b.

Note. In the special case that b is prime this agrees with the Legendre symbol. Warning that it is no longer the case that $\left(\frac{a}{b}\right) = 1$ implies that a is a square modulo b. On the other hand, of course $\left(\frac{a}{b}\right) = -1$ implies that a is not a square modulo b.

Lecture 13 Friday 02/11/18

The key properties of the Jacobi symbol are deduced from those of the Legendre symbol in the following lemma.

Lemma 43.

1. We have $\left(\frac{a_1}{b}\right)\left(\frac{a_2}{b}\right) = \left(\frac{a_1 a_2}{b}\right)$ and $\left(\frac{a}{b_1}\right)\left(\frac{a}{b_2}\right) = \left(\frac{a}{b_1 b_2}\right)$.
2. $\left(\frac{a}{b}\right)$ depends only on a modulo b.
3. $\left(\frac{a^2}{b}\right) = 1$.
4. $\left(\frac{-1}{b}\right) = (-1)^{(b-1)/2}$.
5. $\left(\frac{2}{b}\right) = (-1)^{\left(b^2-1\right)/8}$.
6. If $a, b > 0$ are both odd then $\left(\frac{a}{b}\right)\left(\frac{b}{a}\right) = (-1)^{((a-1)/2)((b-1)/2)}$.

Proof. All of these statements are true for Legendre symbols, that is for b prime, and a prime in 6. The first three parts are immediate from the definition and the corresponding results for the Legendre symbol. The same is true of the last three from 1 and the corresponding statements for Legendre symbols. We give the details for $\left(\frac{2}{b}\right) = (-1)^{\left(b^2-1\right)/8}$. It is enough to show that if it holds for b_1, b_2, then it holds for $b_1 b_2$. Since $\left(\frac{2}{b_1 b_2}\right) = \left(\frac{2}{b_1}\right)\left(\frac{2}{b_2}\right)$, we need to show that $(-1)^{\left(b_1^2-1\right)/8}(-1)^{\left(b_2^2-1\right)/8} = (-1)^{\left((b_1 b_2)^2-1\right)/8}$, or equivalently that

$$\left(b_1^2 - 1\right) + \left(b_2^2 - 1\right) \equiv \left(b_1 b_2\right)^2 - 1 \mod 16,$$

which is equivalent to $\left(b_1^2 - 1\right)\left(b_2^2 - 1\right) \equiv 0 \mod 4$, which is true because $b_1^2 \equiv b_2^2 \equiv 1 \mod 4$. $\square$

The last quadratic reciprocity property means that we can compute Jacobi symbols, and thus Legendre symbols, in a similar way to computing (a, b) with Euclid's algorithm, although we also have to take care of powers of two in the numerator.

Example. 9283 is prime, and we can compute $\left(\frac{7411}{9283}\right)$ as follows.

$$\begin{aligned} \left(\tfrac{7411}{9283}\right) &= -\left(\tfrac{9283}{7411}\right) = -\left(\tfrac{1872}{7411}\right) = -\left(\tfrac{16}{7411}\right)\left(\tfrac{117}{7411}\right) = -\left(\tfrac{117}{7411}\right) = -\left(\tfrac{7411}{117}\right) = -\left(\tfrac{40}{117}\right) \\ &= -\left(\tfrac{8}{117}\right)\left(\tfrac{5}{117}\right) = -\left(\tfrac{2}{117}\right)\left(\tfrac{5}{117}\right) = \left(\tfrac{5}{117}\right) = \left(\tfrac{117}{5}\right) = \left(\tfrac{2}{5}\right) = -1. \end{aligned}$$

So 7411 is not a square modulo 9283.

7 Sums of squares

Here we answer the question, which integers are representable as a sum of two squares or four squares?

7.1 Sums of two squares

Definition 44. We say that $n \in \mathbb{Z}$ is a **sum of two squares** if there exist $x, y \in \mathbb{Z}$ such that $n = x^2 + y^2$.

Of course there is an algorithm to work this out, which is just to use brute force, as $|x|$ and $|y|$ are both at most $\sqrt{n}$.

Example. $21 \equiv 1 \mod 4$, but 21 is not a sum of two squares. On the other hand, we will see that all primes which are 1 modulo 4 are sums of two squares.

Example. In a negative direction, note that any square is congruent to 0 or 1 modulo 4, so if n is a sum of two squares, then n cannot be congruent to 3 modulo 4.

It is convenient to make use of the following.

Definition 45. The ring of **Gaussian integers**, denoted $\mathbb{Z}[i]$, is the subring of $\mathbb{C}$ consisting of all complex numbers of the form $a+bi$, where $a, b \in \mathbb{Z}$.

Here are some brief recollections about this. To see that $\mathbb{Z}[i]$ is a subring one must of course check that it is closed under addition and multiplication. This is easy. There is a natural **norm** $N : \mathbb{Z}[i] \to \mathbb{Z}_{\geq 0}$ defined by $N(z) = z\overline{z}$. We have $N(a+bi) = a^2 + b^2$. On the other hand, since $\overline{zw} = \overline{z}\overline{w}$, we have $N(zw) = (zw)(\overline{zw}) = (z\overline{z})(w\overline{w}) = N(z)N(w)$. Let $z = a+bi$, $w = c+di$. We then have $zw = (ac-bd)+(ad+bc)i$. Applying the formula $N(zw) = N(z)N(w)$, we find that $\left(a^2+b^2\right)\left(c^2+d^2\right) = (ac-bd)^2 + (ad+bc)^2$. We record this as the following.

Lemma 46. If m and n are sums of two squares, then so is mn.

In light of this result it makes sense to focus first on which primes are a sum of two squares. Indeed, we have the following.

Theorem 47 (Fermat's two square theorem)**.** Every prime congruent to 1 modulo 4 is a sum of two squares.

Lemma 46 and Theorem 47 together allow you to give a complete classification of the integers which are sums of two squares, in terms of their prime factorisations. We will prove this shortly. Recall that in fact $\mathbb{Z}[i]$ with its norm N is a Euclidean domain.

Definition 48. A ring R is a **Euclidean domain** if it is an integral domain, that is $ab = 0$ gives $a = 0$ or $b = 0$, and there exists a function $N : R \to \mathbb{Z}_{\geq 0}$ such that for all $a, b \in R$, with $b \neq 0$, there exist $q, r \in R$ such that

1. $a = qb + r$, and
2. either $r = 0$, or $N(r) < N(b)$.

In particular, in a Euclidean domain R you can carry out Euclid's algorithm, hence the name. In particular, it follows by the same proof that we gave for $\mathbb{Z}$ that prime elements and irreducible elements are the same, and that, up to reordering and multiplication by units, which are the elements ± 1, $\pm i$ with $N(z) = 1$, we have unique factorisation for every element as a product of irreducibles. $\mathbb{Z}[i]$ together with N is a Euclidean domain. By definition, we can rephrase the question of which $n \in \mathbb{Z}$ are representable as a sum of two squares by asking which $n \in \mathbb{Z}$ are norms in $\mathbb{Z}[i]$, so there exists $z \in \mathbb{Z}[i]$ with $N(z) = n$. Since the norm is multiplicative, $N(zw) = N(z)N(w)$, and every element of $\mathbb{Z}[i]$ is a product of primes, we can answer this question by asking what the primes in $\mathbb{Z}[i]$ are, and what their norms are. (TODO Exercise: Show that the units in $\mathbb{Z}[i]$ are ± 1, $\pm i$) Lecture 14

Two elements of $\mathbb{Z}[i]$ are **associates** if their ratio is a unit, that is z, w are associates if $z = uw$ for $u \in \{\pm 1, \pm i\}$. $\mathbb{Z}[i]$ is a Euclidean domain, so in particular we have unique factorisation into primes. Tuesday 06/11/18

Lemma 49. Let p be a prime in $\mathbb{Z}[i]$. Then there exists a prime $q \in \mathbb{Z}$ such that either $N(p) = q$ or $N(p) = q^2$. In the latter case, p is an associate of q, that is $p = uq$ for some unit u. Moreover, given q a prime in $\mathbb{Z}$, there exists a prime $p \in \mathbb{Z}[i]$ such that $N(p) = q$ if and only if q is a sum of two squares.

Proof. Let $n = N(p)$, and factor $n = q_1^{s_1} \dots q_r^{s_r}$ as a product of integer primes. Since by definition, $n = p\overline{p}$, we have that $p \mid n = q_1^{s_1} \dots q_r^{s_r}$ in $\mathbb{Z}[i]$, and so since p is prime, $p \mid q_i$ for some i. Let $q = q_i$. Then we have $p \mid q$ gives $q = pv$ for some v, so $N(p)N(v) = N(pv) = N(q) = q^2$. If $N(p) = 1$, then p would be a unit, a contradiction. Since $N(p)$ is not 1, so $N(p) \mid q^2$ gives $N(p) = q$ or $N(p) = q^2$, as claimed. First suppose that $N(p) = q^2$. Since p divides q we have $q = pv$. Since $N(p) = N(q) = q^2$ we must have $N(v) = 1$, so v is a unit and p is an associate of q, by definition. Suppose $N(p) = q$. Writing $p = a+bi$, we see that $q = a^2 + b^2$. Conversely, if $q = a^2 + b^2$, then $q = (a+bi)(a-bi)$. Since p divides q, we have either p divides $(a+bi)$ or p divides $(a-bi)$, so $N(p) \mid N(a+bi) = q$ or $N(p) \mid N(a-bi) = q$. In either case $N(p)$ divides q and must thus equal q. ☐

Corollary 50. The primes in $\mathbb{Z}[i]$ are either of the form $a+bi$, where $a^2+b^2 \in \mathbb{Z}$ is a prime, or an associate of q, where $q \in \mathbb{Z}$ is a prime that is not a sum of two squares.

Theorem 51. If $p = 2$ or $p \equiv 1 \mod 4$, then p is a sum of two squares.

Proof. By Corollary 50, we just have to show that p is not a prime in $\mathbb{Z}[i]$. There exists n such that $n^2 \equiv -1 \mod p$. If $p = 2$ obvious, and if $p \equiv 1 \mod 4$, $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = 1$ by Euler's criterion, so -1 is a QR modulo p. That is, $p \mid n^2+1 = (n+i)(n-i)$. If p were prime, then $p \mid n+i$ or $p \mid n-i$, that is there exist $c, d \in \mathbb{Z}$ such that $n \pm i = p(c+di)$, so $1 = pd$, a contradiction. It follows that p is not prime, so by Corollary 50, p is a sum of two squares. ☐

This proof looks non-constructive, but in fact it is not. In practice, if you want to go from $n^2+1 \equiv 0 \mod p$ to finding a, b such that $a^2+b^2 = p$, simply use Euclid's algorithm in $\mathbb{Z}[i]$ to compute a GCD $(n+i, p) = a+bi$ of p and $n+i$. Since p divides neither $n+i$ nor $n-i$, this GCD is not a unit, nor is it an associate of p, so its norm is neither 1 nor p^2. Therefore this GCD has norm exactly p. Since we have also shown that primes congruent to 3 modulo 4 are not a sum of two squares, even modulo 4, we have the following complete characterisation of sums of two squares.

