# Math 124: Number Theory

Hahn Lheem (as Course Assistant)
Taught by Mark Kisin

Fall 2023

## Contents

# 0 Preface

This class is from the Fall 2023 semester. Meeting times are Monday and Friday from 12-1:15pm in SC221. The main textbook is Ireland-Rosen's *A Classical Introduction to Modern Number Theory*.[1] There are no prerequisites to this course, so "if you don't understand something, it's [Kisin's] fault, not yours." Bottom line: don't be afraid to ask questions!

Problem sets will be assigned approximately weekly. Kisin's office hours will be 2-3pm on Wednesdays in SC232. Office hours and section times for the course assistants can be found on Canvas. There will be both a midterm and final exam, which will be "absolutely routine if you do all of the homework." (The problems will be taken from the homework.)

If you see anything wrong or unclear, let me know at hahnlheem@college.harvard.edu!

# 1 09/08 - Unique Prime Factorizations for Integers

A little bit more review of the syllabus. From Ireland-Rosen, the basic plan is to cover chapters 1-8, 10, 11, 16, and 17. The topics include:

- unique factorization,
- congruences,
- Quadratic Reciprocity,
- equations over finite fields,
- Diophantine equations.

The magic of number theory is that very sophisticated tools and results can be developed from the most simple techniques. This means that the beginning of the course may feel slow, perhaps even underwhelming, but it comes together very nicely (and quickly!) in the latter half of the semester. So hold your horses.

## 1.1 Notation

Always good to establish from the getgo.

- $\mathbb{Z} := \{0, \pm 1, \pm 2, \dots\}$
- $\mathbb{N}^+ = \{1, 2, 3, \dots\}$

[1] Kisin believes this textbook should be titled "An Elementary Introduction to Classical Number Theory"; I share this sentiment.

- $\in$ indicates an element of a set, e.g. $5 \in \mathbb{N}^+$
- For $a, b \in \mathbb{Z}$, $a \mid b$ means "$a$ divides $b$"

## 1.2 Integers have unique prime factorization

A longstanding theme of number theory is that integers are understood from the primes. This makes sense: every number can be factorized into primes. Thus, it is important to have an unambiguous definition of prime numbers from the start.

**Definition 1.1** (Prime Number)**.** An integer $p \in \mathbb{N}^+$ is a **prime number** if $p > 1$ and its only divisors are $\pm 1, \pm p$, i.e. if $a \in \mathbb{Z}$ such that $a \mid p$, then $a$ is either $\pm 1$ or $\pm p$.

As mentioned already, primes are important because every integer factors into primes:

**Theorem 1.2** (Unique factorization of integers)
If $n \in \mathbb{N}^+$, then $n$ is a product of primes in a unique way. In other words, we can uniquely express $n = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$, where $p_1 < p_2 < \cdots < p_s$ are prime and $a_i \in \mathbb{N}^+$.

**Remark 1.3.** For $n = 1$, we take the empty product, i.e. when $s = 0$.

We're all familiar with this theorem, perhaps we use it every day. But just how "obvious" is this theorem? Is it immediate from the definition?

The theorem statement seems to incorporate two parts to it. First, we need to show that a prime factorization actually exists in the first place, and then assert that it is unique. This distinction is important: there are some worlds where existence holds, but the factorization is not unique! (For those who have taken Math 122, this may be familiar. Otherwise, an example I like to give is that you can write $6 = 2 \cdot 3$, but if you allow for terms with $\sqrt{-5}$ then suddenly you can also write $6 = (1 + \sqrt{-5})(1 - \sqrt{-5})$. So the factorization is not unique.)

## 1.3 Proving Uniqueness

Because this uniqueness condition is harder to satisfy, it is harder to prove. The existence condition is almost immediate from the definition, which we now demonstrate with the following lemma:

**Lemma 1.4**
Every $n \in \mathbb{N}^+$ is a product of primes.

*Proof.* We prove by (strong) induction on $n$. For $n = 1$, we take the empty product, aka the product of no primes, so our base case is satisfied.

Now suppose the statement is true for all $1 \leq m < n$. Consider $n$. If $n$ is prime, take the obvious factorization $n = n$, done. Otherwise, if $n$ is not prime, then $n$ must have some divisor $a$ where $1 < a < n$. The quotient $n/a$ must also be an integer, call $b$, so $n = a \cdot b$. But since $a, b < n$, by our inductive hypothesis, both $a$ and $b$ are a product of primes! Therefore, $n = a \cdot b$ is also a product of primes, as desired. □

So the subtle thing we are using here that makes this work is the very convenient fact that the integers have a *well-ordering*, i.e. given any two integers $a, b$, we can compare their sizes. This is used in the fact that the two divisors $a, b$ are both less than $n$, which allows us to activate the inductive hypothesis.

Again, this may not seem the most enlightening stuff now, but when we grow up from $\mathbb{N}^+$ to some other set of numbers and try to prove similar statements, the work we've done here will become very handy.

Now, we want to prove that the factorization is unique, up to ordering. For instance, we can factor $36 = 6 \cdot 6 = 2 \cdot 3 \cdot 2 \cdot 3$, or $36 = 4 \cdot 9 = 2 \cdot 2 \cdot 3 \cdot 3$, which are the same up to ordering of the factors. We want to show this is always the case, no matter which $n$ we choose.

We first develop a definition for greatest common divisor, an important concept that will come up again and again.

**Definition 1.5.** If $a_1, a_2, \ldots, a_n \in \mathbb{Z}$, define $(a_1, a_2, \ldots, a_n) = \{a_1x_1 + a_2x_2 + \cdots + a_nx_n \mid x_i \in \mathbb{Z}\}$.

**Remark 1.6.** For those familiar with ring theory, this is the *ideal* in $\mathbb{Z}$ generated by $a_1, a_2, \ldots, a_n$.

**Definition 1.7.** If $a, b \in \mathbb{Z}$, an integer $d$ is called a **greatest common divisor (gcd)** of $a, b$ if

1. $d \mid a$ and $d \mid b$, and
2. whenever $c \in \mathbb{Z}$ such that $c \mid a$ and $c \mid b$, then $c \mid d$.

Might seem weird at first, but if you think about it at first, it makes perfect sense.

**Remark 1.8.** You might be wondering why we don't replace the second condition of gcd with a more familiar condition "whenever $c \mid a$ and $c \mid b$, then $c \leq d$. There are a few reasons why we don't want to do this, the first being that this takes away some factors which would qualify as gcd. For instance, $\gcd(36, 60) = \pm 12$, but using the $\leq$ condition would remove the $-12$ possibility. This may not be a

concern now, but it would cause problems when we go beyond the integers.

The second reason is that it's just not necessary. Why do we need to invoke the well-ordering of the integers when we don't need to? This is a good thing to practice in math: if you don't need something, don't use it. You'll thank yourself in the long run.

Okay, we have this definition. Now we want to know this gcd always exists.

**Lemma 1.9** (Existence of gcd)
If $a, b \in \mathbb{Z}$, then $(a, b) = d\mathbb{Z} = \{d \cdot n \mid n \in \mathbb{Z}\}$ for some $d \in \mathbb{Z}$, and $d$ is a gcd for $a$ and $b$.

*Proof.* Let's take care of the stupid case first. If $a = b = 0$, then $\gcd(a, b) = 0$. Now assume $a \neq 0$. Let $d$ be the smallest positive element of $(a, b)$. We will show $(a, b) = d\mathbb{Z}$.

When we want to show equality of two sets, it is customary to show that one contains the other, and vice versa. One inclusion is immediate: given our choice of $d$ above, it is clear that $d \cdot \mathbb{Z} \subseteq (a, b)$. ($\subseteq$ means "contains".) So we are left to prove $(a, b) \subseteq d \cdot \mathbb{Z}$.

Suppose $e \in (a, b)$. We invoke the Division Algorithm, which tells us that $e = q \cdot d + r$ where $q, r \in \mathbb{Z}$ and $0 \leq r < d$. (This is just saying when you divide $e$ by $d$, you get a quotient $q$ with a remainder $r < d$.) Then, $r = e - qd$, so $r \in (a, b)$ since $d, e \in (a, b)$.

But since we chose $d$ to be the smallest positive element of $(a, b)$, this forces $r = 0$, which means $e = q \cdot d \in d \cdot \mathbb{Z}$. Since this works for any $e \in (a, b)$, we conclude $(a, b) \subseteq d \cdot \mathbb{Z}$, and equality of the two sets follows.

Now we must address the last part of the result: showing that $d = \gcd(a, b)$. Showing that $d$ is a common factor is immediate: $a, b \in (a, b) = d\mathbb{Z}$, so $d \mid a$ and $d \mid b$. Also note that $d \in (a, b)$, so there are integers $x, y$ such that $d = ax + by$. Thus, if an integer $c \in \mathbb{Z}$ satisfies $c \mid a$ and $c \mid b$, then $c \mid ax + by = d$, so $d$ is a gcd of $a, b$. □

Just as a cook must know the ingredients in a recipe and not just the steps, we should understand what are the key things being used in our proofs. Here, the main content of the proof unravels from the Division Algorithm, a simple/intuitive yet very important result in number theory.

We continue to develop some more definitions and results for our proof of unique prime factorization.

**Definition 1.10** (Coprime)**.** If $a, b \in \mathbb{Z}$, we say $a, b$ are **coprime** if $(a, b) = \mathbb{Z}$. (Note: from above, this means $\gcd(a, b) = 1$, i.e. if $c \mid a$ and $c \mid b$, then $c = \pm 1$.)

**Remark 1.11.** Notation: if $(a, b) = d\mathbb{Z}$, i.e. $d = \gcd(a, b)$, then we abuse notation by suppressing the gcd and writing $(a, b) = d$.

**Lemma 1.12**
If $a \mid bc$ and $(a, b) = 1$, then $a \mid c$.

*Proof.* This follows from previous definitions. As $(a, b) = \mathbb{Z}$ (in particular, $1 \in (a, b)$), there exists $r, s \in \mathbb{Z}$ such that $1 = ra + sb$. Thus, $c = rac + sbc$. Clearly, $a \mid rac$, and from assumption, $a \mid s \cdot bc$, so $a \mid c$ as desired. □

What we will use in our proof of unique factorization is a specific version of this where $a = p$ is a prime.

**Corollary 1.13**
If $p$ is prime and $p \mid bc$, then $p \mid b$ or $p \mid c$.

*Proof.* As the only factors of $p$ are $\pm 1, \pm p$, we either have $(p, b) = 1$, in which case $p \mid c$ from above, or $(p, b) = p$, in which case $p \mid b$. □

One more definition, for ease of notation.

**Definition 1.14** (Order)**.** If $p$ is prime and $n \in \mathbb{Z}$, then $\operatorname{ord}_p n$ is the largest integer $a$ such that $p^a \mid n$.

A fact about orders:

**Corollary 1.15**
If $p$ is prime, and $a, b \in \mathbb{Z}$, then $\operatorname{ord}_p(ab) = \operatorname{ord}_p(a) + \operatorname{ord}_p(b)$.

*Proof.* Let $\alpha = \operatorname{ord}_p a$, $\beta = \operatorname{ord}_p b$. By definition, $a = p^\alpha \cdot a'$ and $b = p^\beta \cdot b'$, where $p \nmid a', b'$. We see $ab = p^{\alpha+\beta} a'b'$. Because $p \nmid a'$, $p \nmid b'$ means $p \nmid a'b'$, we get $\operatorname{ord}_p ab = \alpha + \beta = \operatorname{ord}_p a + \operatorname{ord}_p b$, as desired. □

Note that we can apply this above corollary repeated, e.g. $\operatorname{ord}_p(abc) = \operatorname{ord}_p(ab) + \operatorname{ord}_p c = \operatorname{ord}_p a + \operatorname{ord}_p b + \operatorname{ord}_p c$. Now, finally, we are ready to prove uniqueness of prime factorization.

*Proof of second part of Theorem 1.2.* Let $n \in \mathbb{N}^+$. Write $n = p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n}$, where the $p_i$'s are distinct primes and $a_i > 0$. (Note that we already proved a prime factorization exists, so this is fair game.) We want to show that the exponent of any prime $p$ in the factorization depends only on $n$, not on the choice of factorization.

Let $a \coloneqq \operatorname{ord}_p n$. By the above corollary, we have $\operatorname{ord}_p n = \operatorname{ord}_p(p_1^{a_1}) + \operatorname{ord}_p(p_2^{a_2}) + \cdots + \operatorname{ord}_p(p_n^{a_n})$. If $p_i \neq p$, then $p$ does not divide $p_i$, so $\operatorname{ord}_p p_i^{a_i} = 0$. On the other

hand, if $p_i = p$, then $\operatorname{ord}_p n = \operatorname{ord}_p p^{a_i} = a_i$, so $a = a_i$. Since $a$ is determined by $n$ alone ($a = \operatorname{ord}_p n$), $a_i$ is only dependent on $n$ as well. In particular, we can write

$$n = \prod_{p \text{ prime}} p^{\operatorname{ord}_p n},$$

and this is the unique factorization. □

### 1.4 Unique Factorization of $k[x]$

I warned you. We graduate from the integers and now move on to polynomials.

Let $k$ be a field. If you don't know what a field is, think either $\mathbb{Q}$ (the rational numbers), $\mathbb{R}$ (the real numbers), or even $\mathbb{C}$ (the complex numbers). A field, very briefly, is a structure with addition and multiplication, where all nonzero elements have a multiplicative inverse. (So for example, $\mathbb{Z}$ is not a field, since the multiplicative inverse of 2 is 1/2 which is not an integer.)

Denote $k[X]$ as the set of polynomials in $X$ over $k$ (i.e. with coefficients in $k$), so formally

$$k[X] = \{f(X) \coloneqq a_0 + a_1 X + \cdots + a_n X^n \mid a_i \in k, a_n \neq 0 \text{ if } n > 0\}.$$

For example, if $k = \mathbb{R}$, then $\sqrt{2} + \pi X + eX^2$ is a polynomial in $\mathbb{R}[X]$. If $f(X) = a_0 + a_1 X + \cdots + a_n X^n$, where $a_n \neq 0$, we say the **degree** of $f$ is $\deg f = n$.

## 2 9/11 - Generalizing Unique Factorization

### 2.1 Unique Factorization of $k[x]$, cont.

Today, we will try to transfer our work in $\mathbb{Z}$ (namely, unique factorization) to our new set $k[X]$. Let's lay out some definitions beforehand to ease our work moving forward:

**Definition 2.1** (Irreducible)**.** Let $f \in k[x]$, and $\deg f > 0$. Then, $f$ is called **irreducible** if whenever $f = gh$, either $\deg g = 0$ or $\deg h = 0$.

**Definition 2.2** (Monic)**.** A polynomial $f(X) = a_0 + a_1 X + \cdots + a_m X^m$ with $m = \deg f > 0$ is called **monic** if $a_m = 1$.

Finally, we specify the units of our ring $k[x]$. In $\mathbb{Z}$, the units are $\{\pm 1\}$, as they are the only two integers with a multiplicative inverse. In $k[x]$, try to convince yourself of the following:

**Exercise 2.3.** The units of $k[x]$ are $(k[x])^\times = k^\times = k - \{0\}$.

Given this, we have the following theorem, analogous to unique prime factorization in $\mathbb{Z}$.

**Theorem 2.4** (analogous to Theorem 1.2)
Every nonzero $f \in k[x]$ has a factorization into irreducibles $f = c \cdot f_1 \cdots f_n$, unique up to $k^\times$.

Turns out the proof of this is almost identical to what we did last time, which is why we went through it in the first place! So our approach will be, like last time, to first prove existence of such a factorization, then prove its uniqueness. We continue expanding our analogy between $\mathbb{Z}$ and $k[x]$ in order to prove this theorem.

| Set | $\mathbb{Z}$ | $k[x]$ |
|---|---|---|
| Units | $\{\pm 1\}$ | $k^\times$ |
| Size | $\lvert n\rvert$ | $\deg f$ |

*Proof of existence.* We may assume $f$ is monic. Like with our proof last time, we can start with showing a factorization exists via induction. Any linear polynomial (i.e. $\deg f = 1$) is irreducible, so the base case is satisfied. Otherwise, suppose factorization exists for all $h$ such that $\deg h < n$. Take a degree-$n$ polynomial $f$. If $f$ is irreducible, great, we have the trivial factorization $f = f$. Otherwise, we can write $f = gg'$, where $\deg g, \deg g' < \deg f$. By the inductive hypothesis, each of $g, g'$ has a factorization into irreducibles, so the product $gg' = f$ does as well. □

Recall the proof of uniqueness in the integer scenario was solo carried by the wonderful thing called the Division Algorithm: given $a, b \in \mathbb{Z}$ $(b \neq 0)$, there exists integers $q, r \in \mathbb{Z}$ such that $a = bq + r$ and $0 \leq r < |b|$. To have the same proof apply to $k[x]$, we need an analogy of this result. Because this was pretty intuitive for the integers, we omitted a proof there, but here we will need to put in the work.

**Lemma 2.5** (Division Algorithm for Polynomials)
If $f, g \in k[x]$, $g \neq 0$, then there exists $q, r \in k[x]$ such that $f = qg + r$ and either $\deg r < \deg g$ or $r = 0$.

*Proof.* Consider the set $S \coloneqq \{f - \zeta g : \zeta \in k[x]\}$. Let $r$ be an element of $S$ with minimal degree. By definition, we can write $r = f - qg$ for some $q \in k[x]$. We have two cases for $g$.

First, suppose $g \in k^\times$. Then, we can set $q = f/g$, in which case $r = 0$. That's good.

Now, suppose deg $> 0$. We wish to show $\deg r < \deg g$. Write $r = ax^\ell + \dots$ and $g = bx^m + \dots$, so $\deg r = \ell$ and $\deg g = m$. Suppose for the sake of contradiction that $\deg r \geq \deg g$. Then, we have $r - ab^{-1}x^{d-m}g = f - (q + ab^{-1}x^{d-m})g$ so it is in $S$, but $\deg(r - ab^{-1}x^{d-m}g) < \deg g$, hence contradicting the minimality of $g$ in $S$. Thus, $\deg r < \deg g$, and we conclude. □

This stuff may look familiar to you, because it is from last lecture.

**Definition 2.6.** If $f_1, \dots, f_n \in k[x]$, denote

$$(f_1, \dots, f_n) \coloneqq \{f_1h_1 + \dots + f_nh_n : h_i \in k[x]\} \subset k[x].$$

(Again, this is the ideal generated by $f_1, \dots, f_n$.)

**Definition 2.7** (GCD)**.** If $f, g \in k[x]$, then $d \in k[x]$ is a **greatest common divisor** (gcd) for $f, g$ if

1. $d|f$ and $d|g$, and
2. if $c \in k[x]$ such that $c \mid f$ and $c \mid g$, then $c \mid d$.

How do we get the GCD? Well...

**Lemma 2.8** (analogous to Lemma 1.9)
If $f, g \in k[x]$, then $(f, g) = (d) = d \cdot k[x]$ and $d$ is a gcd for $f, g$.

*Proof.* Let $d$ be an element of minimal degree in $(f, g)$. It is clear $(d) \subseteq (f, g)$, so we now prove the reverse inclusion. If $c \in (f, g)$, then by Division Algorithm (Lemma 2.5), we can write $c = qd + r$ where $\deg r < \deg d$ or $r = 0$. Rewrite as $r = c - qd$. Because $c, d \in (f, g)$, we have $r \in (f, g)$ as well. This forces $r = 0$, because otherwise we would have $\deg r < \deg d$, contradicting the minimality of $d$. Thus, when $r = 0$, we have $c = qd$, namely $c \in (d)$. This means $(f, g) \subseteq (d)$, so equality follows. This completes the first part of this lemma.

Now we show that this choice of $d$ is a gcd of $f, g$. First, note that $(f, g) = (d)$, so in particular both $f \in (d)$ and $g \in (d)$, meaning $d \mid f$ and $d \mid g$, respectively. Let $c \in k[x]$ where $c \mid f$ and $c \mid g$. Then, $c \mid af + bg$ for any $a, b \in k[x]$. But note $d \in (f, g)$, so $d = a'f + b'g$ for some $a', b' \in k[x]$ by definition! Thus, $c \mid d$, as desired. □

We continue to make headway.

**Definition 2.9** (Relatively Prime)**.** We say $f, g \in k[x]$ are **relatively prime** if $(f, g) = (1) = k[x]$.

**Lemma 2.10** (analogous to Lemma 1.12)
If $f, g \in k[x]$ are relatively prime, and $h \in k[x]$ such that $f \mid gh$, then $f \mid h$.

*Proof.* Since $(f, g) = (1)$, we may write $1 = af + bg$ for some $a, b \in k[x]$. Then, $h = afh + bgh$. Clearly, $f \mid afh$, and $f \mid b \cdot gh$, so $f \mid h$, as desired. □

**Corollary 2.11**
If $p \in k[x]$ is irreducible and $p \mid fg$, then $p \mid f$ or $p \mid g$.

*Proof.* Let $(p, f) = (d)$, so $p = d \cdot d'$ for some $d' \in k[x]$. By definition of irreducible, either $(d) = (1)$ or $(d') = (1) \implies (d) = (p)$. For the latter, $f \in (d) = (p)$, so $p \mid f$. Otherwise, if $(d) = (1)$, then $p \mid g$ by the above lemma, and we conclude. □

**Definition 2.12** (Order)**.** For $p \in k[x]$ irreducible, $f \in k[x]$, define the **order** $\operatorname{ord}_p f$ as the largest integer $a$ such that $p^a \mid f$.

**Corollary 2.13** (analogous to Corollary 1.15)
If $f, g \in k[x]$, then $\operatorname{ord}_p(fg) = \operatorname{ord}_p f + \operatorname{ord}_p g$.

This is left as an exercise, but you should just follow the proof for what we did in $\mathbb{Z}$.