Theorem 52. $n \in \mathbb{Z}$ is a sum of two squares if and only if its prime factorisation only contains primes congruent to 3 modulo 4 to even powers, that is of the form

$$n = 2^a \prod_{p_i \equiv 1 \mod 4} p_i^{r_i} \prod_{q_i \equiv 3 \mod 4} q_i^{2s_i}.$$

Proof. Suppose n is of the above form, and write

$$z = (1+i)^a (a_1+b_1 i)^{r_i} \dots (a_k+b_k i)^{r_k} q_1^{s_1} \dots q_h^{s_h},$$

where $p_i = N(a_i+b_i i)$. Then these are all sums of two squares, so $N(z) = n$ is a sum of two squares by Lemma 46. Conversely, if $n = a^2+b^2$ is a sum of two squares, we can write $a+bi$ as a product of primes in $\mathbb{Z}[i]$. Then $n = N(a+bi)$ is the product of the norms of these primes, and we already saw that the norms of primes in $\mathbb{Z}[i]$ are either 2, a prime which is 1 modulo 4, or the square of a prime which is 3 modulo 4. Then $N(a+bi)$ has the claimed form. ☐

7.2 Sums of four squares - the ring of quaternions

We have used the arithmetic of the ring $\mathbb{Z}[i]$ to determine precisely which integers are a sum of two squares. On the other hand, it is a fact, first proved by Lagrange, that every positive integer is a sum of four integer squares. This fact is connected with the arithmetic of a non-commutative ring, which we now describe.

Definition 53. The ring $\mathbb{H}$ of **quaternions** is the ring whose elements are formal sums $a+bi+cj+dk$, with $a, b, c, d \in \mathbb{R}$. Addition is given by the rule

$$(a+bi+cj+dk) + (A+Bi+Cj+Dk) = (a+A) + (b+B)i + (c+C)j + (d+D)k.$$

Multiplication is given by the rules

$$i^2 = j^2 = k^2 = -1, \qquad ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j,$$

extended by $\mathbb{R}$-linearity and the distributive law.

Let $z = a+bi+cj+dk$ be a quaternion. The **conjugate** z^* of z is the quaternion $a-bi-cj-dk$. Note that if z and w are quaternions, then $(zw)^* = w^*z^*$. The norm $N(z)$ of z is defined by $N(z) = zz^* = a^2+b^2+c^2+d^2$. Note that this is a real number, and hence commutes with all elements of $\mathbb{H}$. Thus we have

$$N(zw) = zw(zw)^* = zww^*z^* = zN(w)z^* = zz^*N(w) = N(z)N(w),$$

because $N(w) \in \mathbb{R}$. As was the case with $\mathbb{Z}[i]$, this gives an expression for the product of two sums of four squares as a sum of four squares. Explicitly, one has

$$\begin{aligned}\left(a^2+b^2+c^2+d^2\right)\left(x^2+y^2+z^2+w^2\right) &= N(a+bi+cj+dk)\,N(x+yi+zj+wk) \\ &= N((a+bi+cj+dk)(x+yi+zj+wk)) \\ &= (ax-by-cz-dw)^2+(ay+bx+cw-dz)^2 \\ &+ (az-bw+cx+dy)^2+(aw+bz-cy+dx)^2.\end{aligned}$$

In particular, if $m, n \in \mathbb{Z}$ are representable as sums of four integer squares, then so is their product mn. Thus to prove Lagrange's theorem, it suffices to prove that every prime is a sum of four integer squares.

7.3 Proof of Lagrange's theorem

Let p be a prime. If $p = 2$, or $p \equiv 1 \mod 4$, then we have already shown that p is a sum of two squares, hence also a sum of four squares. It thus remains to prove that primes congruent to 3 modulo 4 are sums of four squares. We will do this via a descent argument. Lecture 15 Wednesday 07/11/18

Lemma 54. Let p be a prime congruent to 3 modulo 4. Then there exist x and y such that $x^2+y^2+1 \equiv 0 \mod p$.

Proof. It suffices to find $a \in \mathbb{Z}$ such that a is a square modulo p and $a+1$ is not, so $\left(\frac{a}{p}\right) = 1$ and $\left(\frac{a+1}{p}\right) = -1$. Since $p \equiv 3 \mod 4$, $\left(\frac{-1}{p}\right) = -1$ by Euler's criterion, so -1 is not a QR modulo p. So we would then have $\left(\frac{-(a+1)}{p}\right) = \left(\frac{a+1}{p}\right)\left(\frac{-1}{p}\right) = 1$, so $-(a+1)$ is a QR modulo p. Taking x such that $x^2 \equiv a \mod p$ and y such that $y^2 \equiv -(a+1) \mod p$. Then $x^2+y^2 \equiv -1 \mod p$ and the claim would follow. Suppose that we cannot do this. Then for each square a modulo p, $a+1$ would also be a square modulo p. In particular since $\left(\frac{1}{p}\right) = 1$, we must have $\left(\frac{2}{p}\right) = \dots = \left(\frac{p-1}{p}\right) = 1$, so every congruence class modulo p would be a square. But we know that there are $(p-1)/2$ values of b with $1 \le b \le p-1$ and $\left(\frac{b}{p}\right) = -1$, a contradiction. Since this does not happen we are done. □

Fix a prime congruent to 3 modulo 4. By Lemma 54, we can find $x, y \in \mathbb{Z}$ such that $x^2+y^2+1 = pr$ for some $r \in \mathbb{Z}$. Since we only care about x and y modulo p and the congruence $x^2+y^2+1 \equiv 0 \mod p$ only depends on x, y modulo p, we can further arrange x, y with $|x|, |y| \le p/2$. Then $\left(x^2+y^2+1\right)/p = r < p$. We are now ready to begin our descent.

Proposition 55. Suppose that for some $1 \le r < p$, we have x, y, z, w such that $x^2+y^2+z^2+w^2 = pr$. Then there exist $x', y', z', w', r' \in \mathbb{Z}$ with $1 \le r' < r$, and $(x')^2+(y')^2+(z')^2+(w')^2 = pr'$.

Proof. There are two cases we must treat separately.

1. First suppose that r is even. Then either all of x, y, z, w have the same parity, so all even or all odd, or two of them are odd and two of them are even. Permuting x, y, z, w as necessary, we can assume without loss of generality x and y have the same parity, as do z and w, so $x \equiv y \mod 2$ and $z \equiv w \mod 2$. Then set
$$x' = \frac{x+y}{2}, \qquad y' = \frac{x-y}{2}, \qquad z' = \frac{z+w}{2}, \qquad w' = \frac{z-w}{2}, \qquad r' = \frac{r}{2}.$$
It is then easy to verify that $(x')^2+(y')^2+(z')^2+(w')^2 = pr'$.

2. Now suppose that r is odd. Choose a, b, c, d such that
$$-\frac{r}{2} < a, b, c, d < \frac{r}{2}, \qquad x \equiv a \mod r, \qquad y \equiv b \mod r, \qquad z \equiv c \mod r, \qquad w \equiv d \mod r.$$
We can get away with strict inequalities here because r is odd. Since $x^2+y^2+z^2+w^2 = pr \equiv 0 \mod r$, our congruences imply that $a^2+b^2+c^2+d^2 \equiv 0 \mod r$. Write $a^2+b^2+c^2+d^2 = rr'$, and

note that $0 \leq r' < r$ since $a^2, b^2, c^2, d^2 < (r/2)^2 = r^2/4$. On the other hand if r' were zero, then $a = b = c = d = 0$, so all of x, y, z, w would be divisible by r. Since $x^2 + y^2 + z^2 + w^2 = pr$, we would have $r^2 \mid rp$, hence $r \mid p$, and since $r < p$, we get $r = 1$, and we are done. Otherwise we have $1 \leq r' < r$. We then have

$$\begin{aligned}(rr')(rp) &= \left(a^2 + b^2 + c^2 + d^2\right)\left(x^2 + y^2 + z^2 + w^2\right) \\ &= (ax + by + cz + dw)^2 + (-ay + bx + cw - dz)^2 \\ &\quad + (-az - bw + cx + dy)^2 + (-aw + bz - cy + dx)^2 .\end{aligned}$$

Note that

$$\begin{aligned} ax + by + cz + dw &\equiv x^2 + y^2 + z^2 + w^2 \equiv 0 \mod r, \\ -ay + bx + cw - dz &\equiv -xy + yx + zw - wz \equiv 0 \mod r, \\ -az - bw + cx + dy &\equiv -xz - yw + zx + wy \equiv 0 \mod r, \\ -aw + bz - cy + dx &\equiv -xw + yz - zy + wx \equiv 0 \mod r. \end{aligned}$$

We thus have integers

$$x' = \frac{ax + by + cz + dw}{r}, \qquad y' = \frac{-ay + bx + cw - dz}{r},$$

$$z' = \frac{-az - bw + cx + dy}{r}, \qquad w' = \frac{-aw + bz - cy + dx}{r},$$

and $(x')^2 + (y')^2 + (z')^2 + (w')^2 = pr'$ as desired.

□

Proposition 55 gives p is a sum of four squares. To complete the proof, one begins with $x^2 + y^2 + 1 = pr$ as above with $z = 1$ and $w = 0$ and repeatedly applies the proposition until $r = 1$.

Remark 56. This descent can be interpreted as a version of Euclid's algorithm in a non-commutative setting. The associated ring is the ring of quaternions of the form $a + bi + cj + dk$, where either $a, b, c, d \in \mathbb{Z}$, or $a, b, c, d \in \mathbb{Z}/2$, that is fractions of the form $r/2$ with r odd. That is,

$$\left\{ \frac{a + bi + cj + dk}{2} \;\middle|\; a \equiv b \equiv c \equiv d \mod 2 \right\}.$$

Note. This ring is non-commutative, and also $5 = (1 - 2i)(1 - 2i) = (1 + 2j)(1 - 2j)$ for example, so you have to be careful with unique factorisation, etc.

7.4 Sums of three squares

At this point it is natural to ask which positive integers are a sum of three integer squares. This turns out to be much more difficult. 7 is the smallest positive integer which is not a sum of three squares. In one direction, no integer congruent to 7 modulo 8 can be a sum of three squares, because the squares modulo 8 are $0, 1, 4$. In fact no integer of the form $4^a(8k + 7)$ for $t, k \in \mathbb{Z}$ is a sum of three squares. (TODO Exercise) Conversely, one has the following.

Theorem 57. Every positive integer is not a sum of three squares if and only if it is of the form $4^a(8k + 7)$.

The proof of this requires tools beyond the scope of the class, such as the Hasse principle for quadratic forms. One good place to read about this is in Serre's a course in arithmetic.

8 Pell's equation

8.1 Pell's equation

Let $d \in \mathbb{Z}_{>1}$ be squarefree, and consider the equation $x^2 - dy^2 = 1$. This is called **Pell's equation**.

Example. For $d = 2$, $(x, y) = (3, 2)$ is a solution. In fact, there are infinitely many solutions, and this is true for any d.

Here is another way of thinking about this, generalizing what we said about the Gaussian integers. We will find it useful to write $x^2 - dy^2 = \left(x + \sqrt{d}y\right)\left(x - \sqrt{d}y\right)$. This suggests that we should look at a ring like $\mathbb{Z}\left[\sqrt{d}\right] = \left\{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\right\}$.

Definition 58. Let $\alpha \in \mathbb{C}$. Then the ring $\mathbb{Z}[\alpha]$ is the smallest subring of $\mathbb{C}$ containing α.

Note. Note that for this course, a ring always contains 1.

As usual one has to check that this makes sense. An alternative definition is to take $\mathbb{Z}[\alpha]$ to be the intersection of all subrings of $\mathbb{C}$ containing α. This intersection clearly contains 0, 1, and α, so it is non-empty. It is also closed under addition and multiplication, since it is an intersection of sets that are closed under these operations. Therefore it is a subring of $\mathbb{C}$ that, by construction is contained in any subring of $\mathbb{C}$ that contains α.

Example. If $\alpha = 1$, then $\mathbb{Z}[\alpha] = \mathbb{Z}$. Let us show that for $\alpha = i$, this agrees with our earlier definition of $\mathbb{Z}[i]$. We have already shown that the set of all integers of the form $a + bi$ is a subring of $\mathbb{C}$. On the other hand, any subring of $\mathbb{C}$ containing i is closed under addition and multiplication, and contains 1, so it contains every complex number of the form $a + bi$. Thus our new definition is consistent with our old one.

Note. For arbitrary α, it is not necessarily true that $\mathbb{Z}[\alpha]$ consists of all complex numbers of the form $a + b\alpha$ for $a, b \in \mathbb{Z}$, although it always contains all complex numbers of that form.

Example.

1. To see this, consider examples like $\mathbb{Z}[\pi]$, the ring of $a_0 + \dots + a_n\pi^n$ for $a_i \in \mathbb{Z}$ and n arbitrary.
2. Also $\mathbb{Z}[1/p]$ for p some prime contains $1/p^n$ for all n, so in fact $\mathbb{Z}[1/p] = \{a/p^n \mid a \in \mathbb{Z},\ n \geq 0\}$.
3. Also $\mathbb{Z}[\beta]$ where $\beta = \sqrt[3]{2}$ is a cube root of 2 is not just the set $\left\{a + b\sqrt[3]{2} \mid a, b \in \mathbb{Z}\right\}$, because the set does not contain $\beta^2 = \left(\sqrt[3]{2}\right)^2 = \sqrt[3]{4}$, which is not of the form $a + b\sqrt[3]{2}$ for any $a, b \in \mathbb{Z}$.

Lecture 16 is a problem class. Lecture 16 Friday 09/11/18

8.2 Quadratic subrings of $\mathbb{C}$

Lecture 17 Tuesday 13/11/18

Definition 59. An element α of $\mathbb{C}$ is an **algebraic integer of degree two**, alternatively a **quadratic algebraic integer**, if there exists a polynomial of the form $P(X) = X^2 + aX + b$ with $a, b \in \mathbb{Z}$ and $\alpha \notin \mathbb{Z}$ such that $P(X)$ has no rational, equivalently integer, roots and $P(\alpha) = 0$.

Example.

1. $\alpha = i$ is an algebraic integer of degree two and a root of $X^2 + 1$ since $i^2 + 1 = 0$.
2. $\alpha = \sqrt{d}$ is a root of $X^2 - d$ for $d > 1$ squarefree.

We thus have the following.

Proposition 60. If α is an algebraic integer of degree two, then $\mathbb{Z}[\alpha]$ is equal to the set of complex numbers $\{x + y\alpha \mid x, y \in \mathbb{Z}\}$.

Proof. Let $a, b \in \mathbb{Z}$ such that $\alpha^2 + a\alpha + b = 0$. Since $\alpha \notin \mathbb{Z}$, we have $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. If $\alpha = r/s$ for $(r, s) = 1$ then $r^2 + ars + bs^2 = 0$, so $s \mid r^2$, so $s \mid 1$, so $\alpha \in \mathbb{Z}$. So if $x, y \in \mathbb{Z}$ and $x + y\alpha = 0$, then $x = y = 0$. Then, for $x, y, X, Y \in \mathbb{Z}$, we have

$$\begin{aligned} (x + y\alpha)(X + Y\alpha) &= xX + (xY + yX)\alpha + yY\alpha^2 \\ &= xX + (xY + yX)\alpha + yY(a\alpha + b) \\ &= (xX + byY) + (xY + yX + ayY)\alpha. \end{aligned}$$

In particular the set of complex numbers of the form $x + y\alpha$ for $x, y \in \mathbb{Z}$ is closed under addition, subtraction, and multiplication and is therefore a subring of $\mathbb{C}$. Since this subring contains α, it contains $\mathbb{Z}[\alpha]$. On the other hand it is clear that this subring is certainly contained in $\mathbb{Z}[\alpha]$, so the two must be equal. ☐

For α an algebraic integer of degree two, we will say that $\mathbb{Z}[\alpha]$ is a **real quadratic** subring of $\mathbb{C}$ if $\alpha \in \mathbb{R}$, and an **imaginary quadratic** subring of $\mathbb{C}$ if $\alpha \notin \mathbb{R}$. We let α^* denote the root of $X^2 + aX + b = 0$ that is not equal to α. If $z = x + y\alpha \in \mathbb{Z}[\alpha]$, write $z^* = x + y\alpha^*$.

Example.

1. $i^* = -i = \bar{i}$.
2. $\sqrt{d}^* = -\sqrt{d}$.

If $\mathbb{Z}[\alpha]$ is imaginary quadratic, $\alpha^* = \overline{\alpha}$ and $z^* = \overline{z}$ and then, as with $\mathbb{Z}[i]$, we can define a norm $N : \mathbb{Z}[\alpha] \to \mathbb{Z}_{\geq 0}$ by setting $N(z) = z\overline{z}$. Note that $\overline{\alpha}$ is also a root of $X^2 + aX + b$ in this case, so we have $\alpha + \overline{\alpha} = -a$, and $\alpha\overline{\alpha} = b$. Thus we have

$$N(x + y\alpha) = (x + y\alpha)(x + y\overline{\alpha}) = x^2 + xy(\alpha + \overline{\alpha}) + y^2\alpha\overline{\alpha} = x^2 - axy + by^2 \in \mathbb{Z}_{\geq 0}.$$

This is multiplicative, $N(zw) = N(z)N(w)$. If $\mathbb{Z}[\alpha]$ is real quadratic, note that α^* is no longer equal to $\overline{\alpha}$. In this case we define

$$N(x + y\alpha) = (x + y\alpha)(x + y\alpha^*) = x^2 - axy + by^2 \in \mathbb{Z}.$$

We thus get a map $N : \mathbb{Z}[\alpha] \to \mathbb{Z}$. We have $(zw)^* = z^*w^*$ gives $N(z)N(w) = zz^*ww^* = (zw)(zw)^* = N(zw)$. This is again multiplicative, but no longer non-negative, so we can have $N(z) < 0$.

Example.

1. In $\mathbb{Z}[\sqrt{2}]$, $N(1 + \sqrt{2}) = (1 + \sqrt{2})(1 - \sqrt{2}) = -1$.
2. In $\mathbb{Z}\left[\sqrt{d}\right]$, $N\left(\sqrt{d}\right) = \left(\sqrt{d}\right)\left(-\sqrt{d}\right) = -d < 0$.

(TODO Exercise: $N(x + y\alpha) = 0$ if and only if $x = y = 0$)

8.3 Factorisation in quadratic rings

In general, quadratic rings are not Euclidean domains, or even unique factorisation domains. The situation is more complicated, see the algebraic number theory class. First, some definitions.

Definition 61. An element of $\mathbb{Z}[\alpha]$ is a unit if it has a multiplicative inverse, that is, if it lies in the group $\mathbb{Z}[\alpha]^\times$ under multiplication. Two elements $z, w \in \mathbb{Z}[\alpha]$ are associates if there exists a unit $u \in \mathbb{Z}[\alpha]^\times$ such that $z = uw$.

Note. If $u \in \mathbb{Z}[\alpha]^\times$ is a unit, there exists v such that $uv = 1$. Taking norms we find that $1 = N(1) = N(u)N(v)$, so $N(u) = \pm 1$. If $\mathbb{Z}[\alpha]$ is imaginary quadratic, then this means $N(u) = 1$. Conversely, if $N(u) = \pm 1$, then $\pm 1 = N(u) = u(u^*)$, so $u(\pm u^*) = 1$, so either u^* or $-u^*$ is a multiplicative inverse of u, so $u \in \mathbb{Z}[\alpha]^\times$. So $\mathbb{Z}[\alpha]^\times = \{z \in \mathbb{Z}[\alpha] \mid N(z) = \pm 1\}$.

8.4 Back to Pell's equation

If $\alpha = \sqrt{d}$ for $d > 1$ squarefree, then $\mathbb{Z}\left[\sqrt{d}\right]$ is a real quadratic subring of $\mathbb{C}$, and its norm form is given by

$$N\left(x + y\sqrt{d}\right) = \left(x + y\sqrt{d}\right)\left(x + y\sqrt{d}\right)^* = \left(x + y\sqrt{d}\right)\left(x - y\sqrt{d}\right) = x^2 - dy^2.$$

Thus the problem of finding integer solutions to Pell's equation is equivalent to finding elements of norm one in $\mathbb{Z}\left[\sqrt{d}\right]$. Since such elements are units, and the norm is multiplicative, these elements form a multiplicative subgroup

$$\mathbb{Z}\left[\sqrt{d}\right]^{\times,1} = \left\{z \in \mathbb{Z}\left[\sqrt{d}\right] \;\middle|\; N(z) = 1\right\} = \left\{x + y\sqrt{d} \;\middle|\; x^2 - dy^2 = 1\right\},$$

the **1-units** of $\mathbb{Z}\left[\sqrt{d}\right]$. (TODO Exercise: If $\mathbb{Z}[\alpha]$ is imaginary quadratic, show that $\mathbb{Z}[\alpha]^\times = \mathbb{Z}[\alpha]^{\times,1}$ is finite. What are the possibilities for this group?) There are certainly two obvious elements of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$, namely ± 1. All others are of the form $x + y\sqrt{d}$, with $x, y \in \mathbb{Z}$ and y non-zero. We have the following.

Lemma 62. Let $x + y\sqrt{d}$ be an element of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$. Then

1. We have $x > 0$, $y > 0$ if and only if $x + y\sqrt{d} > 1$.
2. We have $x > 0$, $y < 0$ if and only if $0 < x + y\sqrt{d} < 1$,
3. We have $x < 0$, $y > 0$ if and only if $-1 < x + y\sqrt{d} < 0$, and
4. We have $x < 0$, $y < 0$ if and only if $x + y\sqrt{d} < -1$.

Proof. It is clear that if $x, y > 0$ then $x + y\sqrt{d} > y\sqrt{d} \ge \sqrt{d} > 1$. But then $x - y\sqrt{d} = 1/\left(x + y\sqrt{d}\right)$ lies between 0 and 1. So replacing y by $-y$, we get $x > 0$ and $y < 0$, so $0 < x + y\sqrt{d} < 1$. Similarly, replacing (x, y) with $(-x, -y)$, $-x + y\sqrt{d}$ lies between -1 and 0, and $-x - y\sqrt{d} < -1$. But since the four cases are mutually exclusive and exhaust all the possibilities for $x, y \neq 0$, the leftward implications from the right hand side hold as well. ☐

Lemma 63. Let $z = x + y\sqrt{d}$, $z' = x' + y'\sqrt{d}$ be two elements of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$ with $z, z' > 1$, that is x, y, x', y' all positive. Then $z > z'$ if and only if $y > y'$.

Proof. We have $z - 1/z = x + y\sqrt{d} - \left(x - y\sqrt{d}\right) = 2y\sqrt{d}$. Since $z - 1/z$ is increasing for z positive, since its derivative is $1 + 1/z^2 > 0$, we have $z > z'$ if and only if $z - 1/z > z' - 1/z'$, if and only if $y > y'$. ☐

Lecture 18 Wednesday 14/11/18

Suppose we have a non-trivial element z of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$, that is $z \neq \pm 1$. Without loss of generality by replacing z by $\pm z^{\pm 1}$, we can take $z > 1$. So by Lemma 62, if $z = x + y\sqrt{d}$, then $x, y > 0$. Then there exists $\epsilon = x + y\sqrt{d} \in \mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$ such that x and y are both positive and y as small as possible. We will call ϵ the **fundamental 1-unit** in $\mathbb{Z}\left[\sqrt{d}\right]$. By the previous lemmas it is the smallest element of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$ that is greater than one.