Now we are ready to complete the proof of Theorem 2.4 by proving uniqueness of factorization.

*Proof of uniqueness.* Let $f \in k[x]$. We already proved it has a factorization into irreducibles, so write $f = c \cdot f_1^{a_1} \cdots f_n^{a_n}$ where the $f_i$'s are irreducible, $a_i \in \mathbb{N}^+$, and $c \in k^\times$. We may assume that $f$ and the $f_i$'s are monic and $c = 1$. But the exponents $a_i$ are uniquely determined by the order of $p$, namely if $p = f_i$, then $a_i = \operatorname{ord}_p f$. But the order is determined solely by $f$, so the factorization is indeed only dependent on $f$, i.e. it is unique. □

Hooray! Time to celebrate.

## 2.2 Euclidean Domains

But what's so special about $\mathbb{Z}$ and $k[x]$? Surely, if this entire thread of results holds for these two rings/sets, then shouldn't we be able to apply the same reasoning to any ring given sufficient properties?

This is why we care about not just the steps of the proofs, but also *what we use*, because then we can extract exactly what information we are using, and set that to be our definition. It turns out that the specific set of properties we need for all of these proofs to translate nicely gives rise to a ring called a **Euclidean domain**. We will now define this.

First, a Euclidean domain is a type of **integral domain**. Most things in life are integral domains, so Kisin does not define this in class, but I will add a short definition.

**Definition 2.14** (Integral Domain)**.** A ring $R$ is an integral domain if $ab = 0 \implies a = 0$ or $b = 0$.

Again, most things are integral domains. For instance, $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ is an integral domain, as are $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} \mid a, b \in \mathbb{Z}\}$ and $\mathbb{Z}[e^{2\pi i/3}] = \{a + b \cdot e^{2\pi i/3} \mid a, b \in \mathbb{Z}\}$. A non-example of an integral domain is $\mathbb{Z}/6\mathbb{Z}$, since $2 \cdot 3 = 0$ but $2, 3 \neq 0$.

**Remark 2.15.** (bypassing rings) If you don't know what a ring is, think of an integral domain $R$ as a subset of $\mathbb{C}$ such that $0, 1 \in R$ and the set is closed under both addition and multiplication. This will be sufficient most of the time for this class, I think.

**Definition 2.16** (Euclidean Domain)**.** An integral domain $R$ is called **Euclidean** if there exists a map $\lambda : R \to \mathbb{Z}_{\geq 0}$ (the non-negative integers) such that for any $a, b \in R$, $b \neq 0$, there exists $q, r \in R$ such that $a = qb + r$ and either $\lambda(r) < \lambda(b)$ or $r = 0$.

Let's look at an example of a Euclidean domain.

**Claim 2.17.** $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ (here, $i = \sqrt{-1}$) is a Euclidean domain.

*Proof.* Let us take the function $\lambda(a+bi) = a^2 + b^2 = (a+bi)(a-bi)$. (This will be true even when $a, b \in \mathbb{R}$, not just in $\mathbb{Z}$.) It is clear that $\lambda$ outputs non-negative integers, as $a^2, b^2 \in \mathbb{Z}_{\geq 0}$. Furthermore, $\lambda$ is multiplicative, i.e. $\lambda((a+bi)(c+di)) = \lambda(a+bi)\lambda(c+di)$. (You can prove this by expanding everything out directly, or using basic facts about conjugation and use $\lambda(a + bi) = (a + bi)(a - bi)$.)

Now we want to show that we can construct a Division Algorithm using $\lambda$. Take $\alpha, \gamma \in \mathbb{Z}[i]$, $\gamma \neq 0$, and let $\alpha/\gamma = r + si$, where $r, s \in \mathbb{R}$. We can choose integers $m, n \in \mathbb{Z}$ such that $|r - m| \geq 1/2$ and $|s - n| \geq 1/2$ (just choose the closest integers to $r$ and $s$ respectively). Let $\delta = m + ni \in \mathbb{Z}[i]$ and $\zeta = \alpha - \gamma\delta$.

If $\zeta = 0$, then $\alpha = \gamma\delta$, which satisfies the form we want. Otherwise, we have

$$\begin{aligned}\lambda(\zeta) = \lambda(\alpha - \gamma\delta) &= \lambda(\gamma)\lambda(\alpha/\gamma - \delta) \\ &\leq 1/2\lambda(\gamma) < \lambda(\gamma)\end{aligned}$$

when $\gamma \neq 0$. The last line follows from the observation

$$\lambda(\alpha/\gamma - \delta) = \lambda((r-m) + (s-n)i) \leq 1/4 + 1/4 = 1/2.$$

The inequality $\lambda(\zeta) < \lambda(\gamma)$ follows the definition of Euclidean domain, as desired. □

# 3 9/15 - Results on Primes

Last time, we wrapped up our discussion of unique factorization into irreducibles. We continue our study of primes, as they can be thought of as the "building blocks" of everything in number theory, but with a slightly different focus.

## 3.1 Infinitely Many Primes in the Integers

The first question we'll answer is

How many primes are there in $\mathbb{Z}$?

We can answer this from a remarkably simple result given by Euclid.

**Theorem 3.1** (Euclid)
There are infinitely many primes.

*Proof.* Suppose there are only finitely many primes, so $\{p_1, \ldots, p_n\}$ is our complete list of distinct primes. Consider $z = p_1p_2 \cdots p_n + 1$. If any $p_i \mid z$, then $p_i \mid z - p_1p_2 \cdots p_n = 1$, so $p_i \nmid z$. But we know $z$ has a prime factorization, so there must exist a prime dividing $z$ but not contained in our list $\{p_1, \ldots, p_n\}$. This contradicts our assumption of finitely many primes. □

There are many proofs of this result, but this was the first one, and I find it particularly nice because of its simplicity. Even better, one can use this technique to prove even stronger statements. For instance, I challenge you to prove:

**Exercise 3.2.** Prove there are infinitely many primes congruent to 1 mod 4. (You can also prove this for 3 mod 4.)

The fact about infinitely many primes in $\mathbb{Z}$ also comes as a consequence of the following result:

**Theorem 3.3**

The sum $\sum_{p \text{ prime}} \frac{1}{p}$ diverges.

Before we prove this, we will prove the following lemma:

**Lemma 3.4**

If $s \in \mathbb{N}^+$, then the sum $\sum_{n=1}^{\infty} \frac{1}{n^s} = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$ diverges if $s = 1$ and converges if $s \geq 2$.

*Proof.* We invoke some stuff from high school calculus. The sum is the left Riemann sum for $\int_1^\infty \frac{1}{(x+1)^s}\,dx$ and the right Riemann sum for $\int_1^\infty \frac{1}{x^s}\,dx$, so the sum is bounded by

$$\int_1^\infty \frac{1}{(x+1)^s}\,dx \leq \sum_{n=1}^{\infty} \frac{1}{n^s} \leq \int_1^\infty \frac{1}{x^s}\,dx.$$

The left integral when $s = 1$ is $\lim_{z\to\infty} \int_1^z (x+1)^{-1}\,dx = \lim_{z\to\infty} \log(x+1)\big|_1^z \to +\infty$, so the sum diverges for $s = 1$. For $s \geq 2$, the right integral converges: $\int_1^z x^{-s}\,dx = -\frac{1}{s-1}\,\frac{1}{x^{s-1}}\Big|_1^z = \frac{1}{s-1}\left(1 - \frac{1}{z^{s-1}}\right)$ which is bounded as $z \to \infty$. □

Now we prove Theorem 3.3.

*Proof.* Let $p_1, \ldots, p_{\ell(n)}$ be all primes $\leq n$. Then, consider the product

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

Each $\left(1 - \frac{1}{p_i}\right)^{-1}$ is the sum of a geometric series in disguise (with initial term 1 and common ratio $\frac{1}{p_i}$), so we can write

$$\begin{aligned}
\lambda(n) &= \prod_{i=1}^{\ell(n)} \left(1 - \frac{1}{p_i}\right)^{-1} \\
&= \prod_{i=1}^{\ell(n)} \left(1 + \frac{1}{p_i} + \frac{1}{p_i^2} + \cdots\right) \\
&= \left(1 + \frac{1}{p_1} + \frac{1}{p_1^2} + \cdots\right)\left(1 + \frac{1}{p_2} + \frac{1}{p_2^2} + \cdots\right)\cdots.
\end{aligned}$$

Clearly, the prime factorization of any integer $m \leq n$ uses only the primes $\leq n$, i.e. the primes in $\{p_1, \ldots, p_{\ell(n)}\}$. Thus, in the expansion of this product, $\frac{1}{m}$ appears for all $m \leq n$. In particular,

$$\lambda(n) \geq 1 + \frac{1}{2} + \cdots + \frac{1}{n},$$

which we know from Lemma 3.4 diverges as $n \to \infty$.

Now we do some clever manipulations. Using the Taylor series expansion of log, which recall is

$$\log(1-x) = -\sum_{m=1}^{\infty} \frac{x^m}{m},$$

we can now write

$$\begin{aligned}
\log \lambda(n) &= -\sum_{i=1}^{\ell(n)} \log\left(1 - \frac{1}{p_i}\right) \\
&= \sum_{i=1}^{\ell(n)} \sum_{m=1}^{\infty} (mp_i^m)^{-1} \\
&= \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_{\ell(n)}} + \sum_{i=1}^{\ell(n)} \sum_{m=2}^{\infty} (mp_i^m)^{-1}.
\end{aligned}$$

Call the remaining double sum at the end as $S$. We will now show $S$ converges.

$$\begin{aligned}
\sum_{m=2}^{\infty} (mp_i^m)^{-1} < \sum_{m=2}^{\infty} (p_i^m)^{-1} &= p_i^{-2} + p_i^{-3} + \cdots \\
&= \frac{p_i^{-2}}{1 - p_i^{-1}} \leq 2p_i^{-2} \\
\implies S &= \sum_{i=1}^{\ell(n)} \sum_{m=2}^{\infty} (mp_i^m)^{-1} \\
&< \sum_{i=1}^{\ell(n)} 2p_i^{-2} < 2\sum_{n=1}^{\infty} \frac{1}{n^2},
\end{aligned}$$

which we know converges by Lemma 3.4.

Now, we are equipped to complete the proof. We know

$$\log \lambda(n) = \frac{1}{p_1} + \frac{1}{p_2} + \cdots + \frac{1}{p_{\ell(n)}} + S,$$

and as $n \to \infty$, $S$ converges (aka it is bounded) but $\log \lambda(n) \to \infty$, so the sum $\frac{1}{p_1} + \cdots + \frac{1}{p_{\ell(n)}} \to \infty$, i.e. the sum $\sum_{p \text{ prime}} \frac{1}{p}$ diverges, as desired. $\square$

## 3.2 Infinitely Many Primes for Polynomials

Just like what we did in the first two days, we will try to extend this question to polynomial rings. So we will pose the question:

> Are there infinitely many primes in $k[x]$, where $k$ is a field?

We're going to revise this question a little bit. Note that if $k$ is an infinite field, say $k = \mathbb{Q}$ or $\mathbb{C}$, then I could take any irreducible polynomial of the form $ax + b$, and there are infinitely many choices for $(a, b)$, so the question is obvious. Thus, we will focus on when $k$ is a **finite** field.

**Example 3.5** (Finite Field)

Perhaps the simplest example of a finite field is the integers modulo a prime. For instance, $p = 5$ is prime, and $\mathbb{Z}/5\mathbb{Z} = \{0, 1, 2, 3, 4\}$, under addition and multiplication modulo 5, is a field.

We actually know how to classify all finite fields. This is not straightforward to prove, but we will provide the fact:

**Fact 3.6.** A finite field has $q = p^r$ elements for some prime $p$ and $r \in \mathbb{N}$. (We denote such a field as $\mathbb{F}_q$.)

Now, we can provide an analogous statement to Theorem 3.3 for $k[x]$. This highlights the usefulness of this theorem, because then we can conclude the infinitude of primes for other rings beyond just the integers.

**Theorem 3.7**

If $|k| = q$, then the sum

$$\sum_{\substack{p(x) \in k[x] \\ p(x) \text{ irreducible}}} q^{-\deg p(x)} = \sum_{p \text{ prime}} \frac{1}{q^{\deg p(x)}}$$

diverges.

The proof of this is very similar in flavor to the proof we provided for Theorem 3.3.

Someone asked about the density of primes, so we'll provide a brief interlude here to address this question.

**Definition 3.8.** If $x \in \mathbb{R}^+$, then let $\pi(x)$ be the number of primes $p$ such that $1 < p \leq x$.

We have a remarkable theorem, so important, it is called the Prime Number Theorem.

**Theorem 3.9** (Prime Number Theorem)
$\pi(x) \sim \frac{x}{\log x}$, i.e.

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

This was conjectured by Gauss at the (impressive) age of 16.[2] It was proven as a theorem by Hadamard and Poussin in 1896 using the Riemann zeta function. In particular, proving strong bounds on the error term for the Prime Number Theorem requires assuming the Riemann Hypothesis, just one of many reasons why the elusive conjecture is so important in math.

We won't prove the Prime Number Theorem, as it is very involved, but we will prove a pretty strong, related result:

**Theorem 3.10**
There exists constants $c_1, c_2 > 0$ such that

$$c_2 \cdot \frac{x}{\log x} < \pi(x) < c_1 \cdot \frac{x}{\log x}.$$

**Definition 3.11.** For $x \in \mathbb{R}^+$, let $\theta(x) = \sum_{p\leq x} \log p$.

We can provide a pretty nice bound for $\theta(x)$.

**Proposition 3.12**
For $x \in \mathbb{R}^+$, $\theta(x) < (4 \log 2) \cdot x$.

*Proof.* Consider the binomial coefficient $\binom{2n}{n} = \frac{(2n!)}{n!n!}$. (If you have never seen this before, this is the number of ways to choose $n$ objects from $2n$ total objects.) We see that this is divisible by all primes $p$ with $n + 1 \leq p \leq 2n$. (They appear in the numerator, but

[2] If you ever feel yourself having too big of an ego, just think of Gauss. On the other hand, if you feel yourself having pretty low self-esteem, just think of the time when Grothendieck needed to use a particular prime in a lecture and said, "Alright, take 57," which became so famous that 57 is now called the Grothendieck prime.

not in the denominator.) Also note $\binom{2n}{n}$ appears in the binomial expansion of $(1+1)^{2n}$, so we have

$$2^{2n} = (1+1)^{2n} > \binom{2n}{n} > \prod_{\substack{n<p\leq 2n \\ p \text{ prime}}} p.$$

Taking the log on both sides, we have

$$2n \log 2 > \sum_{\substack{n<p\leq 2n \\ p \text{ prime}}} \log p = \theta(2n) - \theta(n). \tag{1}$$

Because $\theta(1) = 0$, we can write

$$\theta(2^m) = \theta(2^m) - \theta(2^0) = \sum_{i=1}^{m} \theta(2^i) - \theta(2^{i-1}).$$

Using Equation 1, we have

$$\begin{aligned} \theta(2^m) &= \sum_{i=1}^{m} \theta(2^i) - \theta(2^{i-1}) \\ &< \log 2(2^m + 2^{m-1} + \cdots + 2) \\ &< (\log 2) \cdot 2^{m+1}. \end{aligned}$$

Thus, for any $x \in \mathbb{R}^+$, we can find $m \in \mathbb{N}$ such that $2^{m-1} \leq x \leq 2^m$, which gives us

$$\theta(x) \leq \theta(2^m) < (\log 2) \cdot 2^{m+1} \leq 4(\log 2) \cdot x,$$

as desired. □

# 4 9/18 - Proving Weaker Version of Prime Number Theorem

Let's pick off from last time. We wanted to prove a weaker version of the Prime Number Theorem, given by Theorem 3.10, and to do this we had a nice result bounding $\theta(x) = \sum_{p\leq x} \log p$ (Proposition 3.12). We proved the proposition at the end of the last lecture.

Well, Proposition 3.12 gives us an upper bound related to the primes less than a given $x$, so with a little bit more work, we can determine a constant $c_1 > 0$ such that $\pi(x) < c_1 \cdot \frac{x}{\log x}$.

## 4.1 Upper Bound on $\pi(x)$

**Corollary 4.1**

There exists a constant $c_1 > 0$ such that $\pi(x) < c_1 \cdot \frac{x}{\log x}$ for $x \geq 2$.

*Proof.* We have

$$\theta(x) = \sum_{1 \leq p \leq x} \log p, \qquad \pi(x) = \sum_{1 \leq p \leq x} 1.$$

How can we relate $\pi(x)$ with $\theta(x)$? One (less fruitful) observation one could make is $\log p \leq \log x$, so $\theta(x) \leq \pi(x) \log(x)$. But we will do something a bit more useful, because after all, we want an upper bound for $\pi(x)$, not a lower one.

We will consider the sum $\sum_{x \geq p \geq \sqrt{x}} \log p$. This may seem a bit out of the blue, but $\log \sqrt{x} = \frac{1}{2} \log x$, so we are just taking the sum from the logarithmically top half of the range $1 \leq p \leq x$. We now have a really nice lower bound for $\theta(x)$ (equivalently, an upper bound for $\pi(x)$):

$$\begin{aligned}
\theta(x) &\geq \sum_{x \geq p \geq \sqrt{x}} \log p \\
&\geq \log \sqrt{x} \cdot \pi(x) - \log \sqrt{x} \cdot \pi(\sqrt{x}) \\
&\geq \log \sqrt{x} \cdot \pi(x) - \sqrt{x} \log \sqrt{x} \\
&= \frac{1}{2} (\log x) \cdot (\pi(x) - \sqrt{x}). \\
\implies \pi(x) &\leq \frac{2\theta(x)}{\log x} + \sqrt{x} \\
&< 8 \log 2 \frac{x}{\log x} + \sqrt{x}.
\end{aligned}$$

So this is basically what we want. To make the $\sqrt{x}$ extra term disappear, observe that $\sqrt{x}$ grows slower than $\frac{x}{\log x}$ (more specifically, one can show $\sqrt{x} < \frac{2x}{\log x}$ for $x \geq 2$), so

$$\pi(x) \leq 8 \log 2 \frac{x}{\log x} + \sqrt{x} < (8 \log 2 + 2) \frac{x}{\log x},$$

so we can take $c_1 = 2 + 8 \log 2$. □

To recap on the above proof, because it is a bit technical: we have this wonderful result from Proposition 3.12, and we want to turn this into an upper bound for $\pi(x)$. This means we have to write $\theta(x)$ as an upper bound of some expression in terms of $\pi(x)$. The trick we employ is to consider the sum that only takes the "top half" of the range $1 \leq p \leq x$, and the magic happens in the computations.

## 4.2 Lower Bound on $\pi(x)$

We are left with constructing a constant $c_2$ for the lower bound, which we address now.

**Proposition 4.2**
There exists a constant $c_2 > 0$ such that $\pi(x) > c_2 \cdot \frac{x}{\log x}$.

*Proof.* We have the nice (combinatorial) result[3]

$$\operatorname{ord}_p n! = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots .$$

To see why this is the case: $\left\lfloor \frac{n}{p^i} \right\rfloor$ counts the number of multiples of $p^i$ at most $n$. Let's consider the sum of the first two terms. All numbers only divisible by $p$ (and not $p^2$) are counted once from the first term, but the multiples of $p^2$ are counted twice, once from the first term and once from the second. In general, the sum will count all multiples of $p^i$ exactly $i$ times, which by definition is what we want for the order.[4]

Now let us consider $\binom{2n}{n}$. We can compute

$$\begin{aligned} \operatorname{ord}_p \binom{2n}{n} &= \operatorname{ord}_p \frac{(2n)!}{n!n!} \\ &= \sum_{j=1}^{t_p} \left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor, \end{aligned}$$

where $t_p$ is the largest integer such that $p^{t_p} \le 2n$, i.e., $t_p = \left\lfloor \frac{\log 2n}{\log p} \right\rfloor$. Let $\left\{ \frac{a}{b} \right\}$ be the fractional part of $\frac{a}{b}$, so $\left\{ \frac{5}{3} \right\} = \frac{2}{3}$. Then, we have

$$\left\lfloor \frac{2n}{p^j} \right\rfloor - 2 \left\lfloor \frac{n}{p^j} \right\rfloor = \begin{cases} 1 & \text{iff } \left\{ \frac{n}{p^j} \right\} \ge \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we have

$$\begin{aligned} 2^n \le \binom{2n}{n} &\le \prod_{p \le 2n} p^{t_p} \\ \implies n \log 2 \le \sum_{p \le 2n} t_p \log p &= \sum_{p \le 2n} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p \\ &= \sum_{p \le \sqrt{2n}} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p + \sum_{\sqrt{2n} < p \le 2n} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p. \end{aligned}$$

[3] Kisin: "I always remember this formula, because it was needed in an olympiad problem while I was in high school, but I had never seen it before, so I just proved it on the spot." Absolute chad.

[4] If this still concerns you, do this explicitly for a specific prime $p$ and integer $n$. Doing examples will help!

We now consider the latter sum separately. If $p > \sqrt{2n}$, then

$$\frac{\log 2n}{\log p} < \frac{\log 2n}{\log \sqrt{2n}} = \frac{\log 2n}{\frac{1}{2}\log 2n} = 2,$$

so $\left\lfloor \frac{\log 2n}{\log p} \right\rfloor = 1$. Substituting this back into our inequality above, we have

$$n \log 2 \leq \sum_{p \leq \sqrt{2n}} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p + \sum_{\sqrt{2n} < p \leq 2n} \log p.$$

Noting that $\lfloor a/b \rfloor \cdot b \leq a$, we can extend the inequality to

$$\begin{aligned} n \log 2 &\leq \sum_{p \leq \sqrt{2n}} \left\lfloor \frac{\log 2n}{\log p} \right\rfloor \log p + \sum_{\sqrt{2n} < p \leq 2n} \log p \\ &< \sqrt{2n} \log(2n) + \theta(2n) \\ \implies \theta(2n) &\geq n \log 2 - \sqrt{2n} \log(2n). \end{aligned}$$

Like in the above proof where we used $\sqrt{x}$ grows slower than $\frac{x}{\log x}$ (specifically, $\sqrt{x} < \frac{2x}{\log x}$), rearranging tells us that $\sqrt{2n}\log(2n)$ grows slower than $n$, i.e. $\frac{\sqrt{2n}\log(2n)}{n} \to 0$ as $n \to \infty$. Thus, the right side of the above inequality is dominated by $n \log 2$, an in particular we have $\theta(2n) > T \cdot n$ for some $T > 0$ when $n >> 0$. If $2n \leq x < 2n - 2$, then

$$\theta(x) \geq \theta(2n) \geq T \cdot n > T \cdot \frac{x-1}{2} > c_2 \cdot x$$

for some $c_2 > 0$ and for all $x \geq 2$. Thus, $c_2 \cdot x < \theta(x) \leq \pi(x) \cdot \log x$, as desired. □

Whew, okay that was a very involved proof with lots of big steps. Let's get back to ground level and deal with things a bit less stressful.

## 4.3 Modular Congruence

**Definition 4.3** (Modular Congruence)**.** Let $n \in \mathbb{N}^+$. If $a, b \in \mathbb{Z}$, we say $a \equiv b (n)$ if $n \mid a - b$.

Sometimes, when $n$ is understood, we will suppress the $(n)$ or the $\bmod n$ and just write $a \equiv b$.

This is what we call an equivalence relation. It is especially nice because it complies with all the algebraic operations we'd want in life: if $a \equiv b$ and $c \equiv d$, then $a + c \equiv b + d$ and $a \cdot c \equiv b \cdot d$. (The latter can be seen via $ac - bd = c(a - b) + b(c - d)$.) You are probably familiar with all of this already from the integers modulo $n$, denoted $\mathbb{Z}/n\mathbb{Z}$.

**Lemma 4.4**
If $(a, n) = (1)$, then $\exists\, x \in \mathbb{Z}$ such that $ax \equiv 1 \pmod{n}$.

*Proof.* This follows very quickly from definitions. If $(a, n) = (1) = \mathbb{Z}$, then $1 \in (a, n)$, so 1 can be expressed as a linear combination of $a$ and $n$. In other words, $\exists\, x, y \in \mathbb{Z}$ such that $ax + ny = 1$. Equivalently, this means $ax \equiv 1 \pmod{n}$, as desired. □

One can think, therefore, of $a$ such that $(a, n) = 1$ as an invertible element modulo $n$. In $\mathbb{Z}/n\mathbb{Z}$, the element corresponding to $a$ has a multiplicative inverse.

**Definition 4.5** (Units of $\mathbb{Z}/n\mathbb{Z}$)**.** The units of $\mathbb{Z}/n\mathbb{Z}$, denoted $(\mathbb{Z}/n\mathbb{Z})^\times$, are the congruence class of $a$ for any $(a, n) = 1$.

**Example 4.6** (Units of $\mathbb{Z}/p\mathbb{Z}$)
Let $p$ be a prime. Then, every nonzero element of $\mathbb{Z}/p\mathbb{Z}$ is a unit: if $p \nmid a$, then $(a, p) = 1$, so the congruence class of $a$ is a unit in $\mathbb{Z}/n\mathbb{Z}$. In particular, $(\mathbb{Z}/p\mathbb{Z})^\times = (\mathbb{Z}/p\mathbb{Z}) \setminus \{0\}$. (By definition, this means $\mathbb{Z}/p\mathbb{Z}$ is a field. Additionally, $(\mathbb{Z}/p\mathbb{Z})^\times$ is a (multiplicative) group, because the product of any two nonzero elements ($p \nmid a, p \nmid b$) stays nonzero ($p \nmid ab$).)

A result we will prove next time is Fermat's Little Theorem, which states that if $x \in (\mathbb{Z}/p\mathbb{Z})^\times$, then $x^{p-1} \equiv 1 \pmod{p}$.

A natural question one might ask is how we can count the number of units of $\mathbb{Z}/n\mathbb{Z}$, or equivalently count the number of $1 \le a < n$ such that $(a, n) = 1$. This is an important quantity in number theory, so it has a specific function related to it, called the **Euler totient function** $\varphi$. For an integer $n$, $\varphi(n) = |(\mathbb{Z}/n\mathbb{Z})^\times| = \{1 \le a < n : (a, n) = 1\}$. We have a nice way of counting this:

**Lemma 4.7**
If $n = p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r}$, then

$$\phi(n) = n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_r}\right) = \prod_{i=1}^{r} p_i^{a_i - 1}(p_i - 1).$$

This gives rise to a more general version of Fermat's Little Theorem (called Euler's Totient Theorem), which we will state and prove next time.

# 5 9/22 - Euler's Totient

## 5.1 Proving Euler's Totient Formula

We will start off today with proving Lemma 4.7 above. We will prove two proofs, one more direct, and the other one using a nifty trick called Möbius inversion, which is used often in number theory.

*Proof 1.* We wish to compute the size of the set $\{1 \leq m \leq n \mid (m,n) = 1\}$. This is now just a counting problem. We first start with all integers from 1 to $n$ – there are $n$ of them. Now, we subtract all multiples of $p_i$ for each $p_i$. This leaves us with

$$n - \frac{n}{p_1} - \frac{n}{p_2} - \cdots - \frac{n}{p_r}.$$

But we have subtracted too much! For instance, we remove the number $p_1p_2$ twice: once for the multiples of $p_1$, and the other for $p_2$. Thus, we must add back in all multiples of $p_ip_j$ for $i \neq j$. This gives us now

$$n - \frac{n}{p_1} - \cdots - \frac{n}{p_r} + \frac{n}{p_1p_2} + \cdots + \frac{n}{p_{r-1}p_r}.$$

Continuing this process of correcting for under/overcounting, we eventually get

$$\begin{aligned} n - \frac{n}{p_1} - \cdots - \frac{n}{p_r} + \frac{n}{p_1p_2} + \cdots + \frac{n}{p_{r-1}p_r} + \frac{n}{p_1p_2p_3} + \cdots \\ = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_r}\right), \end{aligned}$$

as desired. □

**Remark 5.1.** Note that this is not actually a proof, this is more of a general argument. There needs to be some work to formalize this. For those familiar with some combinatorics, this is essentially invoking the Principle of Inclusion-Exclusion.

## 5.2 Möbius Inversion

The next proof will be a lot more formal. First, we will introduce a new operation on two functions on the integers. This operation is sometimes called (Dirichlet) **convolution**.

**Definition 5.2** (Convolution)**.** Let $f, g : \mathbb{N} \to \mathbb{C}$ be two functions. We define the

convolution $f * g$ of $f$ and $g$ as

$$(f * g)(n) = \sum_{\substack{d_1 d_2 = n \\ d_1, d_2 \in \mathbb{N}}} f(d_1) g(d_2).$$

One can show that this operation is both associative and commutative (just write it out, it's not bad at all).

We introduce three particular functions.

1. Consider the function $I : \mathbb{N} \to \mathbb{C}$ where $I(n) = 1$ for all $n \in \mathbb{N}$. Then, for any $f : \mathbb{N} \to \mathbb{C}$, convolution gives
$$(f * I)(n) = (I * f)(n) = \sum_{d|n} f(d).$$
2. Let $\varphi : \mathbb{N} \to \mathbb{C}$ send $\varphi(1) = 1$ and $\varphi(n) = 0$ for all $n \neq 1$. Then, for any $f : \mathbb{N} \to \mathbb{C}$, we have $f * \varphi = f$.
3. The last is called the **Möbius function**, denoted $\mu$, and it is defined by
$$\mu(n) = \begin{cases} (-1)^s & n = p_1 \cdots p_s, \text{ each } p_i \text{ distinct} \\ 0 & \text{otherwise.} \end{cases}$$
Some examples: $\mu(p) = -1$ for any prime $p$, $\mu(10) = (-1)^2 = 1$, and $\mu(28) = 0$ since it has two factors of 2. In particular, if $n$ is divisible by a perfect square, then $\mu(n) = 0$.

Considering that we want to use a technique called Möbius inversion, it makes sense that the Möbius function will be an object of interest. We begin with one nice property of it:

**Lemma 5.3**

If $n > 1$, then $\sum_{d|n} \mu(d) = 0$.

*Proof.* Let $n = p_1^{a_1} \cdots p_s^{a_s}$, where the $p_i$'s are distinct. Then,

$$\begin{aligned} \sum_{d|n} \mu(d) &= \sum_{0 \leq b_i \leq a_i} \mu(p_1^{b_1} \cdots p_s^{b_s}) \\ &= \sum_{b_i \in \{0,1\}} \mu(p_1^{b_1} \cdots p_s^{b_s}) \\ &= 1 - s + \binom{s}{2} - \binom{s}{3} + \cdots + (-1)^s \binom{s}{s} \\ &= (1-1)^s = 0. \end{aligned}$$

Here, the second equality follows from the fact that any integer divisible by a square evaluates to 0, the third line follows from simply counting how many tuples $(b_1, \ldots, b_s)$ have $i$ 1's for $0 \le i \le s$, and the last line follows from the Binomial Theorem. □

Now, we introduce **Möbius Inversion**.

**Theorem 5.4** (Möbius Inversion)
If $f : \mathbb{N} \to \mathbb{C}$ and $F = f * I$, i.e., $F(n) = \sum_{d|n} f(d)$, then

$$f = F * \mu, \qquad \text{i.e., } f(n) = \sum_{d|n} F(d)\mu(n/d).$$

*Proof.* The statement may come as a surprise, but the proof is actually not bad at all. We will use all three functions $I, \varphi, \mu$ that we had before. First, note that

$$(\mu * I)(n) = \sum_{d|n} \mu(d)I(n/d) = \sum_{d|n} \mu(d),$$

which by the previous lemma is 0 for $n > 1$. One can easily compute it is 1 for $n = 1$, so in fact $\mu * I = \varphi$. In this case, we have

$$\begin{aligned} F &= f * I \\ \implies F * \mu &= f * I * \mu \\ &= f * \mu * I \\ &= f * \varphi = f, \end{aligned}$$

as desired. □

Recall we want to prove Lemma 4.7 in a different way using Möbius inversion. Here is the first step towards this proof, which gives the flavor of being related to $\mu$.

**Proposition 5.5**
$\sum_{d|n} \phi(d) = n$. In other words, $\phi * I$ is the identity map $\mathrm{id} : n \mapsto n$.

*Proof.* Consider the fractions $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}$. Write all of them in lowest terms. Given any $d \mid n$, there must be exactly $\phi(d)$ fractions with denominator $d$. (For example, if $n = 15$, then the fractions with denominator 3 are $5/15 = 1/3$ and $10/15 = 2/3$. Indeed, $\phi(3) = 3 - 1 = 2$.) But there are $n$ fractions in total, so $\sum_{d|n} \phi(d) = n$ follows as a counting argument. □

Now, Lemma 4.7 follows as a consequence of Proposition 5.5.

**Corollary 5.6**
If $n = p_1^{a_1} \cdots p_s^{a_s}$, then $\phi(n) = n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_s}\right)$.

*Proof.* Proposition 5.5 gives $\phi * I = \mathrm{id}$. Möbius Inversion now tells us $\phi = \mathrm{id} * \mu$, so

$$\begin{aligned}
\phi(n) &= \sum_{d|n} \mu(d)\,\mathrm{id}(n/d) \\
&= \sum_{d|n} \mu(d)\frac{n}{d} \\
&= n - \frac{n}{p_1} - \cdots - \frac{n}{p_s} + \frac{n}{p_1 p_2} + \cdots \\
&= n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_s}\right)
\end{aligned}$$

as we had before, so we conclude. □

## 5.3 Euler's Totient Theorem

Disregarding this explicit formula for $\phi(n)$, we can provide some nice results involving $\phi(n)$.

**Theorem 5.7** (Euler's Totient Theorem)
If $h \in (\mathbb{Z}/n\mathbb{Z})^\times$, then $h^{\phi(n)} \equiv 1 \pmod{n}$.

This is actually a consequence of a more general fact in group theory called Lagrange's Theorem. If you know some group theory, you can replace $(\mathbb{Z}/n\mathbb{Z})^\times$ as any finite group $G$, $h$ with some element of $G$, and $\phi(n)$ with the size $|G|$ of $G$, and the statement would still be true as equality in the group.

Actually, it looks like Kisin wants to talk about this result in its group-theoretic generality, so we will talk a little bit about groups. For our purposes, we will consider a **subgroup** of $(\mathbb{Z}/n\mathbb{Z})^\times$ as a subset $H \subseteq (\mathbb{Z}/n\mathbb{Z})^\times$ such that $1 \in H$ and for any $h_1, h_2 \in H$, their product $h_1 h_2 \in H$ is also in $H$.

A few more definitions, unfortunately all called order. (This is not to be confused with the more strictly number-theoretic definition of order we provided way back in Definition 1.14.)

**Definition 5.8** (Order of an *element*)**.** Let $h \in H$. The smallest $a \in \mathbb{N}^+$ such that $h^a = 1$ is called the **order of** $h$.

If we consider all the powers of $h$ in a set $\{1, h, h^2, \dots, h^i, \dots\}$, then if $H$ is finite, the set must start repeating elements at some point. (The set is contained in $H$, so it can only contain finitely different elements.) The order is just the smallest exponent at which the set begins to repeat itself.

**Definition 5.9** (Order of a *subgroup*)**.** If $H \subseteq G$ is a subgroup (here, $G = (\mathbb{Z}/n\mathbb{Z})^\times$), then $|H|$ is called the **order of** $H$.

One can see that these two definitions of order are related. Consider the subgroup generated by an element $h \in H$, that is, the set of all powers of $h$. We will notate as $\langle h \rangle = \{1, h, h^2, \dots\}$. It follows that the order of $h$ agrees with the order of $\langle h \rangle$.

Now we prove Theorem 5.7 via the following Proposition.

**Proposition 5.10**
$|H|$ divides $\phi(n) = |(\mathbb{Z}/n\mathbb{Z})^\times|$.

*Proof.* This proof might seem a little weird; I think this is because Kisin is trying to explain a proof in group theory without defining new terminology. For the more purely group theory explanation, see the last paragraph.

We can define an equivalence relation between elements in $(\mathbb{Z}/n\mathbb{Z})^\times$ as follows. If $g_1, g_2 \in (\mathbb{Z}/n\mathbb{Z})^\times$, then we say $g_1 \sim g_2$ if there exists some $h \in H$ such that $g_1 \sim g_2 h$. For example, if we take $h = -1$, then $a \sim -a$ for any $a \in (\mathbb{Z}/n\mathbb{Z})^\times$. Another example: all elements of $H$ are equivalent to each other.

Any equivalence relation produces equivalence classes. We can consider the equivalence class of some $a \in (\mathbb{Z}/n\mathbb{Z})^\times$; this is given by the set $\{ah \mid h \in H\}$. Each element in this set is distinct (if $ah_1 = ah_2$, then $h_1 = h_2$), so this equivalence class has exactly $|H|$ elements. Note that if $b$ is in this equivalence class, i.e., $b = ah'$ for some $h' \in H$, then the equivalence class of $b$ is the same as the equivalence class of $a$. (I will leave this as an exercise, but reach out if you have questions.)

Now the finish line is in sight. Each element of $(\mathbb{Z}/n\mathbb{Z})^\times$ belongs to an equivalence class (namely, its own), and each equivalence class has size $|H|$. Therefore, $\phi(n) = |(\mathbb{Z}/n\mathbb{Z})^\times|$ is equal to $|H|$ times the number of equivalence classes. The latter number is clearly an integer, so it follows that $|H|$ divides $\phi(n)$, as desired.

For people familiar with group theory, we are simply considering the cosets of $H$ in $(\mathbb{Z}/n\mathbb{Z})^\times$. Each coset has size $|H|$, the number of cosets is clearly an integer, and every element in $(\mathbb{Z}/n\mathbb{Z})^\times$ is contained in a coset of $H$. It follows that $|H|$ times the number of cosets is $\phi(n)$. □

Now the proof of Theorem 5.7 comes easily.

*Proof of Theorem 5.7.* Let $a$ be the order of $h$, so $h^a = 1$. Then, $H = \{1, h, \dots, h^{a-1}\}$

and $a = |H|$ divides $\phi(n)$ by the above Proposition. Thus, $h^{\phi(n)} = (h^a)^{\phi(n)/a} = 1$, done. □

Restricting our attention to when $n = p$ is prime, we get Fermat's Little Theorem.

**Corollary 5.11** (Fermat's Little Theorem)
If $p$ is prime, then $h^{p-1} \equiv 1 \pmod{p}$.

*Proof.* Directly follows from Euler's Totient Theorem using $\phi(p) = p - 1$. □

Note that what Fermat's Little Theorem is telling us, in terms of orders, is that the order of any element in $(\mathbb{Z}/p\mathbb{Z})^\times$ divides $p - 1$. But it is a neat fact about $(\mathbb{Z}/p\mathbb{Z})^\times$ that there actually exists an element whose order is exactly $p - 1$:

**Theorem 5.12**
If $p$ is prime, then there exists some $h \in (\mathbb{Z}/p\mathbb{Z})^\times$ such that $h$ has order $p - 1$. In other words, $(\mathbb{Z}/p\mathbb{Z})^\times = \{1, h, h^2, \ldots, h^{p-2}\}$ has one generator, so it is cyclic.

**Definition 5.13** (Primitive Root)**.** An element $h \in (\mathbb{Z}/p\mathbb{Z})^\times$ with order $p-1$ (i.e., a generator of the subgroup of units) is a **primitive root** mod $p$.

**Example 5.14** (Primitive Root)
Let's identify primitive roots mod $p$ for small values of $p$. For $p = 3$, we have $(\mathbb{Z}/3\mathbb{Z})^\times = \{1, 2\}$. The element 2 generates the set, since $2^1 = 2$ and $2^2 = 4 = 1$. For $p = 5$, one can see 2 is also a primitive root; 2 is also a primitive root for $p = 11$. (I promise 2 is not always a primitive root; for instance, in $p = 7$, the powers of 2 are $2^1 = 2, 2^2 = 4, 2^3 = 8 = 1$, so it does not cover everything in $(\mathbb{Z}/7\mathbb{Z})^\times$.)

A natural question one may ask after seeing the above examples is whether 2 is a primitive root for infinitely many primes. This is actually an unsolved problem! There is a conjecture by Artin, though, which claims that if $a \neq -1$ is not a square, then $a$ is a primitive root mod $p$ for infinitely many $p$. So we expect for 2 to be a primitive root for infinitely many primes.

# 6 09/25 - Unit Groups

## 6.1 Proving Existence of Primitive Root

Theorem 5.12 is quite remarkable, as it gives $(\mathbb{Z}/p\mathbb{Z})^\times$ perhaps the nicest structure possible. We will work towards proving this amazing fact.

**Lemma 6.1**

If $k$ is a field (e.g., $\mathbb{Z}/p\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$) and $f(x) \in k[x]$ is a monic polynomial of degree $n$, then $f(x) = 0$ has at most $n$ solutions in $k$.

*Proof.* We will induct on $n = \deg f$. If $n = 1$, then $f(x)$ is of the form $f(x) = x - a$ for some $a \in k$. Clearly, the only root is $x = a$, so there is exactly one solution.

Before we move on to the inductive step, we will develop a useful condition for divisibility. Suppose $f(x)$ is a polynomial and $\alpha \in k$ such that $f(\alpha) = 0$. Then, since $k[x]$ has a Division Algorithm (recall Lemma 2.5), we can write $f(x) = (x-\alpha)q(x)+r(x)$, where $\deg r < \deg(x-\alpha) = 1$. This forces $\deg r = 0$, i.e., $r(x) = c$ is a constant function. But then $0 = f(\alpha) = r(\alpha) = c$, which means $(x - \alpha) \mid f(x)$.

Now we proceed with the inductive step. Suppose the statement is true for polynomials of degree $n - 1$, and let $\deg f = n$. If $f$ has no roots, then the Lemma is clearly satisfied. Otherwise, choose some $\alpha$ such that $f(\alpha) = 0$. By the above, $(x - \alpha) \mid f(x)$, so $f(x) = (x - \alpha) \cdot f_1(x)$. But now $\deg f_1$, so our inductive hypothesis tells us that $f_1$ has at most $n - 1$ roots. The conclusion follows. □

**Exercise 6.2.** (For fun) Out of the fields $\mathbb{Z}/p\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, can you find which ones produce *exactly* $n$ roots for a degree $n$ polynomial? (Such fields are called **algebraically closed**, and they are useful because, well, we can always factor a polynomial into linear factors.)

Note that the above lemma tells us that the polynomial $x^{p-1} - 1 = 0$ has at most $p - 1$ roots, but Fermat's Little Theorem (5.11) tells us that it actually has exactly $p - 1$ distinct roots. We can strengthen this observation:

**Corollary 6.3**

If $d \mid p - 1$, then $x^d = 1 \pmod p$ has exactly $d$ solutions in $\mathbb{Z}/p\mathbb{Z}$.

*Proof.* Fermat's Little Theorem tells us that every element $a \in (\mathbb{Z}/p\mathbb{Z})^\times$ satisfies $a^{p-1} = 1$, so we can factor

$$x^{p-1} - 1 = \prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} (x - a).$$

(We can do this because from the proof above, each $(x - a) \mid x^{p-1} - 1$, and as the $(x - a)$ linear factors are coprime, their product must collectively divide $x^{p-1} - 1$. Comparing degrees and leading coefficients, it follows that the two are in fact equal, hence the factorization.)