Example. For $d = 2$, in $\mathbb{Z}\left[\sqrt{2}\right]$ we have $y = 2$, $x = 3$. So $\epsilon = 3 + 2\sqrt{2}$.

Proposition 64. Suppose that $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1} \neq \{\pm 1\}$, so there exists a non-trivial element of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$. Then every element of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$ is of the form $\pm\epsilon^n$ for some $n \in \mathbb{Z}$, where ϵ is the fundamental 1-unit.

Conversely, $N(\pm\epsilon^n) = N(\pm 1)\, N(\epsilon)^n = 1$.

Proof. Let z be an element of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$, $z \neq \pm 1$. After negating z or replacing z by $1/z$, as necessary, we can assume that $z > 1$. Since ϵ is also greater than one, there exists $n \geq 0$ such that $\epsilon^n \leq z < \epsilon^{n+1}$. Then $1 \leq z\epsilon^{-n} < \epsilon$, so $N\left(z\epsilon^{-n}\right) = N(z)\, N(\epsilon)^{-n} = 1$. Then $z\epsilon^{-n}$ is a 1-unit in $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$. Since ϵ is the smallest 1-unit greater than one by choice, and Lemma 63, we have $z\epsilon^{-n} = 1$, that is $z = \epsilon^n$. $\square$

Example. For $d = 2$, $\epsilon = 3 + 2\sqrt{2}$. $\epsilon^2 = \left(3 + 2\sqrt{2}\right)^2 = 17 + 12\sqrt{2}$, and $17^2 - 2\,(12)^2 = 1$.

8.5 Constructing the fundamental 1-unit

In fact, we will show that there are always elements of $\mathbb{Z}\left[\sqrt{d}\right]^{\times,1}$ other than ± 1, so that we are always in the situation of the preceding proposition. The key idea is to note that $x^2 - dy^2 = 0$ for $x, y > 0$ precisely when x/y is a square root of d, and thus, when d is squarefree, $x^2 - dy^2 = 1$ when x/y is in some sense as close as possible to $\sqrt{d}$. This suggests we should think about approximating $\sqrt{d}$ by rational numbers. So one way to try to find 1-units is to find rational numbers which are good approximations to $\sqrt{d}$. $\left|x - y\sqrt{d}\right| = 1/\left|x + y\sqrt{d}\right|$, which is small. Want to make $\left|x/y - \sqrt{d}\right|$ as small as possible for y of a given size. More generally, it is clear that if $x, y \in \mathbb{Z}$, and $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, then we can make $|x/y - \alpha|$ as small as we want. However, in order to do so we might need to make y large. We will thus be interested in approximations to α where the error $|x/y - \alpha|$ is small compared to C/y^n for various C, n fixed. As n gets larger it will become harder and harder to find such approximations. $n = 0$ is trivial. When $n = 1$ and $C = 1$ the situation is very easy. For any α, and any y, there exists x such that x/y is as close to α as possible, and $|x/y - \alpha| < 1/y$. When $n = 2$ things are much less obvious, but in fact we have the following important result that there always exist infinitely many x, y with $|x/y - \alpha| < 1/y^2$, due to Dirichlet.

Theorem 65 (Dirichlet)**.** Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, and let $Q \in \mathbb{Z}_{>1}$. Then there exist $p, q \in \mathbb{Z}$, with $1 \leq q < Q$, such that $|p - q\alpha| < 1/Q$.

Proof. For $1 \leq k \leq Q - 1$, let $a_k = \lfloor k\alpha \rfloor$, so that $0 < k\alpha - a_k < 1$. Partition the interval $[0, 1]$ into Q subintervals of length $1/Q$, $[0, 1/Q], \dots, [(Q-1)/Q, 1]$. One of these intervals contains some pair of elements of the set

$$\{0, \alpha - a_1, \dots, (Q-1)\,\alpha - a_{Q-1}, 1\},$$

which contains $Q + 1$ elements. The difference between these two elements is of the form $p - q\alpha$, where $p, q \in \mathbb{Z}$ and $1 \leq q < Q$, and this difference is less than $1/Q$. $\square$

Corollary 66. For any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, there are infinitely many pairs $p, q \in \mathbb{Z}$ such that $|\alpha - p/q| < 1/q^2$.

Proof. Certainly there exists p for $q = 1$. It suffices to show, given any p, q with $|\alpha - p/q| < 1/q^2$, that we can find another p', q' with $|\alpha - p'/q'| < 1/\left(q'\right)^2$ and $|\alpha - p'/q'| < |\alpha - p/q|$. Suppose given such a p, q, and choose Q such that $1/Q < |\alpha - p/q|$. Then by Theorem 65, there exist p', q' with $1 \leq q' < Q$, and $|\alpha - p'/q'| < 1/Qq' < 1/\left(q'\right)^2$. Also $|\alpha - p'/q'| < 1/Qq' \leq 1/Q < |\alpha - p/q|$, as required. The claim follows. $\square$

We can now show the following.

Theorem 67. For any squarefree $d > 1$, there is a non-trivial solution $x, y \neq 0$ to $x^2 - dy^2 = 1$.

Proof. Corollary 66 gives us infinitely many pairs (p_i, q_i) such that $p_i, q_i > 0$ and $\left|p_i/q_i - \sqrt{d}\right| < 1/q_i^2$, that is $\left|p_i - q_i\sqrt{d}\right| < 1/q_i$. Note that then

$$\left|p_i + q_i\sqrt{d}\right| \leq \left|p_i - q_i\sqrt{d}\right| + 2q_i\sqrt{d} < 1/q_i + 2q_i\sqrt{d} < 3q_i\sqrt{d}.$$

We thus have

$$\left|N\left(p_i + q_i\sqrt{d}\right)\right| = \left|p_i + q_i\sqrt{d}\right|\left|p_i - q_i\sqrt{d}\right| < \left(3q_i\sqrt{d}\right)(1/q_i) = 3\sqrt{d}.$$

Thus for some M between $-3\sqrt{d}$ and $3\sqrt{d}$ there are infinitely many pairs (p_i, q_i) such that $N\left(p_i + q_i\sqrt{d}\right) = M$ for infinitely many i. Since there are finitely many congruence classes modulo M, there is some pair (p_0, q_0) such that there are infinitely many pairs (p_i, q_i) with

$$N\left(p_i + q_i\sqrt{d}\right) = M, \qquad p_i \equiv p_0 \mod M, \qquad q_i \equiv q_0 \mod M,$$

for infinitely many i. Now for (p_i, q_i) and (p_j, q_j) any two distinct such pairs of this form, that is

$$N\left(p_i + q_i\sqrt{d}\right) = N\left(p_j + q_j\sqrt{d}\right) = M, \qquad p_i \equiv p_j \mod M, \qquad q_i \equiv q_j \mod M,$$

consider the quotient

$$\frac{\left(p_i - q_i\sqrt{d}\right)}{\left(p_j + q_j\sqrt{d}\right)} = \frac{\left(p_i - q_i\sqrt{d}\right)\left(p_j + q_j\sqrt{d}\right)}{M} = \frac{(p_ip_j - dq_iq_j) + (p_iq_j - p_jq_i)\sqrt{d}}{M}.$$

The congruence conditions show that $p_iq_j \equiv p_jq_i \mod M$ and $p_ip_j - dq_iq_j \equiv p_i^2 - dq_i^2 = M \equiv 0 \mod M$. So $\left(p_i - q_i\sqrt{d}\right) / \left(p_j - q_j\sqrt{d}\right) \in \mathbb{Z}\left[\sqrt{d}\right]$, and it has norm $N\left(\left(p_i - q_i\sqrt{d}\right) / \left(p_j - q_j\sqrt{d}\right)\right) = M/M = 1$ by multiplicativity of the norm, as required. ☐

Lecture 19 Friday 16/11/18

8.6 The equation $x^2 - dy^2 = -1$

Note that we can also apply these techniques to solving $x^2 - dy^2 = -1$. Solutions correspond to elements of the norm negative one in $\mathbb{Z}\left[\sqrt{d}\right]$. If $u = x + y\sqrt{d} \in \mathbb{Z}\left[\sqrt{d}\right]^\times$ is one given solution, then since this is a unit, the others are given by $\pm u\epsilon^n$ for $n \in \mathbb{Z}$, where ϵ is the fundamental 1-unit. $N(v) = -1$ if and only if $N(v) = N(u)$, if and only if $N(v/u) = 1$.

Note. However, unlike for 1 units there may be no -1-units at all.

Example. Consider $d = 3$, where the equation has no solutions modulo three and thus no integer solutions, but when $d = 2$ we can take $x = y = 1$ as a solution.

9 Continued fractions

9.1 Rational continued fractions

Given $p/q \in \mathbb{Q}$, we can write

$$\frac{p}{q} = a_0 + r_0, \qquad a_0 = \left\lfloor \frac{p}{q} \right\rfloor \in \mathbb{Z}, \qquad 0 \le r_0 < 1.$$

If r_0 is not zero, then we can write

$$\frac{1}{r_0} = a_1 + r_1, \qquad a_1 = \left\lfloor \frac{1}{r_0} \right\rfloor \in \mathbb{Z}, \qquad 0 \le r_1 < 1.$$

We then have

$$\frac{p}{q} = a_0 + r_0 = a_0 + \frac{1}{a_1 + r_1}.$$

Continuing in this way, as long as r_i is non-zero we set

$$\frac{1}{r_i} = a_{i+1} + r_{i+1}, \qquad a_i = \left\lfloor \frac{1}{r_i} \right\rfloor \in \mathbb{Z}, \qquad 0 \le r_{i+1} < 1.$$

The denominators of the r_i are strictly decreasing, so eventually some $r_n = 0$ and we have

$$\frac{p}{q} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}.$$

This expression is called the **continued fraction expansion** of p/q. It is closely related to Euclid's algorithm. Indeed, it is not hard to see that the a_i are the quotients q_i from Euclid's algorithm applied to the pair p, q.

Note. Each $a_i \in \mathbb{Z}$ and for $i \geq 1$, $a_i \geq 1$.

Example.

$$\frac{40}{19} = 2 + \frac{2}{19}, \qquad \frac{19}{2} = 9 + \frac{1}{2} \qquad \Longrightarrow \qquad \frac{40}{19} = 2 + \cfrac{1}{9 + \cfrac{1}{2+0}}.$$

9.2 Infinite continued fractions

Now let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. As above, we can write

$$\alpha = a_0 + r_0, \qquad a_0 = \lfloor \alpha \rfloor \in \mathbb{Z}, \qquad 0 \leq r_0 < 1,$$

and then for each i set

$$\frac{1}{r_i} = a_{i+1} + r_{i+1}, \qquad a_{i+1} = \left\lfloor \frac{1}{r_i} \right\rfloor \in \mathbb{Z}, \qquad 0 \leq r_{i+1} < 1.$$

By definition, $a_i \geq 1$ if $i > 0$.

Note. Unlike in the rational case, this sequence will never terminate, as r_i is always irrational and thus never zero.

We write

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots}}},$$

and call this the continued fraction expansion of α.

Note. So far this is just a formal expression. In fact, we can make mathematical sense of this expression, but it requires some justification.