Note that if $d \mid p - 1$, then $x^d - 1 \mid x^{p-1} - 1$. (This is a strictly algebraic fact; for example, $x^{15} - 1 = (x^5 - 1)(x^{10} - x^5 + 1)$.) Write $x^{p-1} - 1 = (x^d - 1) \cdot g(x)$, where

$\deg g = (p-1) - d$. Now we invoke the above lemma. Since $\deg(x^d - 1) = d$, the equation $x^d - 1 = 0$ has at most $d$ roots. But $\deg g = (p-1) - d$, so $g$ has at most $(p-1) - d$ roots, which means $x^d - 1 = 0$ has at *least* $(p-1) - ((p-1) - d) = d$ roots. Therefore, $x^d - 1 = 0$ must have *exactly* $d$ roots, and we conclude. □

Recall we're doing all of this to prove Theorem 5.12: there exists an element in $(\mathbb{Z}/p\mathbb{Z})^\times$ of order $p-1$. It turns out that the above Corollary gives us enough "restricting conditions" to force this to be true.

Let's elaborate more in this big-picture argument. Suppose no such primitive root (element with order $p-1$) exists. We already know the order of any element must divide $|(\mathbb{Z}/p\mathbb{Z})^\times| = p - 1$. (This is the point of Proposition 5.10, and it comes as a consequence of Euler's Totient Theorem, Theorem 5.7.) But Corollary 6.3 gives us, for any divisor $d \mid p-1$, an exact number for how many elements have order $d$. A quick computation, invoking some of the work from §5.2, will show us that counting over all such elements for all divisors $d < p-1$ is not enough, so there must be an element with order $p-1$.

*Proof of Theorem 5.12.* Let $\psi(d)$ be the number of elements of $(\mathbb{Z}/p\mathbb{Z})^\times$ of order $d$. Note that an element satisfies $x^d - 1 = 0$ if it has order dividing $d$, i.e., if $c$ is the smallest positive integer such that $x^c - 1 = 0$, then $c \mid d$. Thus, we have $d = \sum_{c|d} \psi(c)$. (At this point, you can see we are set up nicely to use Möbius Inversion, Theorem 5.4.)

Recall (the remarkably clever) Proposition 5.5, which tells us $d = \sum_{c|d} \phi(c)$, where $\phi$ is the Euler Totient function. Our desire now is to show $\psi = \phi$. By Möbius Inversion, taking $f$ to be either $\psi$ or $\phi$ and $F = \mathrm{id}$, we have

$$\phi(d) = \sum_{c|d} \mu(c) \cdot d/c = \psi(d),$$

so $\psi(p-1) = \phi(p-1) > 0$. The conclusion follows. □

**Remark 6.4.** Note that not only does this show the existence of a primitive root, it also tells us exactly how many primitive roots there are! This is given by $\psi(p-1)$, which at the end we saw is just $\phi(p-1)$. One can see this more directly, though: once we know $(\mathbb{Z}/p\mathbb{Z})^\times$ is a primitive root, say $a$, then we know

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

Then, it is not hard to show directly that for any $m$ such that $(m, p-1) = 1$, $a^m$ is also a primitive root mod $p$. The number of $m$ such that $(m, p-1) = 1$ is $\phi(p-1)$ by definition.

## 6.2 Structure of Unit Groups

The punchline of the first part of this lecture is that for prime $p$, the subgroup of units $(\mathbb{Z}/p\mathbb{Z})^\times$ has a *cyclic* structure. We can now ask the following natural question:

> What is the structure of the unit group $(\mathbb{Z}/n\mathbb{Z})^\times$ for any $n$?

The answer comes from a significant result in number theory called the Chinese Remainder Theorem.[5]

**Theorem 6.5** (Chinese Remainder Theorem/Sunzi's Theorem)
Let $m_1, m_2, \ldots, m_s \in \mathbb{N}^+$ be pairwise coprime, i.e., $(m_i, m_j) = 1$ for all $i \neq j$. Choose some $a_i \in \mathbb{Z}/m_i\mathbb{Z}$ for each $1 \leq i \leq s$. Then, there exists some $a \in \mathbb{Z}$ such that $a \equiv a_i \pmod{m_i}$ for each $i$, and this solution is unique as an element of $\mathbb{Z}/(m_1 \cdots m_s)\mathbb{Z}$.

*Proof.* Since all $m_i$'s are pairwise coprime, for any prime $p$, $p$ divides at most one $m_i$. Denote $n = m_1 \cdots m_s$ and, for each $i$, denote $n_i = n/m_i = m_1 \cdots m_{i-1} m_{i+1} \cdots m_s$. By construction, $(m_i, n_i) = 1$, so there exist integers $(r_i, s_i)$ such that $r_i m_i + s_i n_i = 1$. Let $e_i = s_i n_i$. Again, by construction, observe

$$e_i \equiv \begin{cases} 0 & \mod m_j \text{ if } j \neq i \\ 1 & \mod m_i. \end{cases}$$

Now, we can prove the existence of such an $a \in \mathbb{Z}$ satisfying all congruences $a_i \pmod{m_i}$. We can simply construct $a = \sum_{i=1}^{s} a_i e_i$; by the above congruences on $e_i$, we have $a \equiv a_i \pmod{m_i}$.

Now, we prove uniqueness modulo $n$. If $a'$ is another such solution, then $a - a' \equiv a_j - a_j \equiv 0 \pmod{m_j}$, so $m_j \mid a - a'$. Since the $m_j$'s are pairwise coprime, it follows that $n = m_1 \cdots m_s \mid a - a'$, i.e., $a \equiv a' \pmod{n}$ as desired. □

Now we can describe the structure of $(\mathbb{Z}/n\mathbb{Z})^\times$ by first decomposing $\mathbb{Z}/n\mathbb{Z}$ following Sunzi's Theorem above.

[5] Note the subtle discrimination going on in the name: any result created by a European/American is credited by name (e.g., Fermat's Little Theorem), but here they fail to give a specific name despite knowing its founder. There is some push in the math community to rename it Sunzi's Theorem, since the result is first known to be stated by Sunzi.

**Corollary 6.6** (Structure of $(\mathbb{Z}/n\mathbb{Z})^\times$)
If $n = p_1^{b_1} \cdots p_s^{b_s}$ where the $p_i$ are distinct primes, then

$$\mathbb{Z}/n\mathbb{Z} \simeq \mathbb{Z}/p_1^{b_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_s^{b_s}\mathbb{Z}$$
$$a \leftrightarrow (a_1, a_2, \ldots, a_s)$$

and, by taking units,

$$(\mathbb{Z}/n\mathbb{Z})^\times \simeq (\mathbb{Z}/p_1^{b_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_s^{b_s}\mathbb{Z})^\times$$
$$a_1^{-1} \leftrightarrow (a_1^{-1}, \ldots, a_s^{-1}).$$

There is no new content in here; the first isomorphism is just a reformulation of Sunzi's Theorem. Note that a unit of the product on the right must be a unit in each component (if it were non-invertible in some $\mathbb{Z}/p_i^{b_i}\mathbb{Z}$, then it cannot be invertible in the product), and so we get the second isomorphism.

We can elaborate even more on this result by describing the structure of these $(\mathbb{Z}/p^b\mathbb{Z})^\times$ unit groups. This is definitely doable, it just takes a little time, so we will leave the following as just a fact and move on to our next topic.

**Fact 6.7.** If $p$ is an odd prime (i.e., $p > 2$) and $b \in \mathbb{N}^+$, then $(\mathbb{Z}/p^b\mathbb{Z})^\times$ is always cyclic.

# 7 09/29 - Quadratic Reciprocity

We technically started this topic in the last 15 minutes of the last lecture, but because this is the next main topic of the course, it felt fitting to just start at new section.

## 7.1 Motivation

Here is the premise of the topic.[6] In general, we are interested in solving polynomial equations. Doing this over $\mathbb{R}$, even $\mathbb{C}$, for linear and quadratic equations is the whole point of the algebra sequence in middle/high school. Doing this in generality is the birthplace of algebraic geometry, one of the most prominent fields in modern mathematics. (Take Math 137 or 232A/B if this piques your interest.)

Number theory cares about things modulo $n$. We can do even better: by Sunzi's Theorem, to study something modulo $n$, it suffices to study it in modulo $p$ for primes $p \mid n$. For instance, if we want to solve $x^3 - 3 = 0 \pmod{30}$, we can solve it in mod 2, 3, and 5, then combine our findings to find solutions modulo 30.

[6] This was not covered in class, I'm just adding this for more context.

Solving linear equations modulo $p$ is easy in some cases and doable in all cases. (We might talk more about this later in the course.) If I give you something like $x - 3 = 0 \pmod{p}$, it is obvious what $x$ can be mod $p$. An equation like $3x \equiv 1 \pmod{7}$ is also completely doable. (Do it!)

Even if I give you something clunky like $Ax \equiv B \pmod{C}$ (take something ridiculous like $A = 11^{1234}$, $B = 420^{420}$, and $C = 7^{7^7}$), we know in general that we can find integers $x$ and $y$ such that $Ax + Cy = (A, C)$ using the Euclidean algorithm/the process you did on your homework. Thus, so long as $(A, C) \mid B$ (which in our example is true, since $(A, C) = (11, 7) = 1$), we have $B = (A, C) \cdot d$, so $A(dx) + C(dy) = (A, C) \cdot d = B \implies A \cdot (dx) = B \pmod{C}$. The point is that solving linear equations is completely understood, and pretty efficient.

The next step is solving quadratic equations. The simplest such equation is of the form $x^2 \equiv a \pmod{p}$. (In fact, every quadratic can be reduced to this form. For instance, if $2x^2 + 3x - 1 \equiv 0 \pmod{7}$, then $2x^2 - 4x = 1 = 8 \pmod{7} \implies x^2 - 2x = 4 \implies (x-1)^2 = 5 \pmod{7}$.) Quadratic reciprocity allows us to answer these questions in a marvelously efficient way. For example, once we lay out the main result, then we can compute problems like these very easily:

> **Exercise 7.1.** Does there exist an $x \in \mathbb{Z}$ such that $x^2 \equiv 37 \pmod{67}$.

Before reading the next section, I invite you to play around with these two baby exercises:

> **Exercise 7.2.** Take the first eight odd primes $p \in \{3, 5, 7, 11, 13, 17, 19, 23, 29\}$. For each of these primes, determine whether $-1$ is a quadratic residue. I'll start: in mod 3, $-1 = 2$, but $1^2 = 2^2 = 1$ modulo 3, so $-1$ is not a quadratic residue. On the other hand, in mod 5, $-1 = 4 = 2^2$, so $-1$ is a quadratic residue. Do this for all primes, and try to see a pattern! The answer may come as a surprise.
>
> For the even more curious, do the same for 2. We saw above that $2 = -1$ is not a quadratic residue modulo 3, and it turns out that 2 is also not a quadratic residue modulo 5. Can you find any patterns?

## 7.2 Quadratic Residues

In light of the above discussion, we proceed with a natural definition.

> **Definition 7.3** (Quadratic Residue)**.** We say $a \in \mathbb{Z}/p\mathbb{Z}$ is a **quadratic residue** modulo $p$ if there exists some $x \in \mathbb{Z}/p\mathbb{Z}$ such that $x^2 \equiv a \pmod{p}$.

We will define a funny-looking, half $\binom{a}{b}$-looking, half fraction-looking symbol as a kind of indicator function on whether or not $a$ is a quadratic residue mod $p$.

**Definition 7.4** (Legendre Symbol)**.** Let $a \in \mathbb{Z}$ and $p$ prime such that $(a, p) = 1$. Then, the **Legendre symbol** is defined as

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

Most of the time, we will restrict our attention to when $a \neq 0$, i.e., $(a, p) = 1$, because if $a = 0$, then we can do the obvious $0^2 = 0$, which is not so interesting.

From Theorem 5.12, we know that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic. Let $h$ be a primitive root modulo $p$. Thus, for any $a \in (\mathbb{Z}/p\mathbb{Z})^\times$, we can write $a = h^i$ for some integer $i$. One can show that $a$ is a quadratic residue if and only if $i$ is even (in which case $a = (h^{i/2})^2$).

Let us elaborate on this a little bit more via the following result.

**Lemma 7.5**

For $a \in \mathbb{Z}$ and $p$ an odd prime such that $(a, p) = 1$, we have

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

This is quite an exciting result! For those who entertained Exercise 7.2, you probably figured there must be a better way to check if a number is a quadratic residue. Well, here we go.

Initially, the fastest way to compute $\left(\frac{a}{p}\right)$ is by going through all $x \in (\mathbb{Z}/p\mathbb{Z})^\times$ and seeing if $x^2 \equiv a \pmod{p}$. In particular, if $a$ is not a quadratic residue, that would require us to go through all $x \in (\mathbb{Z}/p\mathbb{Z})^\times$. For each $x$, we have two computations – square $x$, then reduce mod $p$ – giving a total of $2(p-1)$ computations. Now, we can compute it just from multiplying $a$ to itself $\frac{p-1}{2}$ times, which is far less computationally.

*Proof.* We will first show $a = h^i$ is a quadratic residue if and only if $i$ is even. If $i$ is even, then $a = (h^{i/2})^2$, as demonstrated above. On the other hand, if $x^2 = a \pmod{p}$, then we can also write $x = h^j$ for some $j$ since $x \in (\mathbb{Z}/p\mathbb{Z})^\times$. But then

$$\begin{aligned} x^2 &\equiv a \pmod{p} \\ \implies h^{2j} &\equiv h^i \pmod{p} \\ \implies h^{2j-i} &\equiv 1 \pmod{p}. \end{aligned}$$

Since the order of $h$ mod $p$ is $p-1$, this implies $p-1 \mid 2j-i$. But for odd $p$, $p-1$ is even, so $2j-i$ must be even as well. In particular, $i$ must be even.

We continue with this equivalence. We have $i$ is even if and only if $p-1 \mid i \cdot \frac{p-1}{2}$, which is equivalent to

$$h^{i \cdot \frac{p-1}{2}} = (h^i)^{\frac{p-1}{2}} = a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Thus, to summarize, there exists an $x$ such that $x^2 \equiv a \pmod{p}$ (i.e., $\left(\frac{a}{p}\right) = 1$) if and only if $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. When no such $x$ exists, i.e., when $\left(\frac{a}{p}\right) = -1$, then $a^{\frac{p-1}{2}} \not\equiv 1$. But note that $\left(a^{\frac{p-1}{2}}\right)^2 = a^{p-1} = 1$ by Fermat's Little Theorem, and this is only possible if $a^{\frac{p-1}{2}} \equiv \pm 1$, so $\left(\frac{a}{p}\right) = -1 \iff a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$. We have now covered both cases, so the conclusion follows. □

This lemma now gives us a really nice characterization of when $-1$ is a quadratic residue modulo $p$, just by plugging in $a = -1$. This answers the first part of Exercise 7.2.

**Corollary 7.6** (Criterion for $\left(\frac{-1}{p}\right)$)
$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$. In other words, $\left(\frac{-1}{p}\right) = 1$ if $p \equiv 1 \pmod{4}$ and $-1$ if $p \equiv 3 \pmod{4}$.

So $-1$ being a quadratic residue mod $p$ is determined by $p$ mod 4, which is a condition that kind of comes out of nowhere at first glance. The condition for $a = 2$ requires a little more care, and we will prove it next time, but considering $p$ mod 8 provides sufficient information.

**Proposition 7.7**
$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$. In other words

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & p \equiv 1, 7 \pmod{8} \\ -1 & p \equiv 3, 5 \pmod{8}. \end{cases}$$

Finally, we state this truly remarkable result, first proved by none other than Gauss. It may look a bit complicated at first, but it makes this question regarding quadratic residues very simple.

**Theorem 7.8** (Quadratic Reciprocity)
Let $p$ and $q$ be odd primes. Then,

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

This makes answering questions like Exercise 7.1 not only doable, but even doable by hand.

*Answer to Exercise 7.1.* We compute

$$\begin{aligned}\left(\frac{37}{67}\right) &= \left(\frac{67}{37}\right) = \left(\frac{30}{37}\right) \\ &= \left(\frac{2}{37}\right)\left(\frac{3}{37}\right)\left(\frac{5}{37}\right) \\ &= (-1)\left(\frac{37}{3}\right)\left(\frac{37}{5}\right) \\ &= (-1)\left(\frac{1}{3}\right)\left(\frac{2}{5}\right) \\ &= (-1)\cdot 1\cdot(-1) = 1,\end{aligned}$$

so indeed, 37 is a quadratic residue mod 67. □

## 7.3 Proof of Quadratic Reciprocity, Step 1

We will work towards proving Theorem 7.8. The proof we follow here is a slick one; the most "elementary" one involving Gauss sums will be demonstrated next week. We use the word "elementary" here with caution, though, because elementary does not necessarily mean easy. As we are using less powerful tools, we have to be more creative, and we'll see that next week's proof of Quadratic Reciprocity requires a lot of jumping through hoops.

Today, we are bargaining a little with elementary methods and higher-power tools. All the steps in each proof, albeit scary-looking, follow pretty smoothly, but we do have to reference a result that goes beyond standard number theory (this is Lemma 7.10). This makes the proof a little less grounded, and thus feel a bit more magical, but bear with us for a little while.

We first lay out a few details. Let $S = \left\{-\frac{p-1}{2}, -\frac{p-3}{2}, \ldots, -1, 1, 2, \ldots, \frac{p-1}{2}\right\}$ be the set of non-zero residues mod $p$. Let $a \in \mathbb{Z}$ be an integer coprime to $p$, so $(p, a) = 1$. Consider the set $\left\{a, 2a, \ldots, \frac{p-1}{2}\cdot a\right\}$. Again, this may seem weird (why are we doing this?), but observe that this set has no duplicates: if $ai \equiv aj \pmod p$, then $a(i-j) \equiv 0 \pmod p$ which is only possible if $i = j$ (as $(p, a) = 1$).

For $1 \le i \le \frac{p-1}{2}$, let $m_i \in \left\{1, \ldots, \frac{p-1}{2}\right\}$ such that $ia \equiv \pm m_i \pmod p$. We will consider all instances where we take the negative sign in this equivalence, i.e., define

$$\mu := \mu(a) = \left|\left\{1 \le i \le \frac{p-1}{2} : ia \equiv -m_i,\ 1 \le m_i \le \frac{p-1}{2}\right\}\right|.$$

Let's get to the point:

**Lemma 7.9** (Gauss)
$\left(\frac{a}{p}\right) = (-1)^{\mu}$.

*Proof.* As proven above, all elements of $\left\{a, 2a, \ldots, \frac{p-1}{2}a\right\}$ are distinct modulo $p$, for if $ai \equiv aj \pmod{p}$, then $p \mid a(i-j) \implies i \equiv j \pmod{p}$, which is not possible when $1 \leq i, j \leq \frac{p-1}{2}$ unless $i = j$.

Even better, we can show all $m_i$'s are distinct. If the sign of $m_i$ and $m_j$ are the same in $ia \equiv \pm m_i$, $ja \equiv \pm m_j$, then we can use the argument before to show $m_i \equiv m_j \implies ai \equiv aj \implies i = j$. Likewise, if the signs for $m_i$ and $m_j$ are distinct, then we have $m_i \equiv -m_j \implies ai \equiv -aj \implies i \equiv -j \pmod{p}$, which is impossible as $1 \leq i, j \leq \frac{p-1}{2}$. Thus, the $m_i$'s all take on values from 1 to $\frac{p-1}{2}$, and they are all distinct, so

$$\left\{m_1, m_2, \ldots, m_{\frac{p-1}{2}}\right\} = \left\{1, 2, \ldots, \frac{p-1}{2}\right\}.$$

Recall $ai \equiv \pm\mu \pmod{p}$, and the number of times the sign is negative is $\mu$ by definition. We now multiply all $m_i$'s together to get

$$a^{\frac{p-1}{2}}\left(\frac{p-1}{2}\right)! = \prod_{i=1}^{\frac{p-1}{2}} ai = (-1)^{\mu}\prod_{i=1}^{\frac{p-1}{2}} m_i.$$

But note from the equality of sets $\left\{m_1, m_2, \ldots, m_{\frac{p-1}{2}}\right\} = \left\{1, 2, \ldots, \frac{p-1}{2}\right\}$, we have $\prod_i m_i = \prod_i i = \left(\frac{p-1}{2}\right)!$. Canceling this out on both sides, we conclude

$$(-1)^{\mu} = a^{\frac{p-1}{2}},$$

so $(-1)^{\mu} = \left(\frac{a}{p}\right)$ by Lemma 7.5, as desired. □

As a corollary, we can prove Proposition 7.7, which says that 2 is a quadratic residue when $p \equiv 1, 7 \pmod{8}$, and not a quadratic residue otherwise.

*Proof of Lemma 7.5.* We use Lemma 7.9, which we just proved. Here, we may compute $\mu$ explicitly. Given $i$ between 1 and $\frac{p-1}{2}$, note $2i < p$, so we have $2i \equiv -m_i$ for some $1 \leq m_i \leq (p-1)/2$ if and only if $2i > \frac{p-1}{2}$, i.e. $i > \frac{p-1}{4}$. Thus, $\mu$ is equal to the number of $1 \leq j \leq (p-1)/2$ such that $j > \frac{p-1}{4}$, which by complementary counting is just $\frac{p-1}{2} - m$ where $m = \left\lfloor \frac{p-1}{4} \right\rfloor$. We do this by casework:

- ($p \equiv 1 \pmod{8}$) We can write $p = 8k+1$. Then, $m = 2k$, so $\frac{p-1}{2} - m = 4k = 2k = 2k$. Thus, $\mu$ is even.
- ($p \equiv 3 \pmod{8}$) We can write $p = 8k+3$. Then, $m = 2k$ again, so $\mu = \frac{p-1}{2} - m = 4k + 1 - 2k = 2k + 1$ is odd.
- ($p \equiv 5 \pmod{8}$) Write $p = 8k+5$. Then, $m = 2k+1$, so $\mu = \frac{p-1}{2} - m = (4k+2) - (2k+1) = 2k+1$ is odd.
- ($p \equiv 7 \pmod{8}$) Write $p = 8k+7$. Then, $m = 2k+1$, so $\mu = \frac{p-1}{2} - m = (4k+3) - (2k+1) = 2k+2$ is even.

Collecting all of these computations, the result follows. □

## 7.4 Proof of Quadratic Reciprocity, Step 2

More magic to ensue. The next lemma is the step where we go beyond just working with the integers, which relieves future computations but requires us to prove something slightly more difficult at the onset.

Consider the function $f(z) = e^{2\pi iz} - e^{-2\pi iz}$. Observe $f(z) = f(z+1)$ and $f(-z) = -f(z)$. If you know $e^{i\theta} = \cos\theta + i\sin\theta$, then see that $f(z) = 2i\sin(2\pi z)$.

**Lemma 7.10**

$$\frac{f(nz)}{f(z)} = \prod_{m=1}^{(n-1)/2} f\left(z + \frac{m}{n}\right) f\left(z - \frac{m}{n}\right).$$

*Proof.* We first prove a smaller lemma.

**Lemma 7.11**

If $n > 0$ is odd, then

$$x^n - y^n - \prod_{k=0}^{n-1} \left(\zeta^k x - \zeta^{-k} y\right),$$

where $\zeta = e^{2\pi i/n}$.

*Proof.* Note that for $\zeta = e^{2\pi i/n}$, we have $(\zeta^k)^n = e^{2\pi ik} = 1$, so all $\zeta^k$'s are a root of the polynomial $z^n - 1 = 0$. But there are $n$ such powers of $\zeta$, and $\deg(z^n - 1) = n$, so they each appear as a root exactly once. In particular,

$$z^n - 1 = \prod_{k=0}^{n-1} (z - \zeta^k).$$

If $z = x/y$, then we can write $x^n - y^n = \prod_{k=0}^{n-1}(x - \zeta^k y)$.

Now for odd $n$, the map $x \mapsto -2x$ is a bijection between $\mathbb{Z}/n\mathbb{Z}$ and itself (it is injective, as $-2a = -2b \implies a = b$, and its inverse is $y \mapsto -1/2 \cdot y$, and $-1/2$ exists

mod $n$ since $n$ is odd). Thus, we have

$$\begin{aligned} x^n - y^n &= \prod_{k=0}^{n-1}(x - \zeta^k y) \\ &= \prod_{k=0}^{n-1}(x - \zeta^{-2k} y) \\ &= (\zeta^n)^{\frac{n-1}{2}} \prod_{k=0}^{n-1}(x - \zeta^{-2k} y) \\ &= \zeta^{1+2+\cdots+(n-1))} \prod_{k=0}^{n-1}(x - \zeta^{-2k} y) \\ &= \prod_{k=0}^{n-1}(\zeta^k x - \zeta^{-k} y), \end{aligned}$$

as desired. □

Great, let us return to the main lemma at hand. We apply the lemma above with $x = e^{2\pi i z}$ and $y = e^{-2\pi i z}$ to get

$$\begin{aligned} f(nz) &= e^{2\pi i n z} - e^{-2\pi i n z} \\ &= \prod_{k=0}^{n-1} e^{2\pi i\left(z+\frac{k}{n}\right)} - e^{2\pi i\left(-\frac{k}{n}-z\right)} \\ &= \prod_{k=0}^{n-1} f\left(z + \frac{k}{n}\right). \end{aligned}$$

From $f(z) = f(z+1)$, we have $f(z + k/n) = f(z + k/n - 1) = f\left(z - \frac{n-k}{n}\right)$. Now, we can rewrite our product above: for $\frac{n+1}{2} \le k \le n-1$, we have $1 \le n - k \le \frac{n-1}{2}$, so

$$\begin{aligned} f(nz) &= f(z + 0/n) \prod_{k=1}^{(n-1)/2} f(z + k/n) \prod_{k=\frac{n+1}{2}}^{n-1} f(z + k/n) \\ &= f(z) \prod_{k=1}^{\frac{n-1}{2}} f\left(z + \frac{k}{n}\right) f\left(z - \frac{k}{n}\right), \end{aligned}$$

which is equivalent to what we want. □

## 7.5 Proof of Quadratic Reciprocity, Step 3

The payout for this lemma is high, as promised.

**Proposition 7.12**

If $p$ is an odd prime and $a \in \mathbb{Z}$ such that $(p, a) = 1$, then

$$\prod_{\ell=1}^{(p-1)/2} f\left(\frac{\ell a}{p}\right) = \left(\frac{a}{p}\right) \prod_{\ell=1}^{(p-1)/2} f\left(\frac{\ell}{p}\right).$$

*Proof.* We can write $\ell a \equiv \pm m_\ell \pmod{p}$ for some $1 \leq m_\ell \leq (p-1)/2$. This tells us that $\frac{\ell a \mp m_\ell}{p}$ is an integer, so using the relations $f(z) = f(z+1)$ and $f(z) = -f(-z)$,

$$f(\ell a/p) = -f(\mp m_\ell/p) = \pm f(m_\ell/p).$$

Multiplying across all $1 \leq \ell \leq (p-1)/2$ on both sides, we have

$$\prod_{\ell=1}^{\frac{p-1}{2}} f(\ell a/p) = (-1)^\mu \prod_{\ell=1}^{\frac{p-1}{2}} f(\ell/p) = \left(\frac{a}{p}\right) \prod_{\ell=1}^{\frac{p-1}{2}} f(\ell/p),$$

where the last equality invokes Lemma 7.9. □

Now why is this useful? Well, we can now prove Quadratic Reciprocity.

*Proof of Theorem 7.8.* Take $p, q$ odd primes, and apply the above Proposition to $a = q$. (Note that $p, q$ are "symmetric" in the sense that they are interchangeable, so whatever we do for $q$, we can do the same for $p$.) The Proposition tells us

$$\prod_{\ell=1}^{\frac{p-1}{2}} f(\ell q/p) = \left(\frac{q}{p}\right) \prod_{\ell=1}^{\frac{p-1}{2}} f(\ell/p),$$

so by Lemma 7.10 (here $z = \ell/p$ and $n = q$,

$$\left(\frac{q}{p}\right) = \prod_{\ell=1}^{\frac{p-1}{2}} \frac{f(\ell q/p)}{f(\ell/p} = \prod_{\ell=1}^{\frac{p-1}{2}} \prod_{m=1}^{\frac{q-1}{2}} f\left(\frac{\ell}{p} + \frac{m}{q}\right) f\left(\frac{\ell}{p} - \frac{m}{q}\right).$$

Switching $q$ and $p$, we can go through the same process by applying the Proposition for $a = p$ to get

$$\begin{aligned}
\left(\frac{p}{q}\right) &= \prod_{\ell=1}^{\frac{q-1}{2}} \prod_{m=1}^{\frac{p-1}{2}} f\left(\frac{\ell}{q} + \frac{m}{p}\right) f\left(\frac{\ell}{q} - \frac{m}{p}\right) \\
&= \prod_{m=1}^{\frac{q-1}{2}} \prod_{\ell=1}^{\frac{p-1}{2}} f\left(\frac{m}{q} + \frac{\ell}{p}\right) f\left(\frac{m}{q} - \frac{\ell}{p}\right) \\
&= (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}} \left(\frac{q}{p}\right),
\end{aligned}$$

which, quite remarkably, is what we wanted. □

# 8 10/02 - Algebraic Numbers & Integers

We will use something called "quadratic Gauss sums" (discussed in next lecture) to provide another proof of Quadratic Reciprocity. The upshot of these techniques is that we can generalize our results to higher dimensions. We begin with some definitions.

## 8.1 Algebraic Numbers

**Definition 8.1** (Algebraic Numbers/Integers)**.** An **algebraic number** is a number $\alpha \in \mathbb{C}$ which is a solution to a polynomial equation of the form

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$$

for $a_{n-1}, \ldots, a_0 \in \mathbb{Q}$. (Equivalently, we can clear denominators and say $\alpha$ satisfies $c_n\alpha^n + c_{n-1}\alpha^{n-1} + \cdots + c_0 = 0$ for integers $c_i$.

We say $\alpha$ is an **algebraic integer** if $a_{n-1}, \ldots, a_0 \in \mathbb{Z}$, or equivalently if $c_n = 1$.

Another way to think about this is that $\alpha$ is an algebraic number if it satisfies some polynomial relation $p(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0 = 0$ for $a_i \in \mathbb{Q}$, and likewise $a_i \in \mathbb{Z}$ for algebraic integers.

**Example 8.2** (Algebraic Numbers/Integers)
$\sqrt{2}$ is an algebraic integer, since $(\sqrt{2})^2 - 2 = 0$. On the other hand, $\sqrt{3}/4$ is an algebraic number, since $(\sqrt{3}/4)^2 - 3/16 = 0$, but this is not an algebraic integer.

In general, we can always find algebraic numbers which are not algebraic integers. We in fact have a nice characterization of when algebraic numbers are algebraic integers. (Spoiler: it's given in the name.)

**Proposition 8.3**
Let $r \in \mathbb{Q}$. Then,

1. $r$ is an algebraic number.
2. if $r$ is an algebraic integer, then $r \in \mathbb{Z}$.

*Proof.* The first is trivial: $r$ satisfies $p(X) = X - r = 0$. We now focus on the second statement. Assume there are $a_i \in \mathbb{Z}$ such that $r^n + a_{n-1}r^{n-1} + \cdots + a_0 = 0$. Write $r = \frac{p}{q}$, where $p, q \in \mathbb{Z}$ and $(p, q) = 1$. Clearing denominators in our polynomial relation,

we have

$$\begin{aligned} & p^n + a_{n-1}p^{n-1}q + \cdots + a_n q^n = 0 \\ \implies & -q(a_{n-1}p^{n-1} + a_{n-2}p^{n-2}q + \cdots + a_n q^{n-1}) = p^n. \end{aligned}$$

But this means $q$ divides $p^n$, which is only possible when $(p, q) = 1$ if $q = \pm 1$. This means $r = \pm p \in \mathbb{Z}$, as desired. □

We have provided characterizations of algebraic numbers and integers as elements. Now, let us consider the set of algebraic numbers (resp. integers). We will show that these have very reasonable and familiar structures: the set of algebraic numbers is a *field*, and the set of algebraic integers is a *ring*.

If you think about it for a little bit, this is quite difficult to do just from the definitions! Even if I give you really simple algebraic integers, say like $\sqrt{2}$ and $\sqrt{3}$, it requires a lot of brainpower to construct a polynomial $p(X)$ such that $p(\sqrt{2} + \sqrt{3}) = 0$. We will find a way to prove this without providing explicit constructions for our polynomial relations.

**Definition 8.4** (Module)**.** Let $V \subseteq \mathbb{C}$ be a subset. Then, $V$ is a $\mathbb{Q}$-vector space (equivalently, a $\mathbb{Q}$-**module**) of finite dimension if

1. $\forall\, x, y \in V$, $x + y \in V$;
2. $\forall\, r \in \mathbb{Q}, x \in V$, $r \cdot x \in V$;
3. $\exists\, \gamma_1, \ldots, \gamma_n \in V$ such that $\forall\, x \in V$, $\exists!(r_1, \ldots, r_n) \in \mathbb{Q}^n$ such that $x = \sum_{i=1}^n r_i \gamma_i$. (The $\gamma_i$'s are the generators of $V$.)

So we have entered the land of linear algebra, which is great, because we know linear algebra really well.[7]

**Proposition 8.5**

Let $V$ be a $\mathbb{Q}$-module. Let $\alpha \in \mathbb{C}$ such that $\alpha \cdot V \subseteq V$. Then, $\alpha$ is an algebraic number.

*Proof.* Let $\gamma_1, \ldots, \gamma_n$ be a basis of $V$. Consider the map

$$\begin{aligned} m_\alpha : V &\to V \\ x &\mapsto \alpha x. \end{aligned}$$

It is easy to see that this is a $\mathbb{Q}$-linear map. Let $M \in M_n(\mathbb{Q})$ be its matrix with respect to the basis $\gamma_1, \ldots, \gamma_n$. Take the characteristic polynomial $P(X) = \det(M - XI_n) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$; since $M \in M_n(\mathbb{Q})$, the coefficients here live in $\mathbb{Q}$.

[7] "We" means the math community at large. I myself am pretty bad at linear algebra, oops.

Now we invoke the Cayley-Hamilton Theorem, one of the most important results in linear algebra. The theorem states that plugging in the matrix into the characteristic polynomial of the matrix gives 0, so here, we have $P(m_\alpha) = m_\alpha^n + a_{n-1}m_\alpha^{n-1} + \cdots + a_0 = 0$. But we know exactly what $m_\alpha^k$ is from definition: $m_\alpha^k(x) = \alpha^k x$. Thus, $P(m_\alpha)(x) = (\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0)x = 0$ for all $x$. Setting $x \neq 0$, this forces the sum in the parentheses to be 0, so $\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$. Hence, $\alpha$ is algebraic, as desired. □

## 8.2 Algebraic Numbers (Integers) form a Field (Ring)

We now put this sick result to use:

**Proposition 8.6**

The set of algebraic numbers is a field.

*Proof.* Let $\alpha, \beta \in \mathbb{C}$ be two algebraic numbers. We want to show three things are algebraic: (1) $1/\alpha$ (so the set has inverses), (2) $\alpha \cdot \beta$ (closed under multiplication), and (3) $\alpha + \beta$ (closed under addition).

We start with (1). We know by definition that $\alpha, \beta$ satisfy the relations

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$$
$$\beta^m + b_{m-1}\beta^{m-1} + \cdots + b_0 = 0$$

for $a_i, b_i \in \mathbb{Q}$. Assume that $a_0 \neq 0$ (otherwise we can divide the first equation by $\alpha$). Dividing by $a_0\alpha^n$, we get a new equation

$$\frac{1}{a_0} + \frac{a_{n-1}}{a_0} \cdot \frac{1}{\alpha} + \cdots + \frac{a_{n-i}}{a_0} \cdot \left(\frac{1}{\alpha}\right)^i + \left(\frac{1}{\alpha}\right)^n = 0,$$

so $1/\alpha$ is an algebraic number.

We will prove (2) and (3) with the same approach. Let $V$ be the $\mathbb{Q}$-module with basis $\{\alpha^k\beta^j : 0 \leq k < n, 0 \leq j < m\}$. By construction, this has finite dimension. Furthermore, any element in $V$ is of the form $v = \sum_{i,j} r_{ij} \cdot \alpha^i\beta^j$, and we have $\alpha v = \sum_{i,j} r_{ij} \cdot \alpha^{i+1}\beta^j$, which is still an element of $V$. (The only thing we have to check is that $\alpha \cdot \alpha^{n-1}$ still lives in $V$, which is true since $\alpha^{1+(n-1)} = \alpha^n = -\sum_{i<n} a_i\alpha^i \in V$.) Likewise, $\beta v \in V$.

But this means $\alpha \cdot V \subseteq V$ and $\beta \cdot V \subseteq V$. Now, we can take sums and products to get $(\alpha + \beta) \cdot V \subseteq V$ and $(\alpha\beta)V \subseteq V$. Proposition 8.5 tells us then that $\alpha + \beta$ and $\alpha\beta$ are algebraic numbers, as desired. □

To show that the set of algebraic integers for a ring, we now provide the proof given in the textbook. It is basically going to be the same, except instead of working over $\mathbb{Q}$, we will work over $\mathbb{Z}$.

**Remark 8.7.** We called a $\mathbb{Q}$-vector space as a $\mathbb{Q}$-module because modules are more general than vector spaces. For example, there are no such things as a $\mathbb{Z}$-vector space since $\mathbb{Z}$ is not a field. However, modules are defined over rings, so it makes sense to talk about a $\mathbb{Z}$-module.

**Definition 8.8** ($\mathbb{Z}$-module)**.** Let $W \subseteq \mathbb{C}$. We say that $W$ is a $\mathbb{Z}$-module if

1. $W$ is an abelian subgroup;
2. if $n \in \mathbb{Z}, \alpha \in W$, then $n \cdot \alpha \in W$;
3. $\exists\, \gamma_1, \ldots, \gamma_n \in W$ such that every element $x \in W$ can be written as $x = \sum_{i=1}^{n} n_i \gamma_i$ for some $n_i \in \mathbb{Z}$.

Akin to Proposition 8.5, we have this analogous result, which uses a similar Cayley-Hamilton argument.

**Proposition 8.9**

Let $W$ be a $\mathbb{Z}$-module. Let $\alpha \in \mathbb{C}$ such that $\alpha W \subseteq W$. Then, $\alpha$ is an algebraic integer.

*Proof.* Let $\gamma_1, \ldots, \gamma_n$ be as in the definition (a set of generators of $W$). Since $\alpha \cdot \gamma_i \in W$, we have that for any $1 \leq i \leq n$, there exists $c_{ij} \in \mathbb{Z}$ such that

$$\alpha \cdot \gamma_i = \sum_{j=1}^{n} c_{ij} \gamma_j$$

$$\iff \sum_{j=1}^{n} (\alpha \delta_{ij} - c_{ij}) \gamma_j = 0,$$

where $\delta_{ij}$ is the Kronecker delta function that gives 1 when $i = j$ and 0 otherwise.

Consider the $n \times n$-matrix $M = (\alpha \delta_{ij} - c_{ij})_{i,j}$. Note the above equality indicates $M \cdot (\gamma_1 \cdots \gamma_n)^\intercal = 0$, which forces $\det M = 0$ since the $\gamma_i$'s are nonzero. But $\det M$ is a polynomial in $\alpha$ with integer coefficients since each $c_{ij} \in \mathbb{Z}$. Note also that each nonzero coefficient of $\alpha$ in $M$ is $\delta_{ii} = 1$, so $\det M$ is a monic polynomial in $\alpha$. Hence, $\alpha$ is an algebraic integer. □

The proof that the set of algebraic integers forms a ring follows straight from this Proposition, similar to what we did for algebraic numbers.

**Corollary 8.10**

The set of algebraic integers is a ring.

*Proof.* Take $W$ to be the $\mathbb{Z}$-module generated by $\{\alpha^i\beta^j \mid 0 \leq i < n, 0 \leq j < m\}$, and proceed as in the proof of Proposition 8.6. □

## 8.3 Properties of Algebraic Numbers

Denote $\Omega$ as the set of algebraic integers. If $w_1, w_2, \gamma \in \Omega$ with $\gamma \neq 0$, then we say $w_1 \equiv w_2 \pmod{\gamma}$ if we can find some $\delta \in \Omega$ such that $w_1 - w_2 = \delta \cdot \gamma$. (Note this is basically how we defined modular congruence in the integers as well: $a \equiv b \pmod{q}$ in the integers if $a - b = qm$ for some $m \in \mathbb{Z}$.)

Suppose we have $a, b, c \in \mathbb{Z}$ such that $a \equiv b \pmod{c}$. Thinking of these are elements of $\Omega \supseteq \mathbb{Z}$, we have $a \equiv b \pmod{c}$ in $\Omega$, so $a - b = c\delta$ for some $\delta \in \Omega$. Thus, $\delta = \frac{a-b}{c} \in \mathbb{Q} \cap \Omega = \mathbb{Z}$ by Proposition 8.3, so in fact over the integers, these two modular congruences agree.

Given this, we can show that the Freshman's Dream $(a+b)^p \equiv a^p + b^p \pmod{p}$ also holds when $a, b \in \Omega$.

**Proposition 8.11**

Let $w_1, w_2 \in \Omega$ and $p \in \mathbb{Z}$ be a prime number. Then, $(w_1 + w_2)^p \equiv w_1^p + w_2^p \pmod{p}$.

*Proof.* The proof follows exactly like the proof for the integers. By the Binomial Theorem,

$$(w_1 + w_2)^p = w_1^p + w_2^p + \sum_{k=1}^{p-1} \binom{p}{k} w_1^k w_2^{p-k}.$$

Since $p \mid \binom{p}{k}$ for $1 \leq k \leq p-1$, the result follows. □

Well, we know an algebraic number satisfies some polynomial relation. Can we find this polynomial relation? I mentioned earlier that this is a bit difficult to do by hand; for instance, $\sqrt{2}$ satisfies $X^2 - 2 = 0$, and $\sqrt{3}$ corresponds to $X^2 - 3 = 0$, but given these two polynomials, it is hard to find the minimal polynomial that has $\sqrt{2} + \sqrt{3}$ as a root. We will "find" this minimal polynomial through some algebra.

Let $\alpha \in \mathbb{Q}$ be an algebraic number. Then, $S = \{P \in \mathbb{Q}[X] \mid P(\alpha) = 0\}$ is an ideal of $\mathbb{Q}[X]$. But $\mathbb{Q}[X]$ is a principal ideal domain (it has a Euclidean algorithm, recall our work for $k[x]$ in Lecture 2), so $S = (f)$ for some irreducible monic $f \in \mathbb{Q}[X]$.

In fact, $f$ is the polynomial of minimal degree such that $f(\alpha) = 0$ and $f$ is monic. We call $f$ the **minimal polynomial** of $\alpha$, and the degree of $f$ is called the **degree of** $\alpha$.

Note that we could have also come to this more directly without considering $\mathbb{Q}[X]$ as a PID. Take the set $S$, take the element $f \in S$ which is monic and has minimal degree. Suppose $g \in S$ as well. Then, the Division Algorithm in $\mathbb{Q}[X]$ tells us that

$g(X) = f(X)q(X) + r(X)$ for some $q, r \in \mathbb{Q}[X]$ such that $\deg r < \deg f$. But then $r(\alpha) = g(\alpha) - f(\alpha)q(\alpha) = 0$, so $r \in S$ as well. This is only possible if $r = 0$ by minimality of $f$, which implies $f \mid g$. Hence, $S = (f)$, and $f$ is the minimal polynomial of $\alpha$ by construction.