Example. Let $\alpha = \sqrt{3}$.

$$a_0 = 1, \qquad r_0 = \sqrt{3} - 1, \qquad \frac{1}{r_0} = \frac{1}{\sqrt{3}-1} = \frac{\sqrt{3}+1}{2} = 1 + \frac{\sqrt{3}-1}{2}.$$

$$a_1 = 1, \qquad r_1 = \frac{\sqrt{3}-1}{2}, \qquad \frac{1}{r_1} = \frac{2}{\sqrt{3}-1} = \sqrt{3} + 1 = 2 + \left(\sqrt{3} - 1\right).$$

$$a_2 = 2, \qquad r_2 = \sqrt{3} - 1 = r_0, \qquad \frac{1}{r_2} = \frac{1}{\sqrt{3}-1} = \frac{1}{r_0},$$

and we carry on getting alternating $a_i = 1$ for $i > 0$ odd and $a_i = 2$ for $i > 0$ even forever.

First, we introduce some useful notation. For $a_0, \ldots, a_n \in \mathbb{R}$, we define

$$[a_0; a_1, \ldots, a_n] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}} \in \mathbb{R},$$

when this expression is well-defined. That is, when it does not involve a division by zero. The value of this expression can be computed by a recurrence relation, as follows.

Lemma 68. Given $a_0, \ldots, a_n \in \mathbb{R}$, define p_i, q_i for $0 \le i \le n$ by

$$p_0 = a_0, \qquad q_0 = 1, \qquad p_1 = a_0 a_1 + 1, \qquad q_1 = a_1, \qquad p_i = a_i p_{i-1} + p_{i-2}, \qquad q_i = a_i q_{i-1} + q_{i-2}.$$

Then $[a_0; a_1, \ldots, a_n] = p_n/q_n$, assuming no q_i is zero.

Proof. We prove this by induction on n. The cases $n = 0$, $a_0 = a_0/1$, and $n = 1$, $a_0 + 1/a_1 = (a_0 a_1 + 1)/a_1$, are clear. Let p_i' and q_i' for $0 \le i \le n-1$ be the numbers defined by the recurrence attached to the sequence $a_0, \ldots, a_{n-2}, a_{n-1} + 1/a_n$. By definition, the sequences defining p_i', q_i' and p_i, q_i agree for $i \le n-2$. The inductive hypothesis tells us that $[a_0; a_1, \ldots, a_{n-2}, a_{n-1} + 1/a_n] = p_{n-1}'/q_{n-1}'$. By definition, $[a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_{n-2}, a_{n-1} + 1/a_n]$. So we only need to show that $p_{n-1}'/q_{n-1}' = p_n/q_n$. By the recurrence defining the p_i' and q_i', we have

$$\begin{aligned}
\frac{p_{n-1}'}{q_{n-1}'} &= \frac{\left(a_{n-1} + \frac{1}{a_n}\right) p_{n-2}' + p_{n-3}'}{\left(a_{n-1} + \frac{1}{a_n}\right) q_{n-2}' + q_{n-3}'} = \frac{\left(a_{n-1} + \frac{1}{a_n}\right) p_{n-2} + p_{n-3}}{\left(a_{n-1} + \frac{1}{a_n}\right) q_{n-2} + q_{n-3}} \\
&= \frac{(a_n a_{n-1} + 1) p_{n-2} + a_n p_{n-3}}{(a_n a_{n-1} + 1) q_{n-2} + a_n q_{n-3}} = \frac{a_n (a_{n-1} p_{n-2} + p_{n-3}) + p_{n-2}}{a_n (a_{n-1} q_{n-2} + q_{n-3}) + q_{n-2}} = \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}} = \frac{p_n}{q_n},
\end{aligned}$$

as claimed. □

Note. If $a_i \ge 1$ for all $i \ge 1$, as will be the case, for instance, if the a_i come from taking a continued fraction expansion of some real number, then the q_i are all non-zero and form a strictly increasing sequence. Indeed, if all a_i are at least one, then $q_i = a_i q_{i-1} + q_{i-2} \ge q_{i-1} + q_{i-2} \ge 2q_{i-2}$, so this sequence increases exponentially fast.

Now suppose we have an infinite sequence $a_0, a_1, \dots \in \mathbb{R}$, and assume that $a_i \ge 1$ for all $i \ge 1$. Define p_i and q_i by the recurrence given above. We call p_i/q_i is the ***i*-th convergent** to the continued fraction

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots}}}.$$

Lemma 69. For all n, we have $p_n q_{n-1} - q_n p_{n-1} = (-1)^{n-1}$.

Proof. This is again by induction on n. The base case $n = 1$ is clear. For inductive step, assume this is true for $n - 1$. Then by defining recurrence for the p_i and q_i, we have

$$\begin{aligned}
p_n q_{n-1} - q_n p_{n-1} &= (a_n p_{n-1} + p_{n-2}) q_{n-1} - (a_n q_{n-1} + q_{n-2}) p_{n-1} \\
&= p_{n-2} q_{n-1} - q_{n-2} p_{n-1} \\
&= -(p_{n-1} q_{n-2} - q_{n-1} p_{n-2}) \\
&= -(-1)^{n-2},
\end{aligned}$$

and the claim follows. □

Note. In particular if $a_i \in \mathbb{Z}$, then $p_i, q_i \in \mathbb{Z}$, and Lemma 69 gives $(p_n, q_n) = 1$.

In general, Lemma 69 gives that $|p_n/q_n - p_{n-1}/q_{n-1}| = 1/q_nq_{n-1}$. If $a_i \geq 1$ for all $i \geq 1$, then the sequence q_i increases exponentially. It follows that $\sum_{i=1}^{n} 1/q_iq_{i-1}$ converges, so that the (p_n/q_n) form a Cauchy sequence, and hence converge to a real number.

Lecture 20 Tuesday 20/11/18

A natural question to ask at this point is, if the sequence $a_0, a_1, \dots$ arises from the continued fraction expansion of $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, is the limit of the convergents p_n/q_n equal to α? This is indeed the case.

Lemma 70. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, and let $a_0, a_1, \dots \in \mathbb{Z}$ be the sequence arising from its corresponding continued fraction expansion. Let p_n/q_n be the n-th convergent. Then $p_n/q_n < \alpha$ if n is even, and $p_n/q_n > \alpha$ if n is odd.

Proof. Once again we use induction on n. The case $n = 0$, $a_0 = \lfloor \alpha \rfloor < \alpha$ and $p_0/q_0 = a_0/1 = a_0$, is clear. Note that $[a_1; a_2, \dots, a_n]$ is the $(n-1)$-th convergent to $1/(\alpha - a_0)$. Thus by the induction hypothesis, if n is odd we have

$$[a_1; a_2, \dots, a_n] < \frac{1}{\alpha - a_0},$$

since $\alpha = a_0 + 1/\dots$. That is,

$$\alpha - a_0 < \frac{1}{[a_1; a_2, \dots, a_n]} \qquad \Longrightarrow \qquad \alpha < a_0 + \frac{1}{[a_1; a_2, \dots, a_n]} = [a_0; a_1, \dots, a_n] = \frac{p_n}{q_n}.$$

If n is even the same argument works with the inequalities reversed. $\square$

Corollary 71. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, and let $a_0, a_1, \dots \in \mathbb{Z}$ be the sequence arising from its continued fraction expansion. Let $p_n/q_n = [a_0; a_1, \dots, a_n]$ be the n-th convergent. Then $|\alpha - p_n/q_n| < 1/q_nq_{n+1}$. In particular the limit p_n/q_n as n approaches infinity is α.

Proof. Since exactly one of $p_n/q_n < \alpha < p_{n+1}/q_{n+1}$ and $p_n/q_n > \alpha > p_{n+1}/q_{n+1}$ holds by Lemma 70, we have

$$\left| \frac{p_n}{q_n} - \alpha \right| < \left| \frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}} \right| \leq \frac{1}{q_nq_{n+1}},$$

by Lemma 69. Since the q_n increase exponentially quickly the second claim is clear. $\square$

Note. Since $1/q_nq_{n+1} < 1/q_n^2$, this gives a second, completely constructive proof of Dirichlet's theorem on rational approximations with (p_n/q_n).

9.3 Best approximations

In fact, there is a sense in which the convergents to the continued fraction expansion of α are the best rational approximations to α. We will make this precise below. Fix $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, and as above define a_i and r_i by setting

$$\alpha = a_0 + r_0, \qquad a_0 = \lfloor \alpha \rfloor \in \mathbb{Z}, \qquad 0 < r_0 < 1,$$

and if $i \geq 1$,

$$\frac{1}{r_i} = a_{i+1} + r_{i+1}, \qquad a_{i+1} = \left\lfloor \frac{1}{r_i} \right\rfloor \in \mathbb{Z}_{\geq 1}, \qquad 0 < r_{i+1} < 1.$$

Lemma 72. For all n, we have

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

Proof. $\alpha = [a_0; a_1, \dots, a_n, 1/r_n]$. Set $p_{n+1} = p_n/r_n + p_{n-1}$, $q_{n+1} = q_n/r_n + q_{n-1}$. Then by Lemma 68, $\alpha = p_{n+1}/q_{n+1}$. $\square$

Corollary 73. For all n, we have $|\alpha q_n - p_n| < |\alpha q_{n-1} - p_{n-1}|$, so that $|\alpha - p_n/q_n| < |\alpha - p_{n-1}/q_{n-1}|$.

Proof. By Lemma 72, we have $\alpha(q_n + q_{n-1}r_n) = p_n + p_{n-1}r_n$, so $\alpha q_n - p_n = r_n(p_{n-1} - \alpha q_{n-1})$, and $r_n < 1$. So $|\alpha q_n - p_n| = r_n |\alpha q_{n-1} - p_{n-1}| < |\alpha q_{n-1} - p_{n-1}|$. Since $q_n > q_{n-1}$,

$$\left| \alpha - \frac{p_n}{q_n} \right| = \frac{1}{q_n} |\alpha q_n - p_n| < \frac{1}{q_n} |\alpha q_{n-1} - p_{n-1}| < \frac{1}{q_{n-1}} |\alpha q_{n-1} - p_{n-1}| = \left| \alpha - \frac{p_{n-1}}{q_{n-1}} \right|.$$

$\square$

Theorem 74. Let $h,k\in\mathbb{Z}$ with $0<|k|<q_{n+1}$. Then $|k\alpha-h|\geq|\alpha q_n-p_n|$, with equality only if $|k|=q_n$. If $|k|\leq q_n$, then $|h/k-\alpha|\geq|p_n/q_n-\alpha|$, with equality if and only if $h/k=p_n/q_n$. In other words, p_n/q_n is the best rational approximation to α with denominator at most q_n.

Proof. By Lemma 69 we can find $u,v\in\mathbb{Z}$ with

$$h=up_n+vp_{n+1},\qquad k=uq_n+vq_{n+1}$$

because the corresponding matrix for these linear equations is such that

$$\begin{pmatrix}h\\k\end{pmatrix}=\begin{pmatrix}p_n&p_{n+1}\\q_n&q_{n+1}\end{pmatrix}\begin{pmatrix}u\\v\end{pmatrix}\qquad\Longleftrightarrow\qquad\begin{pmatrix}u\\v\end{pmatrix}=\begin{pmatrix}p_n&p_{n+1}\\q_n&q_{n+1}\end{pmatrix}^{-1}\begin{pmatrix}h\\k\end{pmatrix}=\frac{1}{(-1)^n}\begin{pmatrix}q_{n+1}&-p_{n+1}\\-q_n&p_n\end{pmatrix}\begin{pmatrix}h\\k\end{pmatrix}.$$

Since by assumption, $0<|k|<q_{n+1}$, we have $u\neq 0$, else $k=vq_{n+1}$, $|v|<1$ is a contradiction. Also u,v must have opposite signs if $v\neq 0$, else $|k|=|uq_n|+|vq_{n+1}|\geq q_n+q_{n+1}>q_{n+1}$. If $v=0$, then $h=up_n$, $k=uq_n$, and we are done. Otherwise $v\neq 0$, then write $k\alpha-h=u(\alpha q_n-p_n)+v(\alpha q_{n+1}-p_{n+1})$. u,v have opposite signs. Since αq_n-p_n and $\alpha q_{n+1}-p_{n+1}$ also have opposite signs by Lemma 70, we see that $u(\alpha q_n-p_n)$ and $v(\alpha q_{n+1}-p_{n+1})$ have the same sign. Since

$$|k\alpha-h|=|u(\alpha q_n-p_n)|+|v(\alpha q_{n+1}-p_{n+1})|>|\alpha q_n-p_n|$$

if $u,v\neq 0$, the claim follows, with equality requiring $v=0$. If furthermore $|k|\leq q_n$ then we can multiply this inequality and the inequality $1/|k|\geq 1/q_n$ to get the claim

$$\frac{1}{|k|}|k\alpha-h|\geq\frac{1}{q_n}|q_n\alpha-p_n|\qquad\Longrightarrow\qquad\left|\alpha-\frac{h}{k}\right|\geq\left|\alpha-\frac{p_n}{q_n}\right|.$$

□

Corollary 75. If $h,k\in\mathbb{Z}$ with $|\alpha-h/k|<1/2k^2$, then $h/k=p_n/q_n$ for some n.