Define the sets

$$\mathbb{Q}[\alpha] = \{P(\alpha) \mid P \in \mathbb{Q}[X]\} \subset \mathbb{C}$$
$$\mathbb{Q}(\alpha) = \left\{\frac{P(\alpha)}{Q(\alpha)} : P, Q \in \mathbb{Q}[X]\right\} \subset \mathbb{C}.$$

Note $\mathbb{Q}[\alpha] \subseteq \mathbb{Q}(\alpha)$. The natural question is, then, when does equality hold? We have one answer to this:

**Proposition 8.12**

If $\alpha$ is an algebraic integer, then $\mathbb{Q}[\alpha] = \mathbb{Q}(\alpha)$ and it is a $\mathbb{Q}$-vector space of dimension equal to the degree of $\alpha$.

*Proof.* Let $f$ be the minimal polynomial of $\alpha$. Let $\frac{P(\alpha)}{Q(\alpha)} \in \mathbb{Q}(\alpha)$, where $P, Q \in \mathbb{Q}[X]$ and $\mathbb{Q}(\alpha) \neq 0$. This means $f \nmid Q$, and since $f$ is irreducible, this means $(f, Q) = 1$. Thus, the Euclidean algorithm in $\mathbb{Q}[X]$ tells us that $\exists\, h, k \in \mathbb{Q}[X]$ such that $fg + Qk = 1$. Substituting $X = \alpha$, we get $f(\alpha)h(\alpha) + Q(\alpha)k(\alpha) = 1 \implies k(\alpha) = 1/Q(\alpha)$, so $P(\alpha)/Q(\alpha) = P(\alpha)k(\alpha) \in \mathbb{Q}[\alpha]$. This proves the first part of the statement.

Notice that $\alpha^n = -a_{n-1}\alpha^{n-1} - \cdots - a_0$, so $\mathbb{Q}[\alpha]$ is generated by $1, \alpha, \ldots, \alpha^{n-1}$. These elements are linearly independent: if not, then $\alpha$ would satisfy some polynomial relation of the form

$$b_{n-1}\alpha^{n-1} + \cdots + b_0 = 0$$

where $b_i \in \mathbb{Q}$. But $g(X) = \sum_{i=0}^{n-1} b_i X^i$ satisfies $g(\alpha) = 0$ and $\deg g < \deg f$, which only complies with the minimality of $f$ if $g = 0$, meaning $b_i = 0$ for all $i$. The conclusion follows. □

## 8.4 Quadratic Character of 2

Recall that this entire discussion was to provide another proof of Quadratic Reciprocity. We return to the land of Quadratic Reciprocity now by providing a new proof of Proposition 7.7, which characterizes when 2 is a quadratic residue mod $p$. Recall we had

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

We will prove this statement using roots of unity. In general, the $n^{\text{th}}$ roots of unity are the complex numbers $z \in \mathbb{C}$ satisfying $z^n = 1$. (If you think about it enough, you can see that $z$ must be of the form $e^{2\pi ik/n}$ for some integer $k$.)

*Proof.* Let $\zeta = e^{2\pi i/8}$, so $\zeta^8 = 1$ and $\zeta$ is a (primitive) eighth root of unity. We can factor $\zeta^8 - 1 = (\zeta^4 - 1)(\zeta^4 + 1)$; since $\zeta^4 - 1 = \neq 0$, we have $\zeta^4 + 1 = 0 \implies \zeta^4 = -1$. Dividing by $\zeta^2$ on both sides, we get $\zeta^2 + 1/\zeta^2 = 0$. Observe, then, that

$$\left(\zeta + \frac{1}{\zeta}\right)^2 = \zeta^2 + 2 + \frac{1}{\zeta^2} = 2.$$

Let $\tau = \zeta + \zeta^{-1}$. Note $\zeta \in \Omega$ since $\zeta^8 - 1 = 0$, and $\tau \in \Omega$ from $\tau^2 - 2 = 0$ from above. Let $p$ be an odd prime. Using $\left(\frac{a}{p}\right) = a^{\frac{p-1}{2}}$, we have

$$\tau^{p-1} = (\tau^2)^{\frac{p-1}{2}} = 2^{\frac{p-1}{2}} \equiv \left(\frac{2}{p}\right) \pmod{p},$$

so $\tau^p = \left(\frac{2}{p}\right)\tau \pmod{p}$. Since $\tau \in \Omega$, we apply Freshman's Dream (Proposition 8.11) to see $\tau^p = (\zeta + \zeta^{-1})^p \equiv \zeta^p + \zeta^{-p} \pmod{p}$. But $\zeta^8 = 1$, so in fact we have

$$\zeta^p + \zeta^{-p} = \begin{cases} \zeta + \zeta^{-1} & p \equiv \pm 1 \pmod 8 \\ \zeta^3 + \zeta^{-3} & p \equiv \pm 3 \pmod 8. \end{cases}$$

We can write these sums in terms of $\tau$. The first one is just $\tau = \zeta + \zeta^{-1}$ by definition, and the second can be written as $\zeta^3 + \zeta^{-3} = -\zeta^{-1} + (-\zeta^{-1})^{-1} = -(\zeta + \zeta^{-1}) = -\tau$ since $\zeta^4 = -1 \implies \zeta^3 = -\zeta^{-1}$. To summarize,

$$\tau^p = \zeta^p + \zeta^{-p} = \begin{cases} \tau & p \equiv \pm 1 \pmod 8 \\ -\tau & p \equiv \pm 3 \pmod 8. \end{cases} = (-1)^{\varepsilon}\tau,$$

where $\varepsilon = \frac{p^2-1}{8}$.

Now we put all of this together. From above, we have

$$(-1)^{\varepsilon}\tau \equiv \tau^p \equiv \left(\frac{2}{p}\right)\tau \pmod{p}.$$

Multiplying by $\tau$ on both sides gives us factors of $\tau^2 = 2$ on the left and the right. But since $p$ is odd, $2 \nmid p$, so we can cancel out the 2's on both sides to be left with $(-1)^{\varepsilon} \equiv \left(\frac{2}{p}\right) \pmod{p}$. But if both $(-1)^{\varepsilon}$ and the Legendre symbol only take on values of $\pm 1$, then this equivalence mod $p$ translates to equivalence as integers. This concludes the proof. □

I think this proof is a bit nicer than the proof we provided previously for this result, since it gives a little more explanation as to why considering $p$ mod 8 is the correct thing to do. Before, we were like "oh, haha, if you look at $p$ mod 8 and go through all the cases then it works, what a happy coincidence har har har" but here the 8 comes naturally from this whole business with considering the eighth root of unity $\zeta$.

# 9 10/06 - Quadratic Gauss Sums

Recall our marvelous work from the end of last class where we took an eighth root of unity $\zeta = e^{2\pi i/8}$ and defined $\tau = (\zeta + \zeta^{-1})^2$. Using these values, we proved the result $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$. We will do something in greater generality in order to achieve Quadratic Reciprocity once more.

Let $p$ be an odd prime. We will now let $\zeta = e^{2\pi i}p$, so $\zeta$ is a $p^{\text{th}}$ root of unity. In particular, $\zeta^p - 1 = 0$.

**Lemma 9.1**

Let $a \in \mathbb{Z}$. then,

$$\sum_{t=0}^{p-1} \zeta^{at} = \begin{cases} p & \text{if } p \mid a \\ 0 & (p,a) = 1. \end{cases}$$

*Proof.* We first consider the case when $p \mid a$. Then, since $\zeta^p = 1$, it follows that $\zeta^a = 1$ as well, and so the result follows.

Now suppose $p \nmid a$. This is similarly easy: this sum is just the sum of a finite geometric series with starting term 1 and common ratio $\zeta$, so we have

$$\sum_{t=0}^{p-1} \zeta^{at} = \frac{(\zeta^a)^p - 1}{\zeta^a - 1} = 0$$

again because $(\zeta^a)^p = (\zeta^p)^a = 1$. □

**Corollary 9.2**

Let $x, y \in \mathbb{Z}$. Then,

$$\frac{1}{p}\sum_{t=0}^{p-1} \zeta^{t(x-y)} = \begin{cases} 0 & \text{if } p \nmid x - y \\ 1 & \text{if } x \equiv y \pmod{p}. \end{cases}$$

Now, we will prove a useful lemma regarding sums of the Legendre symbol.

**Lemma 9.3**

$\sum_{t=0}^{p-1} \left(\frac{t}{p}\right) = 0$.

*Proof.* Ha syke, this is a homework problem. (Proof was given in class, though, so come to class!) The idea, though, is to count the number of times $\left(\frac{t}{p}\right) = \pm 1$ respectively. □

## 9.1 Gauss Sum

We now introduce the main ingredient in this approach towards Quadratic Reciprocity: the Gauss sum.

**Definition 9.4** (Gauss Sum)**.** Let $a \in \mathbb{Z}$. The sum

$$g_a = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right) \zeta^{at}$$

is called a **Gauss sum**.

We now prove some fundamental properties of this Gauss sum.

**Proposition 9.5**

$g_a = \left(\frac{a}{p}\right) g_1$.

*Proof.* If $p \mid a$, then $\zeta^a = 1$, so $g_a = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right) = 0$ by Lemma 9.3. Now, assume that $p \nmid a$. We have an isomorphism from $\mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ where $t \mapsto at$, so we can change variables to get

$$\begin{aligned}
\left(\frac{a}{p}\right) g_a &= \sum_{t=0}^{p-1} \left(\frac{a}{p}\right)\left(\frac{t}{p}\right) \zeta^{at} \\
&= \sum_{t=0}^{p-1} \left(\frac{at}{p}\right) \zeta^{at} \\
&= \sum_{s=0}^{p-1} \left(\frac{s}{p}\right) \zeta^{s} = g_1, \\
\implies g_a &= \left(\frac{a}{p}\right)\left(\frac{a}{p}\right) g_a = \left(\frac{a}{p}\right) g_1,
\end{aligned}$$

as desired. □

**Author's Note 9.6.** Some notation remarks. We will let $g = g_1$, and I will denote $\mathbb{Z}/p\mathbb{Z}$ as $\mathbb{F}_p$ (the $\mathbb{F}$ stands for "field" because this is a field with $p$ elements). Whenever I suppress the indices of a sum (e.g., if I just write $\sum_x$), then assume $x$ goes from 0 to $p-1$.

Moving forward,

**Proposition 9.7**
$g^2 = (-1)^{\frac{p-1}{2}} p$.

*Proof.* Let $a \not\equiv 0 \pmod{p}$. Then,

$$\begin{aligned} g_a g_{-a} &= \left(\frac{a}{p}\right)\left(\frac{-a}{p}\right) g^2 \\ &= \left(\frac{-1}{p}\right) g^2 \\ &= (-1)^{\frac{p-1}{2}} g^2. \end{aligned}$$

Now taking the sum over all $a$, we have

$$\begin{aligned} (p-1)\cdot(-1)^{\frac{p-1}{2}} g^2 &= \sum_{a=0}^{p-1} g_a g_{-a} \\ &= \sum_{a=0}^{p-1} \sum_x \sum_y \left(\frac{x}{p}\right)\left(\frac{y}{p}\right) \zeta^{a(x-y)} \\ &= \sum_{0\le x,y<p} \left(\frac{xy}{p}\right) \sum_{a=0}^{p-1} \zeta^{a(x-y)} \\ &= \sum_x p\left(\frac{x^2}{p}\right) = p(p-1), \end{aligned}$$

from which we conclude our desired result. The second-to-last equality follows from Corollary 9.2. □

We will denote $(-1)^{\frac{p-1}{2}} p = p^*$ to make things less clunky. So, to rewrite the above Proposition, $g^2 = p^*$.

## 9.2 Second Proof of Quadratic Reciprocity

Quadratic Reciprocity relates two primes, so let's now introduce $q \neq p$ another odd prime. We begin our second proof of Quadratic Reciprocity.

*Proof of Quadratic Reciprocity, Theorem 7.8.* Let $q \neq p$ be another odd prime. We then have

$$\begin{aligned} g^{q-1} &= (g^2)^{\frac{q-1}{2}} = (p^*)^{\frac{q-1}{2}} \\ &\equiv \left(\frac{p^*}{q}\right) \pmod{q}, \\ \implies g^q &= \left(\frac{p^*}{q}\right) g \pmod{q}. \end{aligned}$$

On the other hand, we can use the definition of $g$ to explicitly expand (using Freshman's Dream, which states $(a+b)^q \equiv a^q + b^q \bmod q$):

$$g^q = \left(\sum_t \left(\frac{t}{p}\right)\zeta^t\right)^q \equiv \sum_t \left(\frac{t}{p}\right)^q \zeta^{qt} \equiv g_q \pmod q,$$

where the last equality follows because $\left(\frac{t}{p}\right)^q = \left(\frac{t}{p}\right)$ since the symbol is either $-1, 0, 1$ and $q$ is odd.

But Proposition 9.5 tells us that $g_q \equiv \left(\frac{q}{p}\right) g \bmod q$, so $g^q \equiv g_q \equiv \left(\frac{q}{p}\right) g$. At the same time, we wrote above that $g^q \equiv \left(\frac{p^*}{q}\right) g$, so we have

$$\begin{aligned}
&\left(\frac{p^*}{q}\right) g \equiv \left(\frac{q}{p}\right) g \pmod q \\
\implies &\left(\frac{p^*}{q}\right) g \cdot g \equiv \left(\frac{q}{p}\right) g \cdot g \\
\implies &\left(\frac{p^*}{q}\right) p^* \equiv \left(\frac{q}{p}\right) p^* && \text{(Proposition 9.7)} \\
\implies &\left(\frac{q}{p}\right) \equiv \left(\frac{p^*}{q}\right) \\
&= \left(\frac{(-1)^{\frac{p-1}{2}} p}{q}\right) && \text{(again Prop 9.7)} \\
&= \left(\frac{-1}{q}\right)^{\frac{p-1}{2}} \left(\frac{p}{q}\right) \\
&\equiv (-1)^{\frac{q-1}{2}\cdot\frac{p-1}{2}} \left(\frac{p}{q}\right), && \text{(Corollary 7.6)}
\end{aligned}$$

which is exactly the statement for Quadratic Reciprocity. □

## 9.3 Kronecker's Result for Quadratic Extensions

We now shift our focus back to these Gauss sums. Our end goal for this section will be to prove the following:

$$g^2 = p^* = \begin{cases} p & p \equiv 1 \pmod 4 \\ -p & p \equiv 3 \pmod 4, \end{cases}$$

or equivalently

$$g = \begin{cases} \pm\sqrt{p} & p \equiv 1 \pmod 4 \\ \pm i\sqrt{p} & p \equiv 3 \pmod 4. \end{cases}$$

We step back a little bit and, even before concerning ourselves with specific values, we ask: what is the sign of $g$?

**Proposition 9.8**

Let $P(X) = X^{p-1} + X^{p-2} + \cdots + 1 = \frac{X^p - 1}{X - 1}$. Then, $P$ is an irreducible polynomial in $\mathbb{Q}[X]$.

**Remark 9.9.** We call $P(X)$ above the $p^{\text{th}}$ **cyclotomic polynomial**. Note also that if we take $\zeta = e^{2\pi i} p$ again, then $P(\zeta) = 0$, so $P$ is the minimal polynomial of $\zeta$ by the Proposition, as it is irreducible.

*Proof.* By contradiction, assume that $P$ is reducible. Then, we can write $P(X) = f(X)g(X)$ where $f, g \in \mathbb{Q}[X]$ are monic and $\deg f, \deg g > 0$. We now have a really strong result:

**Exercise 9.10.** For such $f, g \in \mathbb{Q}[X]$ above, it follows that $f, g \in \mathbb{Z}[X]$.

This is Exercise 4 in Chapter 6 of Ireland-Rosen, but one way to approach this is noting that (1) the coefficients of $f$ and $g$ are polynomials in $\zeta^k$ (these are the roots of $P(X)$), (2) each $\zeta^k$ is an algebraic integer, and the set of algebraic integers $\Omega$ forms a ring, and (3) $\Omega \cap \mathbb{Q} = \mathbb{Z}$.

Using this exercise, though, we can write $P(X+1) = f(X+1)g(X+1)$, but we can explicitly write $P(X+1)$ as

$$P(X+1) = \frac{(X+1)^p - 1}{X} = \sum_{k=1}^{p} \binom{p}{k} X^{k-1}.$$

Taking this modulo $p$, we have $X^{p-1} \equiv f(X+1)g(X+1) \bmod p$, which means in mod $p$, we have $f(X+1) \equiv X^r$ and $g(X+1) \equiv X^s$ for some $r, s > 0$. Taking $X = 0$, we see that $p \mid f(1), g(1)$, so $p^2 \mid f(1)g(1) = P(1) = p$, which is a contradiction. The conclusion follows. □

**Proposition 9.11**

Taking $\zeta = e^{2\pi i/p}$ again,[a]

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

[a] Apparently Gauss thought about this almost every day for four years before being able to prove it. Look at Gauss, man, so inspirational.

*Proof.* Consider again $P(X)$ from the above Proposition; since every $\zeta^t$ is a root of $P$, we can write $P(X) = X^{p-1} + \cdots + 1 = \prod_{t=1}^{p-1}(X - \zeta^t)$. Plugging in $X = 1$, we have

$$p = P(1) = \prod_{t=1}^{p-1}(1 - \zeta^t).$$

Now we do something that is a bit ad hoc, which is only reasonable if it took a chad like Gauss four years to come up with this. Observe that the set $\left\{\pm(4k-2) : 1 \leq k \leq \frac{p-1}{2}\right\}$ is a complete set of residues mod $p$. This is the case because the expression $4k - 2$, where we take the values $1 \leq k \leq \frac{p+1}{4}$, gives the 2 mod 4 residues $2, 6, \ldots,$, while those greater than $\frac{p+1}{4}$ give the 3 mod 4 residues. Taking the negatives of everything covers the rest.

Given this, though, we can now rewrite our product, splitting along when we take $+(4k-2)$ or $-(4k-2)$:

$$\begin{aligned} p &= \prod_{k=1}^{\frac{p-1}{2}}(1 - \zeta^{4k-2}) \prod_{k=1}^{\frac{p-1}{2}}(1 - \zeta^{-(4k-2)}) \\ &= \prod_{k=1}^{\frac{p-1}{2}} \zeta^{2k-1}\left(\zeta^{-(2k-1)} - \zeta^{2k-1}\right)\zeta^{-(2k-1)}\left(\zeta^{2k-1} - \zeta^{-(2k-1)}\right) \\ &= (-1)^{\frac{p-1}{2}} \prod_{k=1}^{\frac{p-1}{2}} \left(\zeta^{2k-1} - \zeta^{-(2k-1)}\right)^2. \end{aligned}$$

Rearranging gives the desired result. □

This essentially takes us to the promised land.

**Proposition 9.12**

Let $p$ be an odd prime, and $\zeta = e^{2\pi i/p}$. Then,

$$\prod_{k=1}^{\frac{p-1}{2}} \left(\zeta^{2k-1} - \zeta^{-(2k-1)}\right) = \begin{cases} \sqrt{p} & p \equiv 1 \pmod 4 \\ i\sqrt{p} & p \equiv 3 \pmod 4. \end{cases}$$

*Proof.* From the above Proposition, we know that the magnitude of the product is going to be $\sqrt{p}$, so we are only concerned about the sign/what power of $i$ we multiply.

We need to get a little bit more involved in the complex numbers this time. You may have seen before $e^{i\theta} = \cos\theta + i\sin\theta$; thus, we have $e^{ix} - e^{-ix} = 2i \cdot \sin(x)$, so we

have

$$\begin{aligned}\prod_{k=1}^{\frac{p-1}{2}}\left(\zeta^{2k-1}-\zeta^{-(2k-1)}\right) &= \prod_{k=1}^{\frac{p-1}{2}}\left(e^{\frac{2i\pi(2k-1)}{p}}-e^{-\frac{2i\pi(2k-1)}{p}}\right)\\ &= \prod_{k=1}^{\frac{p-1}{2}} 2i\sin\left(\frac{(4k-2)\pi}{p}\right)\\ &= i^{\frac{p-1}{2}}\prod_{k=1}^{\frac{p-1}{2}} 2\sin\left(\frac{(4k-2)\pi}{p}\right).\end{aligned}$$

Let's look at when sin is negative. In general, $\sin x$ is negative when $\pi < x < 2\pi$, so $\sin\frac{(4k-2)\pi}{p}$ is negative if $\frac{p+2}{4} < k \leq \frac{p-1}{2}$. Counting by case, this gives $\frac{p-1}{4}$ negative terms if $p \equiv 1 \pmod 4$ and $\frac{p-3}{4}$ negative terms if $p \equiv 3 \pmod 4$.

Now we proceed by case mod 4. If $p \equiv 1 \pmod 4$, then the sign of the product is $i^{\frac{p-1}{2}}(-1)^{\frac{p-1}{4}} = (-1)^{\frac{p-1}{4}}(-1)^{\frac{p-1}{4}} = (-1)^{\frac{p-1}{2}} = 1$. This agrees with what we have in our proposition statement.

If $p \equiv 3 \pmod 4$, then the sign is equal to

$$i^{\frac{p-1}{2}}(-1)^{\frac{p-3}{4}} = i(-1)^{\frac{p-3}{4}}(-1)^{\frac{p-3}{4}} = i,$$

which is also what we have in the proposition. This concludes the proof. □

We have seem to gone a long ways away from our temporary home of Gauss sums, but now we have found a road to circle back around. Recall Proposition 9.7 tells us that $g^2 = (-1)^{\frac{p-1}{2}}p$. But this is exactly the expression we have from Proposition 9.11, so we have

$$\begin{aligned}g^2 = (-1)^{\frac{p-1}{2}}p &= \prod_{k=1}^{\frac{p-1}{2}}\left(\zeta^{2k-1}-\zeta^{-(2k-1)}\right)^2\\ \implies g &= \varepsilon\prod_{k=1}^{\frac{p-1}{2}}\left(\zeta^{2k-1}-\zeta^{-(2k-1)}\right).\end{aligned}$$

You are probably on the edge of your seat at this point whether $\varepsilon$ is positive or negative. Kronecker gives us the answer.

**Theorem 9.13** (Kronecker)
$\varepsilon = 1$.

*Proof.* Let

$$f(x) = \sum_{j=1}^{p-1} \left(\frac{j}{p}\right) x^j - \varepsilon \prod_{k=1}^{\frac{p-1}{2}} \left(x^{2k-1} - x^{p-(2k-1)}\right).$$

Note that when $x = \zeta$, we are just computing $f(\zeta) = g - g = 0$. One can also use Lemma 9.3 to deduce $f(1) = 0$. This means that $f$ is divisible by the minimal polynomials of $\zeta$ and 1, respectively; in particular, $X^{p-1} + \cdots + 1 \mid f$ and $X - 1 \mid f$. This means

$$(X-1)(X^{p-1} + \cdots + 1) = X^p - 1 \mid f,$$

so we can write $f(X) = (X^p - 1)g(X)$ for some $g(X)$. Substituting $x = e^z$ in the above expression, we have

$$\sum_{j=1}^{p-1} \left(\frac{j}{p}\right) e^{jz} - \varepsilon \prod_{k=1}^{(p-1)/2} \left(e^{z(2k-1)} - e^{z(p-(2k-1))}\right) = (e^{pz} - 1)g(e^z). \tag{2}$$

We can identify $e^{jz}$ with its Taylor series expansion $\sum_{k=0}^{\infty} \frac{(jz)^k}{k!}$, in which case the sum can be expressed as

$$\sum_{j=1}^{p-1} \left(\frac{j}{p}\right) e^{jz} = \sum_{k=0}^{\infty} \frac{z^k}{k!} \sum_{j=1}^{p-1} \left(\frac{j}{p}\right) j^k.$$

We will now identify the $z^{(p-1)/2}$ coefficient in Equation 2. The sum contributes a coefficient of $\frac{1}{((p-1)/2)!} \sum_{j=1}^{p-1} \left(\frac{j}{p}\right) j^{\frac{p-1}{2}}$. But note that mod $p$, we have $j^{\frac{p-1}{2}} \equiv \left(\frac{j}{p}\right)$, so the coefficient mod $p$ reduces to

$$\frac{1}{((p-1)/2)!} \sum_{j=1}^{p-1} \left(\frac{j}{p}\right)^2 = \frac{p-1}{\left(\frac{p-1}{2}\right)!}.$$

Meanwhile, if we consider the Taylor series of the difference $e^{z(2k-1)} - e^{z(p-(2k-1))}$, the constant term of 1 cancels out so the minimal term is $z(2k-1) - z(p-(2k-1))$. Taking the product as $1 \le k \le (p-1)/2$, we see that the coefficient of $z^{(p-1)/2}$ in the product must be the product of the linear part of each term, namely

$$\prod_{k=1}^{(p-1)/2} ((2k-1) - (p-(2k-1))) = \prod_{k=1}^{(p-1)/2} (4k - p - 2).$$

Reducing this mod $p$ as well, we have

$$\begin{aligned}\prod_{k=1}^{(p-1)/2}(4k-p-2) &\equiv \prod_{k=1}^{(p-1)/2}(4k-2) = 2^{(p-1)/2}\prod_{k=1}^{(p-1)/2}(2k-1)\\ &= \frac{2^{(p-1)/2}(p-1)!}{2\cdot 4\cdots(p-1)} = \frac{2^{(p-1)/2}(p-1)!}{2^{(p-1)/2}\left(\frac{p-1}{2}\right)!}\\ &\equiv -\frac{1}{\left(\frac{p-1}{2}\right)!} \pmod p,\end{aligned}$$

where the last line follows from Wilson's Theorem.

We now consider the coefficient of $z^{(p-1)/2}$ for $(e^{pz}-1)g(e^z)$. Note that $f(X)\in\mathbb{Z}[X]$, so $g(X)\in\mathbb{Z}[X]$ as well. But all coefficients of $e^{pz}-1=\sum_{k\geq 1}(pz)^k/k!$ are 0 mod $p$, hence all coefficients in the expansion of $(e^{pz}-1)g(e^z)$ vanish mod $p$.

Collecting our results, this means the coefficient of $z^{(p-1)/2}$ for the sum and product must agree mod $p$, i.e.,

$$\frac{p-1}{\left(\frac{p-1}{2}\right)!} \equiv -\varepsilon\frac{1}{\left(\frac{p-1}{2}\right)!} \pmod p.$$

It is clear now that $\varepsilon=+1$. □

# 10 10/13 - Finite Fields

We now study a really beautiful and fruitful topic: finite fields. We are quite familiar with perhaps the simplest kind of finite field: $\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}$, the field with $p$ elements. One can show that any two fields with $p$ elements are isomorphic to each other. This gives us a way of classifying all finite fields of a certain prime order – namely, there is only one up to isomorphism. Looking outward at all possible sizes, we pose the question

Can we classify all finite fields?

Just as a note, a field is a set closed under addition and multiplication, and it has the important property that every nonzero element has a multiplicative inverse. For example, $\mathbb{Q}$ and $\mathbb{Z}/7\mathbb{Z}$ are fields because, for instance, $2^{-1}=1/2\in\mathbb{Q}$ and $2^{-1}=4$ in $\mathbb{Z}/7\mathbb{Z}$ (since $2\cdot 4\equiv 1 \pmod 7$), but $2^{-1}=1/2\notin\mathbb{Z}$ so $\mathbb{Z}$ is not a field.

We first provide a result that tells us exactly what the size of a finite field can be.

**Lemma 10.1** (Finite fields have prime power order)
If $k$ is a finite field, then $\mathbb{F}_p\subseteq k$ is a subfield for some prime $p$ and $|k|=p^r=q$ for some $r\in\mathbb{Z}$. (In this case, we often write $k=\mathbb{F}_q$.)

*Proof.* As $k$ is finite, the set under addition is a finite abelian group, so for any $x$,

$$|k| \cdot x = \underbrace{x + \cdots + x}_{|k| \text{ times}} = 0.$$

Denote $q \coloneqq |k|$. Let $m$ be the minimal positive integer such that $m \cdot 1 = 0$. We claim that $m$ is prime. If $m$ is not prime, then we can write $m = a \cdot b$ for some integers $a, b > 1$. But then

$$\begin{aligned} m \cdot 1 &\coloneqq \underbrace{1 + \cdots + 1}_{m \text{ times}} \\ &= \underbrace{1 + \cdots + 1}_{a \text{ times}} \underbrace{1 + \cdots + 1}_{b \text{ times}} \\ &= (a \cdot 1)(b \cdot 1), \end{aligned}$$

and in general if $rs = 0$ in a field, then either $r = 0$ or $s = 0$. But then we have either $a \cdot 1 = 0$ or $b \cdot 1 = 0$, which contradicts the minimality of $m$. Thus, $m = p$ is a prime, so $k$ contains any $\underbrace{1 + \cdots + 1}_{\ell \text{ times}}$ for $\ell \leq p$, i.e. it contains $\mathbb{F}_p$.

The fact that $|k|$ is some power of $p$ comes from the fact that $k$ is an $\mathbb{F}_p$-vector space. We can check the axioms of a vector space manually: given $x, y \in k$ and $a, b \in \mathbb{F}_p$, we have (1) $a \cdot x$ is just multiplication in $k$; (2) $a \cdot (x+y) = a \cdot x + a \cdot y$; (3) $a(b \cdot x) = (a \cdot b) \cdot x$; (4) $(a + b) \cdot x = a \cdot x + b \cdot x$.[8]

Now, thinking of $k$ as an $\mathbb{F}_p$-vector space, we will find an $\mathbb{F}_p$-basis for $k$. In other words, we will find a minimal set of elements $r_1, \ldots, r_n \in k$ such that any $a \in k$ can be written as $a = a_1 r_1 + \cdots + a_n r_n$ for some $a_i \in \mathbb{F}_p$. We can show that the minimality of this set implies these $a_i$ are unique.[9] If $a = a_1 r_1 + \cdots + a_n r_n = b_1 r_1 + \cdots + b_n r_n$ with not all $a_i = b_i$ (WLOG rearrange basis elements so that $a_1 \neq b_1$), then we have

$$\begin{aligned} 0 &= (a_1 - b_1) r_1 + (a_2 - b_2) r_2 + \cdots + (a_n - b_n) r_n \\ (b_1 - a_1) r_1 &= (a_2 - b_2) r_2 + \cdots + (a_n - b_n) r_n \\ \implies r_1 &= (b_1 - a_1)^{-1} ((a_2 - b_2) r_2 + \cdots + (a_n - b_n) r_n), \end{aligned}$$

which is legal because we assumed $a_1 \neq b_1 \implies b_1 - a_1 \neq 0$ but it is an element of $\mathbb{F}_q$, so there exists an inverse. But then we can express $r_1$ in terms of $r_2, \ldots, r_n$, so the set $r_1, \ldots, r_n$ is no longer minimal. Thus, the $a_i$'s are unique.

This provides us with a bijection of sets (in fact, an isomorphism of additive groups) $k \cong \underbrace{\mathbb{F}_p \times \cdots \times \mathbb{F}_p}_{n \text{ times}}$ where $a \mapsto (a_1, \ldots, a_n)$.[10] Thus, $|k| = |\mathbb{F}_p \times \cdot \times \mathbb{F}_p| = p^n$ as desired.

□

[8] Honestly I'm not sure if this covers all the vector space axioms, but the main point is that everything is sunshine and rainbows because $\mathbb{F}_p$ literally lives in $k$, so $k$ is an $\mathbb{F}_p$-vector space.

[9] This is just saying basis elements are linearly independent, for those who have seen this stuff before.

[10] Note this is not a field isomorphism; in fact, $\mathbb{F}_p \times \cdots \times \mathbb{F}_p$ not only fails to be a field, but it fails to be an integral domain. Convince yourself of this!

This is really great news and a strong result off the bat. However, this tells us practically nothing else about the field $k$. The only elements we have a hold on is the copy of $\mathbb{F}_p$ living inside $k$, but otherwise we are a bit lost. (What even are the basis elements of $k$?) We will now prove some things which tell us more information about $\mathbb{F}_q$ (where $q = p^n$).

**Lemma 10.2**

Suppose $k$ is a finite field with $q = p^n$ elements. Then, the polynomial $X^q - X \in \mathbb{F}_p[X]$ factors as

$$X^q - X = \prod_{\alpha \in k} (X - \alpha).$$

*Proof.* Note $k^\times = k \setminus \{0\}$ has size $q - 1$. Since $k^\times$ is itself a multiplicative group, if $y \in k^\times$, then $y^{q-1} = 1$. Equivalently, $\alpha^q - \alpha = 0$ for any $\alpha \in k$. Thus, $\alpha$ is a root of $X^q - X = 0$, so we have $(X - \alpha) \mid X^q - X$. All of the $(X - \alpha)$ terms are relatively prime, so it follows that

$$\prod_{\alpha \in k} (X - \alpha) \Bigg| X^q - X.$$

But both the left and right hand sides have degree $q$, so they must be equal. □

Whenever we talk about a new object, we always care about maps related to the object. Finite fields have a really, really useful map that comes with them, called the **Frobenius automorphism**. An automorphism is just a bijection which is also a (field) homomorphism: if $\sigma$ is an automorphism, then $\sigma$ is a bijection and it satisfies both $\sigma(x + y) = \sigma(x) + \sigma(y)$ and $\sigma(xy) = \sigma(x)\sigma(y)$.

**Lemma 10.3** (Frobenius automorphism)

Let $k$ be a finite field with $q = p^n$ elements. Then, the map

$$\begin{aligned} \sigma : k &\to k \\ \alpha &\mapsto \alpha^p \end{aligned}$$

is an automorphism, called the Frobenius automorphism. Furthermore, the set of elements fixed by $\sigma$, given by $k^\sigma = \{x \in k \mid \sigma(x) = x\}$, is exactly $\mathbb{F}_p \subset k$.

*Proof.* We first verify that $\sigma$ is a homomorphism. We easily observe $\sigma(xy) = (xy)^p = x^p y^p = \sigma(x)\sigma(y)$. We also have

$$\begin{aligned} \sigma(x + y) = (x + y)^p &= x^p + \binom{p}{1} x^{p-1} y + \cdots + \binom{p}{p-1} x y^{p-1} + y^p \\ &= x^p + y^p = \sigma(x) + \sigma(y), \end{aligned}$$

where the second line follows from the fact $p \mid \binom{p}{m}$ for $1 \leq m \leq p-1$, so all of the middle terms vanish in $k$.

Now we show $\sigma$ is injective. It suffices to show that $\sigma$ is injective, because then it shows $|\operatorname{Im}\sigma| = |k|$, but the range of $\sigma$ is just $k$, so $\operatorname{Im}\sigma = k$, i.e. $\sigma$ is also surjective. If $\sigma(x) = \sigma(y)$, then $\sigma(x-y) = (x-y)^p = 0$. But this is only possible if $x - y = 0$, i.e. $x = y$.

Finally, we show $k^\sigma = \mathbb{F}_p$. If $\beta \in k^\sigma$, then by definition $\sigma(\beta) - \beta = \beta^p - \beta = 0$. But we know any $\alpha \in \mathbb{F}_p$ satisfies $X^p - X = 0$ by Fermat's Little Theorem, so as in the above lemma, we have

$$\prod_{\alpha \in \mathbb{F}_p} (X - \alpha) | X^p - X.$$

Comparing degrees, we see that elements of $\mathbb{F}_p$ are the only roots of $X^p - X$, so $\beta \in \mathbb{F}_p$ as desired. □

## 10.1 Construction of Finite Fields

Now we work towards constructing these finite fields, since again, we have no hold on its elements yet besides the elements of $\mathbb{F}_p$. This discussion will develop in conjunction with working towards the following result:

**Theorem 10.4** (Finite fields of order $q$ are unique)
Let $q = p^n$. Then, there exists a field $k$ with $q$ elements, and $k$ is unique up to isomorphism.

Note that because of the existence of the Frobenius automorphism, $k$ is *not* unique up to canonical isomorphism. However, it turns out that any automorphism of $k$ is simply some power of the Frobenius automorphism. (This is why we say the Frobenius automorphism is really important.) This is a taste of a beautiful study called *Galois theory*, which on the ground is about these automorphisms of fields but can be used to prove wildly vast things! (In fact, the study arose from proving that there is no quintic formula.)

We now proceed with the construction of a finite field with $q$ elements. Let $k$ be a field and $f(x) = k[x]$ is an irreducible monic polynomial. We will consider $f[x]$ modulo $f$: we say $g, h \in k[x]$ are equivalent (denoted $g \sim_f h$) if $f \mid g - h$. (Think of this as just $g \equiv h \pmod{f}$.) This now gives us a construction:

**Proposition 10.5**
The set of equivalence classes of $k[x]$ under the equivalence relation $\sim_f$ is a field.

*Proof.* We first show that the set has addition and multiplication. I will also suppress

the $f$ in $\sim_f$ so I can type faster lol. If $h_1 \sim g_1$ and $h_2 \sim g_2$, then $f \mid h_1 - g_1$ and $f \mid h_2 - g_2$, so $f \mid (h_1 + h_2) - (g_1 + g_2) \implies h_1 + h_2 \sim g_1 + g_2$. Similarly,

$$f \mid h(h_2 - g_2) + g_2(h_1 - g_1) = h_1h_2 - g_1g_2 \implies h_1h_2 \sim g_1g_2.$$

Now we show that the set is a field. Take some nonzero $g \in k[x]/\sim$. Since $f$ is irreducible, this just means $f \nmid g$. Consider the ideal $(f, g) = \{af + bg \mid a, b \in k[x]\}$. Recall from the first few lectures (somewhere around Theorem 2.4) that $k[x]$ is a principal ideal domain (a GCD exists between any two elements), so $(f, g) = (h)$ for some $h \in k[x]$. But this means $f \in (h)$, or equivalently $h \mid f$, which is only possible given $f$ is irreducible when $\deg h = 0$ or $h = f$. If $h = f$, then $g \in (h) \implies f = h \mid g$, a contradiction to our choice of $g$. Thus, $\exists\, a, b \in k[x]$ such that $af + bg = 1$, so $bg = 1 - af \sim 1$. This shows $k[x]/\sim$ is a field.

(If this was a lot to handle, this is the tl;dr of the argument. If $g$ lives in a nonzero equivalence class, then since $f$ is irreducible, $f$ and $g$ are coprime, so their gcd is 1. That means we have some polynomials $a, b$ such that $af + bg = 1$, so $bg \equiv 1 \pmod f$, i.e. $b$ is the inverse of $g$.) □

Okay, we constructed a field. What does this field look like? How many elements does it have?

**Corollary 10.6**

If $f$ has degree $d$, then $k[X]/\sim_f$ is a vector space of $k$ with dimension $d$, with a basis given by $1, X, X^2, \ldots, X^{d-1}$.

*Proof.* Since $k[X]$ is a Euclidean domain, it has a Division Algorithm. Thus, for any $g \in k[X]$, we can write $g = q \cdot f + r$ for some $q, r \in k[X]$ and $\deg r < \deg f$. This means $g \sim_f r$, but $r$ is written using just $1, X, \ldots, X^{d-1}$ (as $\deg r < \deg f = d$), so we are done. □

This notation $k[X]/\sim_f$ is a little clunky. Sometimes, we write $k[X]/\sim_f$ as $k(\alpha)$, where $\alpha$ is a "solution" of $f(X) = 0$. I find this a particularly useful perspective, because polynomials are a bit clunky to actually work with, but I can work with numbers. For instance, if we expand our view a little to consider polynomial rings over $\mathbb{Z}$, for instance, then we can consider $f(X) = X^2 + 1$. Then, we could show in a similar fashion that $\mathbb{Z}[X]/\sim_f$ is a ring. But what we're actually doing here is we're adjoining an element $X$ satisfying $X^2 + 1 = 0$. We have another name for such an element! We call it $i = \sqrt{-1}$. So $\mathbb{Z}[X]/\sim_f$ is actually just $\mathbb{Z}[i]$, written in a way where you don't have to write down some element out of thin air.

Another perspective, for those who know stuff about rings and ideals, is that this is actually just the quotient $k[X]/(f)$, where $(f)$ is again the ideal generated by $f$. If $f$ is irreducible, then $(f)$ is maximal, which is equivalent to saying $k[X]/(f)$ is a field. You probably never showed that when $k$ is a finite field, this quotient field is actually finite; this is what we did today.

## 10.2 Existence of $\mathbb{F}_q$

Now we prove another quite remarkable fact. We showed earlier that $X^q - X = 0$ factors into $q$ linear factors in $k[X]$, since every $\alpha \in k$ satisfies $\alpha^q - \alpha = 0$. But what if we just considered the factorization of $X^q - X = 0$ in $\mathbb{F}_p[X]$? A remarkable thing happens:

**Theorem 10.7**

Let $q = p^n$. Then,

$$X^q - X = X^{p^n} - X = \prod_{d|n} F_d(X) \in \mathbb{F}_p[X],$$

where $F_d$ is the product of all monic irreducible polynomials in $\mathbb{F}_p[X]$ of degree $d$.

This gives us access to irreducible polynomials! This may not seem so cool at first, but I challenge you to find an irreducible polynomial of, say, $\mathbb{F}_7[X]$ of degree 4 and see who's laughing by the end of your computation. We care about these irreducible polynomials because, well, we use them to construct our finite fields. In line with this, we have the existence of a finite field of order $p^n$:

**Corollary 10.8**

There exists an irreducible polynomial $f \in \mathbb{F}_p[X]$ of degree $n$.

*Proof.* Assume Theorem 10.7. Let $N_d$ be the number of irreducible polynomials in $\mathbb{F}_p[X]$ of degree $d$. Then, looking at the degrees of both sides of the equation in the Theorem, we have $q = p^n = \sum_{d|n} d \cdot N_d$. We will now apply Möbius inversion to $f(d) = d \cdot N_d$; this gives us

$$f(n) = nN_n = \sum_{d|n} \mu(n/d)p^d.$$

This gives us an explicit form for $N_n$, and one can check that the sum on the right is nonzero (it is a sign of distinct powers of $p$ up to sign), so $N_n \neq 0$ and indeed there exists an irreducible polynomial of degree $n$ in $\mathbb{F}_p[X]$. □

**Corollary 10.9**

There exists a finite field of order $p^n$.

*Proof.* From the above corollary, there is an irreducible polynomial $f \in \mathbb{F}_p[X]$ of degree $n$. Then, $\mathbb{F}_p[X]/\sim_f$ is your desired field. □

We will give a proof for Theorem 10.7 in the next lecture.

# 11 10/16 - Finite Fields, continued

## 11.1 Completing Proof of Existence

As promised, we will now prove Theorem 10.7.

*Proof of Theorem 10.7.* First, we will show that no factor of $X^q - X$ appears twice, namely if $f \in \mathbb{F}_p[X]$ with $\deg f > 0$ such that $f \mid X^{p^n} - X$, then $f^2 \nmid X^{p^n} - X$. Suppose the contrary, so $f^2 \cdot g = X^{p^n} - X$. We will reach a contradiction shortly after a brief interlude on derivatives.

In finite fields, we have a notion of a *formal derivative*. In high school calculus, you learn the derivative via the limit definition, but afterwards, you forget about the limit and just manipulate symbols. For example, you can prove $\frac{d}{dx}x^2 = 2x$ via the limit definition, but you probably know this as just "oh, the rule for the derivative of $x^n$ is just $nx^{n-1}$."