Proof. Without loss of generality $k\geq 1$, and we can choose n with $q_n\leq k<q_{n+1}$. Then we have

$$\begin{aligned}\left|\frac{p_n}{q_n}-\frac{h}{k}\right|&\leq\left|\frac{p_n}{q_n}-\alpha\right|+\left|\alpha-\frac{h}{k}\right|=\frac{1}{q_n}|\alpha q_n-p_n|+\frac{1}{k}|\alpha k-h|\\&\leq\left(\frac{1}{q_n}+\frac{1}{k}\right)|\alpha k-h|=k\left(\frac{1}{q_n}+\frac{1}{k}\right)\left|\alpha-\frac{k}{h}\right|<\frac{1}{2k}\left|\frac{1}{q_n}+\frac{1}{k}\right|\leq\frac{1}{kq_n},\end{aligned}$$

by Theorem 74, so our hypothesis gives $|p_n/q_n-h/k|<1/kq_n$. But this forces $p_n/q_n-h/k=0$, as required. □

Lecture 21 Wednesday 21/11/18

9.4 Returning to Pell's equation

Pell's equation is $X^2-dY^2=1$. If (x,y) is a solution, then $\left|\sqrt{d}-x/y\right|$ is small. We can use Corollary 75 to give us a better algorithm for finding solutions to Pell's equation, and as a bonus it lets us solve $x^2-dy^2=-1$ too, when that has solutions.

Proposition 76. Let $d>1$ be squarefree, and let p_n/q_n be the sequence of convergents for the continued fraction for $\sqrt{d}$. If $x,y>0$ and $x^2-dy^2=\pm 1$, then $x=p_n$, $y=q_n$ for some n.

Proof. Note that it is enough to show that $x/y=p_n/q_n$, as then since $(p_n,q_n)=1$, this implies that $x=rp_n$, $y=rq_n$ for some $r\geq 1$, and then $1=x^2-dy^2=r^2\left(p_n^2-dq_n^2\right)$ is divisible by r^2, so $r=1$.

1. Suppose firstly that $x^2-dy^2=1$. By Corollary 75, it suffices to prove that $\left|\sqrt{d}-x/y\right|<1/2y^2$. Then $x-y\sqrt{d}=1/\left(x+y\sqrt{d}\right)>0$, so $x>y\sqrt{d}$, and $x/y>\sqrt{d}$, and we have

$$\left|\frac{x}{y}-\sqrt{d}\right|=\frac{x}{y}-\sqrt{d}=\frac{1}{y}\left(x-y\sqrt{d}\right)=\frac{1}{y}\left(\frac{1}{x+y\sqrt{d}}\right)<\frac{1}{y}\left(\frac{1}{y\sqrt{d}+y\sqrt{d}}\right)=\frac{1}{\left(2\sqrt{d}\right)y^2}<\frac{1}{2y^2},$$

and the claim follows from Corollary 75.

2. Now suppose that $x^2 - dy^2 = -1$. Here we will have to use a trick. Rewrite the equation as $y^2 - (1/d)\, x^2 = 1/d$. Then $y - \left(1/\sqrt{d}\right) x = (1/d) / \left(y + \left(1/\sqrt{d}\right) x\right) > 0$, so $y > \left(1/\sqrt{d}\right) x$, and so

$$\left|\frac{y}{x} - \frac{1}{\sqrt{d}}\right| = \frac{y}{x} - \frac{1}{\sqrt{d}} = \frac{1}{x}\left(y - \frac{x}{\sqrt{d}}\right) = \frac{1}{x}\left(\frac{\frac{1}{d}}{y + \frac{x}{\sqrt{d}}}\right) < \frac{1}{x}\left(\frac{\frac{1}{d}}{\frac{x}{\sqrt{d}} + \frac{x}{\sqrt{d}}}\right) = \frac{\frac{1}{\sqrt{d}}}{2x^2} < \frac{1}{2x^2},$$

so that Corollary 75 implies that y/x is a convergent to the continued fraction of $1/\sqrt{d}$. But $0 < 1/\sqrt{d} < 1$, so we see that if $[a_0; a_1, \ldots]$ is the continued fraction of $\sqrt{d}$, then the continued fraction for $1/\sqrt{d}$ is of the form $[0; a_0, a_1, \ldots]$. Next step is $1/\left(1/\sqrt{d}\right) = \sqrt{d}$. So if $\sqrt{d} = [a_0; a_1, a_2, \ldots]$ then $1/\sqrt{d} = [0; a_0, a_1, \ldots]$, since

$$\sqrt{d} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\cdots}}, \qquad \frac{1}{\sqrt{d}} = 0 + \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{\cdots}}}.$$

So the convergents for $1/\sqrt{d}$ are just the q_n/p_n. So $y/x = q_n/p_n$ for some n, and $x/y = p_n/q_n$ as required.

□

So we now have an algorithm. We just have to go through the convergents, and see if they satisfy $x^2 - dy^2 = \pm 1$. However, we do not yet know just how far along we have to go. In fact, it turns out that there the continued fractions for $\sqrt{d}$ have a particularly nice form. Let us do some more examples.

Example.

1. $\sqrt{3} = [1; 1, 2, 1, 2, \ldots] = \left[1; \overline{1, 2}\right]$.
2. $\sqrt{2} = 1 + \left(\sqrt{2} - 1\right)$, $1/\left(\sqrt{2} - 1\right) = \sqrt{2} + 1 = 2 + \left(\sqrt{2} - 1\right)$, so $\sqrt{2} = \left[1; \overline{2}\right]$.
3. $\sqrt{5} = 2 + \left(\sqrt{5} - 2\right)$, $1/\left(\sqrt{5} - 2\right) = \sqrt{5} + 2 = 4 + \left(\sqrt{5} - 2\right)$, so $\sqrt{5} = \left[2; \overline{4}\right]$.
4. $\sqrt{7} = \left[2; \overline{1, 1, 1, 4}\right]$.
5. $\sqrt{13} = \left[3; \overline{1, 1, 1, 1, 6}\right]$.
6. $\sqrt{43} = \left[6; \overline{1, 1, 3, 1, 5, 1, 3, 1, 1, 12}\right]$.
7. $\sqrt{n^2 + 1} = \left[n; \overline{2n}\right]$.

We can start to see a pattern, or in fact several patterns. The following could be proved in a couple more lectures using similar techniques to the ones we have been using.

Definition 77. The continued fraction expansion $[a_0; a_1, \ldots]$ of $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is **eventually periodic** if there exist $N, d \in \mathbb{Z}_{>0}$ such that $a_n = a_{n+d}$ for all $n \geq N$. We say that it is **periodic** if we can take $N = 0$, so there exists $d \in \mathbb{Z}_{>0}$ such that $a_n = a_{n+d}$ for all n.

Remark 78. The following are facts.

1. The continued fraction of $\sqrt{d}$ is eventually periodic.
2. In fact, it is of the form $\left[a_0; \overline{a_1, \ldots, a_{m-1}, 2a_0}\right]$.
3. The sequence $a_1, \ldots, a_{m-1}$ is symmetric, that is $a_i = a_{m-i}$ for $1 \leq i \leq m - 1$.
4. The n for which $p_n^2 - dq_n^2 = \pm 1$ are precisely the n for which $n \equiv -1 \mod m$, in which case if $n = lm - 1$, then $p_n^2 - dq_n^2 = (-1)^{lm}$.

5. In particular, the fundamental 1-unit is given by $p_{m-1} + q_{m-1}\sqrt{d}$ if m is even and $p_{2m-1} + q_{2m-1}\sqrt{d}$ if m is odd.

6. There is a solution to $x^2 - dy^2 = -1$ if and only if m is odd, in which case the solutions n for which $p_n^2 - dq_n^2 = -1$ for $(x, y) = (p_n, q_n)$ are precisely the $n \equiv m - 1 \mod 2m$.

Example.

1. Let us see how to use this to find solutions for the example of $x^2 - 43y^2 = \pm 1$. Since $\sqrt{43} = \left[6; \overline{1, 1, 3, 1, 5, 1, 3, 1, 1, 12}\right]$ we have $m = 10$, which is even, so there will not be a solution to $x^2 - 43y^2 = -1$. Smallest solution for $x^2 - 43y^2 = 1$ is p_9, q_9. We can use the usual recurrence of Lemma 68.

i	0	1	2	3	4	5	6	7	8	9
a	6	1	1	3	1	5	1	3	1	1
p	6	7	13	46	59	341	400	1541	1941	3482
q	1	1	2	7	9	52	61	235	296	531

It turns out that $p_9 = 3482$, and indeed $3482^2 - 43\left(531\right)^2 = 1$ is the smallest solution.

2. Similarly we can do $x^2 - 13y^2 = \pm 1$. This time $\sqrt{13} = \left[3; \overline{1, 1, 1, 1, 6}\right]$ and $m = 5$ so there is a solution to $x^2 - 13y^2 = -1$, and it will come from p_4, q_4, and p_9, q_9 is the smallest solution for $x^2 - 13y^2 = 1$. In this case the recurrence is really easy.

i	0	1	2	3	4	5	6	7	8	9
a	3	1	1	1	1	6	1	1	1	1
p	3	4	7	11	18	119	137	256	393	649
q	1	1	2	3	5	33	38	71	109	180

$18^2 - 13\left(5\right)^2 = -1$ is the smallest solution, and $649^2 - 13\left(180\right)^2 = 1$. $N\left(18 + 5\sqrt{13}\right) = -1$, so $N\left(\left(18 + 5\sqrt{13}^2\right)\right) = 1$. Note also that it follows from our facts that this is the fundamental 1-unit, that is $p_9 + q_9\sqrt{13}$. $\left(18 + 5\sqrt{13}\right)^2 = 649 + 180\sqrt{13}$ as expected, and this is a much faster way of calculating q_9.

Lecture 22 Friday 23/11/18

9.5 Periodic continued fractions

While we are not going to prove that $\sqrt{d}$ has an eventually periodic continued fraction, we do prove the following easier converse statement.

Definition 79. $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is a **quadratic irrational** if it there is a polynomial $ax^2 + bx + c$, with $a, b, c \in \mathbb{Q}$ not all zero, that has α as a root.

Proposition 80. Suppose that α has an eventually periodic continued fraction expansion. Then α is a quadratic irrational.

Proof.

1. We first show this when the continued fraction of α is periodic. We then have a $d \geq 1$ such that $a_{n+d} = a_n$ for all n. Then
$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\cdots + \cfrac{1}{a_{d-1} + \cfrac{1}{\alpha}}}}.$$
Simplifying the right hand side, or just applying Lemma 72 to $[a_0; a_1, \ldots, a_{d-1}, \alpha]$, we find $w, x, y, z \in \mathbb{Z}$ such that
$$\alpha = \frac{x\alpha + y}{z\alpha + w}.$$
Then $\left(z\alpha + w\right)\alpha - \left(x\alpha + y\right) = 0$, that is $z\alpha^2 + \left(w - x\right)\alpha - y = 0$. Since α is irrational z cannot be zero, so we conclude that α is a quadratic irrational.

2. If α is only eventually periodic there is a β with periodic continued fraction expansion such that we have
$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{\dots + \cfrac{1}{a_N + \cfrac{1}{\beta}}}}.$$
Then β is a quadratic irrational. To complete the proof, it thus suffices to show that if γ is a quadratic irrational, so are $1/\gamma$ and $\gamma + n$ for any $n \in \mathbb{Z}$. Note that if γ is a root of $aX^2 + bX + c = 0$, then $1/\gamma$ is a root of $cX^2 + bX + a = 0$, and $\gamma + n$ is a root of $a\left(X - n\right)^2 + b\left(X - n\right) + c = 0$, so these claims are clear.

$\square$

In fact, the converse is also true. All quadratic irrationals have eventually periodic continued fraction expansions, but we will not prove this.

10 Diophantine approximation

10.1 Liouville's theorem

We have shown, in two different ways, that for any $\alpha \in \mathbb{R} \setminus \mathbb{Q}$, there are infinitely many $p/q \in \mathbb{Q}$ with $\left|p/q - \alpha\right| < 1/q^2$. What happens if we ask for something stronger?

Example. Can we finite infinitely many p/q with $\langle p/q - \alpha \rangle < 1/q^e$ for some $e > 2$?

It turns out that for many irrational α, the answer is no. In particular, let us make the following definition.

Definition 81. Let $d \in \mathbb{Z}_{\geq 1}$. Then $\alpha \in \mathbb{C}$ is **algebraic of degree** d if there exists a degree d, not necessarily monic, polynomial $P\left(x\right)$, with integer coefficients, such that $P\left(\alpha\right) = 0$, and no such polynomial of degree less than d.

Example.

1. $d = 1$ is $\mathbb{Q}$.
2. $d = 2$ is quadratic irrationals.

We then have the following.

Theorem 82 (Liouville's theorem)**.** Let $\alpha \in \mathbb{R}$ be algebraic of degree d. Then for any $e \in \mathbb{R}_{>d}$, there are at most finitely many $p/q \in \mathbb{Q}$ such that $\left|p/q - \alpha\right| < 1/q^e$.

Proof. Let $P\left(x\right)$ be a polynomial of degree d, with coefficients in $\mathbb{Z}$, such that $P\left(\alpha\right) = 0$. Choose $\epsilon > 0$ such that $P\left(x\right)$ has no roots other than α on the closed interval $\left[\alpha - \epsilon, \alpha + \epsilon\right]$. Write $P\left(x\right) = \left(x - \alpha\right)Q\left(x\right)$. $Q\left(x\right)$ is a monic polynomial with real coefficients, of degree $d - 1$. Since in particular $Q\left(x\right)$ is a continuous, real-valued function, there exists $K \in \mathbb{R}$ such that $\left|Q\left(x\right)\right| \leq K$ on the compact set $\left[\alpha - \epsilon, \alpha + \epsilon\right]$. Now suppose we have p/q with $\left|\alpha - p/q\right| < 1/q^e$. There are only finitely many q such that $1/q^e \geq \epsilon$, so we may assume that q is large enough that $1/q^e < \epsilon$. Now on one hand, since $\left|p/q - \alpha\right| < 1/q^e < \epsilon$, $p/q \in \left[\alpha - \epsilon, \alpha + \epsilon\right]$, we have
$$\left|P\left(\frac{p}{q}\right)\right| = \left|\left(\frac{p}{q} - \alpha\right)Q\left(\frac{p}{q}\right)\right| \leq \left|\frac{p}{q} - \alpha\right|K < \frac{1}{q^e}K.$$
On the other hand, since P has degree d and integer coefficients, the denominator of $P\left(p/q\right)$, when written in lowest terms, is a divisor of q^d. But p/q is not a root of $P\left(x\right)$, so $P\left(p/q\right)$ is non-zero and hence $P\left(p/q\right) \geq 1/q^d$. Note that $P\left(p/q\right) \neq 0$, or we could replace P by P' with $P\left(x\right) = \left(qx - p\right)P'\left(x\right)$. Putting the inequalities together, we get
$$\frac{1}{q^e} > \left|P\left(\frac{p}{q}\right)\right| \geq \frac{1}{q^d}.$$
Rewriting, we find $K > q^{e-d}$, so $K^{1/(e-d)} > q$. Since $e - d > 0$ there are only finitely many q for which this is possible, so only finitely many p/q. $\square$

10.2 Constructing transcendentals

Recall that $\alpha \in \mathbb{C}$ is **algebraic** if it is algebraic of degree d, so there is a polynomial $P(x)$, with integer coefficients, such that $P(\alpha) = 0$, and it is called **transcendental** otherwise. The set of polynomials $P(x)$ with integer coefficients is countable, since each such polynomial has finitely many roots the set of algebraic numbers is countable. Since $\mathbb{R}$ is uncountable this means that transcendental numbers exist, and in a very strong sense, almost every real number is transcendental. In spite of this it is very hard to give an example of a single real number that is provably transcendental. In fact, e and π are examples of transcendental numbers, but this is much harder than what we will do. Liouville's theorem gives one approach to proving that a given number is transcendental, by showing that it admits too many good rational approximations. If we can show that for any $e > 0$, there exist infinitely many p/q such that $|\alpha - p/q| < 1/q^e$, then Liouville's theorem tells us that α cannot be algebraic of any degree, and hence must be transcendental.

Example. Define $\alpha \in \mathbb{R}$ by

$$\alpha = \sum_{n=0}^{\infty} \frac{1}{10^{n!}}.$$

This clearly converges. We can find rational approximations to α simply by truncating the series. For each k, let a_k be the sum

$$\alpha_k = \sum_{n=1}^{k} \frac{1}{10^{n!}}.$$

On one hand, α_k is rational, with denominator $q = 10^{k!}$. On the other hand, we have

$$\begin{aligned} |\alpha - \alpha_k| = \sum_{n=k+1}^{\infty} \frac{1}{10^{n!}} &= \frac{1}{10^{(k+1)!}} \left(1 + \frac{1}{10^{(k+2)!-(k+1)!}} + \frac{1}{10^{(k+3)!-(k+1)!}} + \dots\right) \\ &< \frac{1}{10^{(k+1)!}} \left(1 + \frac{1}{10} + \frac{1}{100} + \dots\right) < \frac{2}{10^{(k+1)!}} = \frac{2}{q^{k+1}}. \end{aligned}$$

Fix $d \in \mathbb{Z}_{>0}$. For any $k > d$, we have $2/q^{k+1} < 1/q^d$. Thus there are infinitely many $p/q = \alpha_k$ such that $|\alpha - p/q| < 1/q^d$. Liouville's theorem thus tells us that α cannot be algebraic of degree d. Since d was arbitrary, take d arbitrarily large, so α must be transcendental.

Lecture 23 Tuesday 27/11/18

10.3 Roth's theorem

Liouville's theorem tells us that algebraic numbers are difficult to approximate well by rationals. In fact, they are even more difficult to approximate than Liouville's theorem would suggest. For instance, we have the following.

Theorem 83 (Roth's theorem)**.** Suppose that $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is algebraic. Then for any $\epsilon > 0$, there are only finitely many $x/y \in \mathbb{Q}$ such that $|\alpha - x/y| < 1/y^{2+\epsilon}$.

In other words, for algebraic numbers, you cannot do any better than Dirichlet's theorem. Roth's theorem is considerably harder to prove than Liouville's, and we will not give a proof in this course. Note that the stronger bound gives proofs that many more additional real numbers are transcendental than Liouville's theorem could.

Example. One can prove with Roth's theorem that the number

$$\beta = \sum_{n=1}^{\infty} \frac{1}{10^{3^n}}$$

is transcendental, which is not possible with Liouville's theorem.

This also shows that higher degree versions of Pell's equation can only have finitely many solutions.

Example. We saw that if $d > 1$ is squarefree, then $x^2 - dy^2 = 1$ has infinitely many solutions with $x, y \in \mathbb{Z}$. Suppose now that $d \in \mathbb{Z}_{>1}$, and consider $x^3 - dy^3 = 1$. Then there are only finitely many pairs $x, y \in \mathbb{Z}$ with $x^3 - dy^3 = 1$.

1. Suppose $d = e^3$ is a cube. Then $x^3 - dy^3 = x^3 - (ey)^3 = 1$, so either $(x, y) = (1, 0)$ or $(x, y) = (0, \pm 1)$ and $d = 1$.

2. Suppose d is not a cube. Then $\sqrt[3]{d} \in \mathbb{R} \setminus \mathbb{Q}$ is algebraic, as it is a root of $X^3 - d = 0$. If $y = 0$ then $x = 1$, while if $y > 0$ then $x > 0$ and if $y < 0$ then $x < 0$. Suppose that $x, y > 0$, so that $x > \sqrt[3]{d}y$, and we have
$$x - \sqrt[3]{d}y = \frac{x^3 - dy^3}{x^2 + x\sqrt[3]{d}y + \sqrt[3]{d^2}y^2} = \frac{1}{x^2 + x\sqrt[3]{d}y + \sqrt[3]{d^2}y^2} < \frac{1}{3\left(\sqrt[3]{d}y\right)^2} = \frac{1}{3\sqrt[3]{d^2}y^2},$$
so that
$$\left|\frac{x}{y} - \sqrt[3]{d}\right| < \frac{1}{3\sqrt[3]{d^2}y^3}.$$
Take any $0 < \epsilon < 1$. Then $3\sqrt[3]{d^2}y^3 < 1/y^{2+\epsilon}$ for any y sufficiently large. So Roth's theorem tells us that there are only finitely many solutions. (TODO Exercise: deal with the case $x, y < 0$ by arguing in the same way that we did when we solved $x^2 - dy^2 = -1$)

Note that Liouville's theorem just fails to prove this result, as $\sqrt[3]{d}$ is algebraic of degree three.

11 Primes in arithmetic progressions

11.1 Primes in arithmetic progressions

A natural question to ask is how the prime numbers are distributed modulo n. Are there infinitely many primes congruent to a modulo n for each a, n? Answer is no in general.

Example. There are finitely many primes congruent to 2 modulo 4, or 0 modulo 2.

It is easy to see that for any a with $(a, n) > 1$, since any number is congruent to a modulo n is divisible by (a, n), there is at most one prime congruent to a modulo n. If $(a, n) = 1$, there is no obvious obstruction.

Example. There are infinitely many primes congruent to 1 modulo 2.

The best possible result is the following.

Theorem 84 (Dirichlet)**.** If $(a, n) = 1$, then there are infinitely many primes congruent to a modulo n.

This was first proven by Dirichlet. The methods involved belong to analytic number theory. See for instance Serre's a course in arithmetic for a proof of this statement. We will instead concern ourselves with special cases of this problem that can be approached by elementary methods, for $a = 1$.

11.2 Elementary results

We first recall the proof that there are infinitely many primes. The structure of this proof will form a template for our arguments.

Theorem 85. There are infinitely many primes.

It suffices to construct, for any finite set S of primes, a prime not in S.

Proof. Let S be a finite set of primes, and let $Q = 1 + \prod_{p \in S} p$. Then $Q > 1$, so Q is divisible by at least one prime q. Then $q \notin S$, so we are done. ☐

It is easy to adapt this proof to show that there are infinitely many primes congruent to 3 modulo 4.

Theorem 86. There are infinitely many primes congruent to 3 modulo 4.

Proof. Let S be a finite set of primes which are congruent to 3 modulo 4, and let $Q = 2 + \prod_{p \in S} p^2$. Then $Q > 1$, and $Q \equiv 3 \mod 4$, so Q is divisible by at least one prime q which has $q \equiv 3 \mod 4$. Then $q \notin S$, so we are done. $\square$

A very similar argument works to show there are infinitely many primes congruent to 5 modulo 6. Handling cases beyond that requires new ideas. For instance, the following holds.

Lemma 87. Let $x \in \mathbb{Z}$ be even, and let p be a prime dividing $x^2 + 1$. Then p is congruent to 1 modulo 4.

Proof. On one hand p is clearly odd. On the other hand, $x^2 + 1 \equiv 0 \mod p$, so $x^2 \equiv -1 \mod p$, so -1 is a QR modulo p and $\left(\frac{-1}{p}\right) = 1$. Hence p is congruent to 1 modulo 4. $\square$

Theorem 88. There are infinitely many primes congruent to 1 modulo 4.

Proof. Let S be a finite set of primes which are congruent to 1 modulo 4, and let $Q = 1 + 4 \prod_{p \in S} p^2 = 1 + \left(2 \prod_{p \in S} p\right)^2$. Then $Q > 1$. By Lemma 87, if q is a prime dividing Q then q is congruent to 1 modulo 4. Then $q \notin S$, so we are done. $\square$

This suggests a strategy for proving there are infinitely many primes congruent to a modulo n for some pair a, n. If we can find a polynomial $P(x)$ such that for any $x \in \mathbb{Z}$, every prime dividing $P(nx)$ is congruent to a modulo n, or at least one, then we can try to mimic the proof that there are infinitely many primes congruent to 1 modulo 4 to show that there are infinitely many primes congruent to a modulo n. When $a = 1$ it turns out this is possible. In fact, it can be shown that this is possible if and only if $a^2 \equiv 1 \mod n$, although this is a hard result. Indeed, the only if direction involves proving a big generalisation of Dirichlet's theorem to number fields. We will prove the case of being $a = 1$ modulo n. The case where n is prime is easier, so we do this first.

Theorem 89. For any prime q, there are infinitely many primes congruent to 1 modulo q.

Definition 90. The **q-th cyclotomic polynomial** $\Phi_q(X)$ is the polynomial $(X^q - 1)/(X - 1) = X^{q-1} + \dots + 1$.

Theorem 91. Let $p \neq q$ be a prime, and let $a \in \mathbb{Z}$. Then $p \mid \Phi_q(a)$ if and only if a has exact order q modulo p.

Proof. a has exact order q modulo p if and only if $a^q \equiv 1 \mod p$ and $a \not\equiv 1 \mod p$. If $p \mid \Phi_q(a)$ then $p \mid a^q - 1$. If also $a \equiv 1 \mod p$, then $\Phi_q(a) \equiv \Phi_q(1) \equiv q \not\equiv 0 \mod p$, a contradiction. Conversely if $a^q \equiv 1 \mod p$, $a \not\equiv 1 \mod p$, then $(a^q - 1)/(a - 1) \equiv 0 \mod p$. $\square$

Lecture 24
Wednesday
28/11/18

Corollary 92. Let $p \neq q$ be prime, and $a \in \mathbb{Z}$. If p divides $\Phi_q(a)$, then $p \equiv 1 \mod q$.

Proof. The order of a modulo p is q by Theorem 91 above. But $a^{p-1} \equiv 1 \mod p$ by Fermat's little theorem. Thus q divides $p - 1$. $\square$

We are now in a position to prove the following.

Theorem 93. Let q be a prime. Then there are infinitely many primes congruent to 1 modulo q.

Proof. Let S be a finite set of primes congruent to 1 modulo q, and let $R = \prod_{p \in S} p$. Consider $\Phi_q(qR) \geq qR + 1 > 1$. Let p be a prime factor of $\Phi_q(qR)$. By Corollary 92, either $p = q$, or $p \equiv 1 \mod q$. But $\Phi_q(qR) = (qR)^{q-1} + \dots + 1 \equiv 1 \mod qR$, so $p \neq q$ and $p \equiv 1 \mod q$. Then $p \notin S$, so we are done. $\square$

11.3 Cyclotomic polynomials

The key to studying the primes congruent to 1 modulo n is a family of polynomials known as the cyclotomic polynomials Φ_n.

Definition 94. Let $n \in \mathbb{Z}_{\geq 1}$. Then the n-th cyclotomic polynomial $\Phi_n(X)$ is the product

$$\Phi_n(X) = \prod_{1 \leq a \leq n,\ (a,n)=1} \left(X - e^{\frac{2\pi a i}{n}} \right).$$

In other words, Φ_n is the monic polynomial whose roots are the primitive n-th roots of unity, all with multiplicity one.

A priori, Φ_n is a polynomial with complex coefficients. However, we observe the following.

Lemma 95. For any n, we have

$$X^n - 1 = \prod_{d \mid n,\ d>0} \Phi_d(X).$$

Proof. Each side is a monic polynomial, so we just need to check that the roots are the same with multiplicities. The roots of $X^n - 1$ are the n-th roots of unity, each with multiplicity one. In other words, they are the primitive d-th roots of unity for some unique d dividing n, each with multiplicity one. These are the same as the roots of the product of the $\Phi_d(X)$. The result follows. $\square$

From this it is easy to deduce the following.

Lemma 96. For any $n \geq 1$, the polynomial $\Phi_n(X)$ has integer coefficients.

Proof. We prove this by strong induction on n. The case $n = 1$, $\Phi_1(X) = X - 1$, is clear. Suppose that the result holds for all $d \mid n$, $d < n$, so $\Phi_d(X)$ has integer coefficients for all $d \mid n$, $d < n$. By Lemma 95, if we set $P(X)$ as the product $P(X) = \prod_{d \mid n,\ 0<d<n} \Phi_d(X)$, we have $X^n - 1 = \Phi_n(X) P(X)$. In particular $P(X)$ is monic with integer coefficients by the induction hypothesis. Let e be the degree $\deg(P(X))$. Write $\Phi_n(X) = \sum_{i=0}^m a_i X^i$ and $P(X) = \sum_{i=0}^e b_i X^i$. Suppose that $\Phi_n(X)$ does not have integer coefficients, so that not all $a_i \in \mathbb{Z}$, and let $q \in \mathbb{Z}$ be maximal such that $a_q \notin \mathbb{Z}$. Since $P(X) = X^e + \dots + a_0$ is monic, the coefficient of X^{q+e} in $\Phi_n(X) P(X)$ is $a_q + \dots + a_{q+e} b_0$, where $a_{q-1} b_{e-1} + \dots + a_{q+e} b_0 \in \mathbb{Z}$. Since $\Phi_n(X) P(X) = X^n - 1 \in \mathbb{Z}[X]$, this is impossible. $\square$

In light of this, we can consider $\Phi_n(X)$ as a polynomial modulo p for various p. In fact, we will show that if p does not divide n, then $\Phi_n(X)$ has no repeated roots modulo p. In order to do this, we need the notion of the derivative of a polynomial with coefficients in an arbitrary field.

Definition 97. Let F be a field, and let $P(X) = \sum_{n=0}^d a_n X^n \in F[X]$ be a polynomial. Then the derivative $P'(X)$ of $P(X)$ is defined as the polynomial $P'(X) = \sum_{n=1}^d n a_n X^{n-1}$.

Note. $(P+Q)' = P' + Q'$ and $(PQ)' = P'Q + PQ'$, as one easily checks.

Lemma 98. Suppose that $P(X)$ has a double root α in F. Then α is a root of both $P(X)$ and $P'(X)$.

Proof. Write $P(X) = (X - \alpha)^2 R(X)$. Then

$$P'(X) = (X-\alpha)^2 R'(X) + 2(X-\alpha) R(X) = (X-\alpha)\left((X-\alpha) R'(X) + 2R(X)\right).$$

$\square$

Corollary 99. If p does not divide n, then $\Phi_n(X)$ has no repeated roots modulo p.

Proof. Since $\Phi_n(X)$ divides $X^n - 1$, it suffices to show that $X^n - 1$ has no repeated roots modulo p. But the derivative of $X^n - 1$ is a non-zero multiple nX^{n-1} of X^{n-1}, since p does not divide n, so its only root is zero, and thus has no common roots with $X^n - 1$. So we are done by Lemma 98. $\square$

Note. If $n = p$, $X^p - 1 \equiv (X-1)^p \mod p$ and $\Phi_p(X) \equiv (X-1)^{p-1} \mod p$.

Theorem 100. Let p be a prime not dividing n, and let $a \in \mathbb{Z}$. Then $p \mid \Phi_n(a)$ if and only if a has exact order n modulo p.

Proof. Suppose that a has exact order n modulo p. Then a is a root of $X^n - 1$ modulo p, but is not a root of $X^d - 1$ modulo p for any $d < n$ dividing n. Since $\Phi_d(X)$ divides $X^d - 1$, it follows that a cannot be a root of $\Phi_d(X)$ for any d dividing n other than $\Phi_n(X)$. Since

$$X^n - 1 = \Phi_n(X) \prod_{d \mid n,\ 0<d<n} \Phi_d(X), \tag{4}$$

we must have a is a root of $\Phi_n(X)$ modulo p, that is $p \mid \Phi_n(a)$. Conversely, suppose that $p \mid \Phi_n(a)$. Then a is a root of $\Phi_n(X)$ modulo p, so by 4, a is a root of $X^n - 1$ modulo p, so the order of a modulo p divides n. We need to show that a is not a root of $X^d - 1$ for any $d \mid n$, $d < n$. Let d be this order, and suppose $d < n$. Writing $X^d - 1 = \prod_{e \mid d} \Phi_e(X)$, we have $a^d - 1 \equiv 0 \mod p$ which means that a would be a root of $\Phi_e(X)$ for some $e \mid d \mid n$. On the other hand by (4), a is a root of both $\Phi_n(X)$ and $\Phi_e(X)$, which are distinct factors of $X^n - 1$. Thus a is a double root of $X^n - 1$ modulo p which is impossible by Corollary 99. Therefore $d = n$ and we are done. $\square$

Corollary 101. Let p be a prime not dividing n, and $a \in \mathbb{Z}$. If $p \mid \Phi_n(a)$, then $p \equiv 1 \mod n$.

Proof. The order of a modulo p is n by Theorem 100 above. Thus n divides $p - 1$, by Fermat's little theorem. $\square$

11.4 Primes congruent to 1 modulo n

We are now in a position to prove the following.

Theorem 102. Let $n \in \mathbb{Z}_{\geq 1}$. Then there are infinitely many primes congruent to 1 modulo n.

Proof. Let S be a finite set of primes congruent to 1 modulo n, and let $R = \prod_{p \in S} p$. For each k, let Q_k be $\Phi_n(knR) \in \mathbb{Z}$. Note that not all Q_k are ± 1, since $\Phi_n(X)$ is a non-constant polynomial. Thus choose k large enough that $Q_k > 1$, so there is a prime p dividing Q_k. Since Q_k divides $(knR)^n - 1$, no prime dividing n or R can divide Q_k. Thus p is not in S, and by Corollary 101 p is congruent to 1 modulo n. $\square$