For finite fields, we define the derivative as *just* this symbolic manipulation. So in $\mathbb{F}_p[X]$, we still write $\frac{d}{dx}x^2 = 2x$, except now it doesn't mean anything more than what we just wrote down. Limits don't make sense here, anyways! But this derivative still obeys all the things we expect from calculus (e.g. Chain Rule, Product Rule, etc.), so we can work with this.

To summarize, we have a map $\frac{d}{dx} : \mathbb{F}_p[X] \to \mathbb{F}_p[X]$ where $X^n \mapsto n \cdot X^{n-1}$. Taking the "derivative" on both sides of $f^2 \cdot g = X^{p^n} - X$, we get

$$2f'(X)f(X)g(X) + f^2(X)g'(X) = p^n X^{p^n-1} - 1 = -1,$$

where the last equality follows because we are working over $\mathbb{F}_p$. But $f$ divides the left hand side, so we must have $f \mid -1$, which is our contradiction. Thus, any factor of $X^{p^n} - X$ has multiplicity one.

Now, it remains to show that the only factors of $X^{p^n} - X$ have degree $d \mid n$. In other words, if $f \in \mathbb{F}_p[X]$ is an irreducible monic, then $f \mid X^{p^n} - X$ if and only if $d = \deg f \mid n$. We will approach this via the following lemma:

**Lemma 11.1**

Let $\ell, m \in \mathbb{N}^+$. If $F$ is a field, then $X^\ell - 1 \mid X^m - 1$ in $F[X]$ if and only if $\ell \mid m$. Likewise, if $a \in \mathbb{N}_{>1}$, then $a^\ell - 1 \mid a^m - 1$ if and only if $\ell \mid m$.

*Proof.* If $\ell \mid m$, then clearly $X^\ell - 1 \mid X^m - 1$. We prove the reverse implication. Let

$m = q\ell + r$ where $0 \leq r < \ell$. Then,

$$\begin{aligned}\frac{X^m - 1}{X^\ell - 1} &= \frac{X^{q\ell+r} - 1}{X^\ell - 1} \\ &= X^r \cdot \frac{X^{q\ell} - 1}{X^\ell - 1} + \frac{X^r - 1}{X^\ell - 1}.\end{aligned}$$

Note $X^\ell - 1 \mid X^{q\ell} - 1$ (we have $X^{q\ell} - 1 = (X^\ell)^q - 1 \equiv 1^q - 1 = 0 \pmod{X^\ell - 1}$), so the first term on the right is a polynomial. However, the last term on the right is not a polynomial since $r < \ell$, unless $r = 0$ in which case $\ell \mid m$, as desired.

The proof for the second part of the statement follows from our work above (the proof is exactly the same). □

We return to the problem at hand. Suppose $f \in \mathbb{F}_p[X]$ is a monic irreducible polynomial with $\deg f = d$. Consider the field constructed from $f$, namely $K = \mathbb{F}_p[X]/\sim_f = \mathbb{F}_p(\alpha)$ (where $\alpha$ is a root of $f$, and more specifically $f$ is the minimal polynomial of $\alpha$). We showed last time (Corollary 10.6) that $K$ is a finite field of dimension $d$ over $\mathbb{F}_p$, so $|K| = p^d$. Since $\alpha \in K^\times$ and $|K^\times| = |K| - 1 = p^d - 1$, it follows that $\alpha^{p^d-1} - 1 = 0$, or $\alpha^{p^d} - \alpha = 0$. But $f$ is the minimal polynomial of $\alpha$, so we have $f(X) \mid X^{p^d} - X$.

This sets us up nicely to complete the proof from here. If $d \mid n$, then Lemma 11.1 tells us that $p^d - 1 \mid p^n - 1$; the Lemma again tells us now that $X^{p^d-1} - 1 \mid X^{p^n-1} - 1$, or $X^{p^d} - X \mid X^{p^n} - X$. Thus, since $f$ divides the left hand side, we have $f(X) \mid X^{p^n} - X$.

Now we wish to show the converse, namely $d \mid n$. If $f(X) \mid X^{p^n} - X$, then we can write $X^{p^n} - X = f(X) \cdot g(X)$. Then, $\alpha^{p^n} - \alpha = f(\alpha)g(\alpha) = 0$. Recall (basically Corollary 10.6) that $1, \alpha, \ldots, \alpha^{n-1}$ is an $\mathbb{F}_p$-basis of $K = \mathbb{F}_p(\alpha)$, so any $\beta$ can be written as $\beta = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_{d-1}\alpha^{d-1}$ for some $a_i \in \mathbb{F}_p$. Taking both sides to the $p^n$-power, the Freshman's Dream (which just says $(a+b)^p = a^p + b^p$ in $\mathbb{F}_p$) tells us that

$$\begin{aligned}\beta^{p^n} &= a_0^{p^n} + a_1^{p^n}\alpha^{p^n} + \cdots + a_{d-1}^{p^n}\alpha^{(d-1)p^n} \\ &= a_0 + a_1\alpha + \cdots + a_{d-1}\alpha^{d-1} = \beta,\end{aligned}$$

so $\beta^{p^n} - \beta = 0$ for any $\beta \in \mathbb{F}_p(\alpha)$. But at the same time, Lemma 10.2 tells us that

$$X^{p^d} - X = \prod_{\beta \in K} (X - \beta),$$

so in particular $\beta$ satisfies $\beta^{p^d} - \beta = 0$. But then any $\beta \in K$ satisfies both $X^{p^d} - X = 0$ and $X^{p^n} - X = 0$, and the only roots of the former are elements of $K$, so we have $X^{p^d} - X \mid X^{p^n} - X$. In other words, $X^{p^d-1} - 1 \mid X^{p^n-1} - 1$, in which case Lemma 11.1 tells us that $p^d - 1 \mid p^n - 1$ and hence $d \mid n$. □

The crux of the above proof relies on Lemma 11.1, which basically gives us a nice divisibility criterion on the exponents given divisibility of polynomials.

## 11.2 Uniqueness of $\mathbb{F}_q$

We have almost completed the story of finite fields. We will now complete the promise given by Theorem 10.4, which not only guarantees existence, but says that the field $\mathbb{F}_q$ is unique up to isomorphism. We now prove the uniqueness.

*Proof of uniqueness, Theorem 10.4.* Let $q = p^n$ and suppose $F$ is a finite field of order $q$. We will show that it is isomorphic to $\mathbb{F}_p[X]/\sim_f$ for some monic irreducible $f \in \mathbb{F}_p[X]$ of degree $n$. Note Theorem 10.7 tells us that $f(X) \mid X^q - X$, but $X^q - X = \prod_{\alpha \in F}(X - \alpha)$, so $\exists\, \alpha \in F$ such that $f(\alpha) = 0$. We can now identify $F$ as $\mathbb{F}_p(\alpha)$ via the following isomorphism:

$$\begin{aligned} \mathbb{F}_p[X]/f &\xrightarrow{\simeq} F \\ X &\mapsto \alpha. \end{aligned}$$

One should be more careful than Kisin here by showing this is actually an isomorphism, but if you know the First Isomorphism Theorem, this is not too bad. Since $f$ is irreducible, it is the minimal polynomial of $\alpha$. Thus, the map $\mathbb{F}_p[X] \to F$ sending $X \mapsto \alpha$ is clearly surjective, and its kernel is exactly $f$, so $\mathbb{F}_p[X]/f \to F$ is an isomorphism, as desired. □

**Proposition 11.2**

If $|F| = q = p^n$, then the subfields of $F$ are in bijection with the divisors of $n$.

*Proof.* Let $E \subseteq F$ be a subfield, and denote $d = \dim_{\mathbb{F}_p} E$. We have $E^\times \subseteq F^\times$ as a multiplicative subgroup, which means $|E^\times| = p^d - 1 \mid p^n - 1 \mid |F^\times|$, which means (again, Lemma 11.1) that $d \mid n$.

To construct a subfield given a divisor $d$ of $n$, consider $E = \left\{\alpha \in F \mid \alpha^{p^d} - \alpha = 0\right\}$. In other words, $E$ is the set of solutions to $X^{p^d} - X = 0$, which we know has $p^d$ distinct solutions. (This is true because $X^{p^d} - X \mid X^{p^n} - X$, and the latter has all distinct roots.) We now show $E$ has a field structure: it is closed under addition because $(\alpha + \beta)^{p^d} = \alpha^{p^d} + \beta^{p^d}$, it is closed under multiplication since $(\alpha\beta)^{p^d} = \alpha^{p^d}\beta^{p^d}$, and it has multiplicative inverses because if $\alpha^{p^d} = \alpha$, then we can take the inverses on both sides and it still satisfies $X^{p^d} = X$. □

## 11.3 Interlude: Galois theory preview

Let us give you a little preview of Galois theory, because what we're doing here when talking about subfields is really discussing the simplest scenario in Galois theory. Recall the Frobenius automorphism $\sigma : F \to F$ sending $\alpha \mapsto \alpha^p$. We saw in Lemma 10.3 that the subfield fixed by $\sigma$, notated $F^\sigma$, is just $\mathbb{F}_p$. If $E \subseteq F$ is a subfield with $|E| = p^d$ (we

just showed then that $d \mid n$), then we can identify $E$ is the subfield of $F$ fixed by $\sigma^d$. In short,

$$E = F^{\sigma^d} = \{\alpha \in F \mid \sigma^d(\alpha) = \alpha^{p^d} = \alpha\}.$$

What Galois theory does is it relates these subfields to subgroups of what we call the **Galois group**, which is just the set of automorphisms of $F$ fixing $\mathbb{F}_p$. In the case of $F$ a field over $\mathbb{F}_p$, it turns out that the only automorphisms of $F$ fixing $\mathbb{F}_p$ are powers of the Frobenius automorphism, so $\mathrm{Aut}(F/\mathbb{F}_p) = \{1, \sigma, \ldots, \sigma^{n-1}\} \simeq \mathbb{Z}/n\mathbb{Z}$. (You may see this as $\mathrm{Gal}(F/\mathbb{F}_p)$; this is because $F$ is what we call a Galois extension over $\mathbb{F}_p$. If you're curious to learn more, take Kisin's Math 123 next semester!)

But what are the subgroups of $\mathbb{Z}/n\mathbb{Z}$? They are simply the multiples of $d$ in $\mathbb{Z}/n\mathbb{Z}$ for $d \mid n$. (For instance, $\{0, 3, 6\}$ is a subgroup of $\mathbb{Z}/6\mathbb{Z}$.) The multiples of $d$ correspond to the subgroup $\{1, \sigma^d, \sigma^{2d}, \ldots, \sigma^{n-d}\}$, or the subgroup generated by $\sigma^d$. But we showed that the field fixed by $\sigma^d$ (hence fixed by this subgroup) is $E$! So there is a bijection between subgroups of $\mathrm{Aut}(F/\mathbb{F}_p)$ and subfields of $F/\mathbb{F}_p$ where a subgroup corresponds to the subfield it fixes. This is really nice, because we have a lot of results in group theory that we can now use.

To conclude this discussion on finite fields, we provide the following nice result:

**Lemma 11.3**
If $|F| = q = p^n$, then $F^\times$ is cyclic. (Hence it is isomorphic to $\mathbb{Z}/(q-1)\mathbb{Z}$.)

The proof follows the proof for when $q = p$ is just a prime (Theorem 5.12), so we omit for brevity.

## 11.4 Proof 2.5 of Quadratic Reciprocity

We shift back to Quadratic Reciprocity, using our new information on finite fields. Let $p, q$ be odd primes. We want to prove again Quadratic Reciprocity. Choose some $n \in \mathbb{N}^+$ such that $q^n \equiv 1 \pmod{p}$ (e.g., $n = p - 1$). LEt $F$ be a finite field of order $q^n$, so $|F^\times| = q^n - 1$. By Lemma 11.3 above, $F^\times$ is cyclic; let $\gamma \in F^\times$ be a generator. Denote $\lambda \coloneqq \gamma^{(q^n-1)/p}$, so $\lambda$ has order exactly $p$. Now consider the Gauss sums

$$\tau_a = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right) \lambda^{at}.$$

Denote $\tau \coloneqq \tau_1$ for ease of notation. Recall the following results we proved back in §9: Proposition 9.5, which says $\tau_a = \left(\frac{a}{p}\right)\tau$, and Proposition 9.7, which says $\tau^2 = (-1)^{(p-1)/2}p$. Denote $p^* \coloneqq (-1)^{(p-1)/2} \cdot p$. We now track through the following equivalent statements:

We have $p^*$ is a square mod $q$ if and only if $\tau \in \mathbb{Z}/q\mathbb{Z}$, as $\tau^2 = p^*$ and the square root of $p^*$ is unique up to sign. But this is equivalent to $\tau^q = \tau$, or equivalently

$$\tau = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right) \lambda^t = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right)^q \lambda^{qt} = \sum_{t=0}^{p-1} \left(\frac{t}{p}\right) \lambda^{qt} = \tau_q = \left(\frac{q}{p}\right) \tau,$$

so $\left(\frac{q}{p}\right) = 1$. To recap, we have $\left(\frac{p^*}{q}\right) = 1 \iff \left(\frac{q}{p}\right) = 1$, so their product must be 1. This gives

$$\begin{aligned} \left(\frac{p^*}{q}\right)\left(\frac{q}{p}\right) &= \left(\frac{(-1)^{(p-1)/2}p}{q}\right)\left(\frac{q}{p}\right) = 1 \\ \implies \left(\frac{(-1)^{(p-1)/2}}{q}\right)\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) &= \left(\frac{-1}{q}\right)^{(p-1)/2}\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) \\ &= (-1)^{\frac{p-1}{2}\frac{q-1}{2}}\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = 1, \end{aligned}$$

and Quadratic Reciprocity follows.

# 12 10/23 - Diophantine Equations

We now transition to the next main part of the class: Diophantine equations. As we go along, the equations that we consider may seem ad hoc. And this is a fair reflection of how math has developed over time: big results are a product of literally many, many years of intense thought. Rather, it is surprising that we are able to assign such a nice narrative to the development of mathematics.

## 12.1 Gaussian Integers, a review

We will begin by reviewing the Gaussian integers, given by $\mathbb{Z}[i]$. This is the set $\{a+bi : a, b \in \mathbb{Z}\}$. We can look at the field of fractions of $\mathbb{Z}[i]$, which are elements of the form $\frac{a+bi}{c+di}$. Rationalizing the denominator, we can express this as $a' + b'i$ for $a', b' \in \mathbb{Q}$. Thus, $\operatorname{Frac}\mathbb{Z}[i] = \mathbb{Q}[i]$, which is a 2-dimensional $\mathbb{Q}$-vector space with basis elements 1 and $i$. (Like with $\mathbb{Z}[i]$, we have $\mathbb{Q}[i] = \{r + is : r, s \in \mathbb{Q}\}$.)

We will review that $\mathbb{Z}[i]$ is a Euclidean domain. We have a norm function $N : \mathbb{Q}[i] \to \mathbb{Q}$ where $N(r+is) = (r+is)\overline{(r+is)} = r^2+s^2$. We can check that this is multiplicative: $N(\alpha\beta) = \alpha\beta\overline{\alpha\beta} = \alpha\overline{\alpha}\beta\overline{\beta} = N(\alpha)N(\beta)$. (Here, we are using the fact $\overline{\alpha\beta} = \overline{\alpha} \cdot \overline{\beta}$, which is easy to check.)

**Proposition 12.1**

$N : \mathbb{Z}[i] \to \mathbb{Z}$ is a Euclidean function.

*Proof.* Let $m, n \in \mathbb{Z}[i]$ with $m, n \neq 0$. Write $n/m = a + bi$ where $a, b \in \mathbb{Q}$. Choose $x, y \in \mathbb{Z}$ such that $|a-x|, |b-y| \leq 1/2$. Letting $q = x+iy$ and $r = (a-x)+(b-y)i$, we have $n/m = q+r$ so $n = m \cdot q + m \cdot r$. By construction, both $m \cdot q$ and $m \cdot r$ are in $\mathbb{Z}[i]$. We can bound $N(m \cdot r) = N(m)N(r) \leq N(m) \cdot ((1/2)^2 + (1/2)^2) = N(m)/2 < N(m)$, so the norm is a Euclidean function. □

As a consequence, we obtain $\mathbb{Z}[i]$ is a UFD, since any Euclidean domain is a UFD. (This was the whole point of the first two or so lectures of the course.)

Now we review a result that you all proved on the problem set, which is an incredible theorem attributed to Fermat.

**Theorem 12.2** (Fermat)
If $p \equiv 1 \pmod 4$ is prime, then $p = a^2 + b^2$ for some $a, b \in \mathbb{Z}$.

*Proof.* We use the fact that $-1$ is a quadratic residue mod $p$ when $p \equiv 1 \pmod 4$. (In general, $\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2}$, which is just 1 when $p \equiv 1 \pmod 4$.) Therefore, this means there exists some integers $s, k \in \mathbb{Z}$ such that $s^2 + 1 = pk$.

Now we claim $p$ is reducible in $\mathbb{Z}[i]$. Suppose not. Then, $p \mid s^2 + 1$ implies either $p \mid s + i$ or $p \mid s - i$. Clearly this cannot be possible, so $p$ is indeed reducible, meaning we can write $p = \alpha \cdot \beta$ where $\alpha, \beta \in \mathbb{Z}[i]$ are not units. Write $\alpha = a + bi$. Then,

$$p^2 = N(p) = N(\alpha)N(\beta) = (a^2 + b^2) \cdot N(\beta).$$

Then, as $a^2+b^2 \mid p^2$, we have $a^2+b^2$ is either 1, $p$, or $p^2$. If it is 1, then $(a+bi)(a-bi) = 1$, which means $\alpha$ is a unit, contradicting our assumption that $\alpha$ is a non-unit. Likewise, if $a^2 + b^2 = p^2$, then $N(\beta) = \beta \cdot \overline{\beta} = 1$, which again means $\beta$ is a unit, giving us a contradiction. It follows $p = a^2 + b^2$, and we conclude. □

## 12.2 Irreducible Elements in Gaussian Integers

Okay, this is great – in fact, really great. But this is saying that primes 1 mod 4 are not irreducible. This begs the question,

What are the irreducible elements in $\mathbb{Z}[i]$?

We actually have a very good starting point to answer this question.

**Lemma 12.3**
If $N(a + bi)$ is a prime integer, then $a + bi$ is irreducible.

*Proof.* We prove the contrapositive. If $a+bi$ is reducible, i.e., $a+bi = \alpha\beta$ for non-units $\alpha, \beta$, then $N(a+bi) = N(\alpha)N(\beta)$ where $N(\alpha), N(\beta) \neq 1$. But then $N(a+bi)$ is not prime, as desired. □

Here, the fact that $N(\alpha) \neq 1$ follows from the fact that $\alpha$ is not a unit. We can actually use this kind of idea to find all units of $\mathbb{Z}[i]$.

**Lemma 12.4**

The units in $\mathbb{Z}[i]$ are $\{\pm 1, \pm i\}$.

*Proof.* If $\alpha = a+bi$ is a unit, then there exists some $\beta \in \mathbb{Z}[i]$ such that $\alpha \cdot \beta = 1$. Taking the norm, we have $N(\alpha)N(\beta) = 1$, which forces $N(\alpha) = 1$ as the norm is always nonnegative. (It is the sum of two squares.) But $a^2 + b^2 = 1$ is only possible when one of them is 0 and the other is $\pm 1$, which corresponds to the units we listed. □

Going back to our original question, we determined that any Gaussian integer with prime norm must be irreducible. But suppose I have an element with norm that is not prime. Then, can it be irreducible? (Basically, is the converse of Lemma 12.3 true?)

To hint at the answer of the above question, let me ask a different question. If $p \in \mathbb{Z}$ is prime, can $p$ be irreducible in $\mathbb{Z}[i]$? And if so, when? Fermat showed that $p$ is not irreducible when $p \equiv 1 \pmod 4$, and we just proved it, so we only need to consider primes 3 mod 4. It turns out that $p \equiv 3 \pmod 4$ is *always* irreducible.

**Proposition 12.5**

Any prime $p \equiv 3 \pmod 4$ is irreducible in $\mathbb{Z}[i]$.

*Proof.* Let $p = \alpha \cdot \beta$, with $\alpha = a+bi$ $(a, b \in \mathbb{Z})$. Since $p \in \mathbb{Z} \subset \mathbb{R}$, $\alpha$ and $\beta$ must be complex conjugates (if you don't believe me, just expand out the product and see what you need for the imaginary part to disappear). Thus, $p = \alpha \cdot \overline{\alpha} = a^2 + b^2$. But squares can only be $0, 1 \bmod 4$, so $a^2 + b^2 \not\equiv 3 \pmod 4$. In particular, it cannot be equal to $p$, so $p$ must be irreducible. □

Now, we have primes 3 mod 4 and Gaussian integers with prime norm as our irreducibles. How many more are there? It turns out that these cover all irreducibles!

**Lemma 12.6**

If $a+bi \in \mathbb{Z}[i]$ is irreducible, where $a, b \neq 0$, then $N(a+bi) = p$ is prime.

Note that if one of $a, b = 0$, then we are just dealing with integers, from which we concluded that the only integers which stay irreducible in $\mathbb{Z}[i]$ are primes 3 mod 4. I

guess we have to be a little careful and make sure to include $p = 2$ in our discussion, but we can factor $2 = (1+i)(1-i) = i(1-i)^2$, so it is reducible in $\mathbb{Z}[i]$.

*Proof.* Let $\pi = a+bi$ be irreducible. If $\pi \cdot \overline{\pi} = N(\pi) = \alpha \cdot \beta$ for some non-units $\alpha, \beta \in \mathbb{Z}$ (i.e., $\alpha, \beta \neq \pm 1$), then either $\pi \mid \alpha$ and $\overline{\pi} \mid \beta$ or the other way around. Without loss of generality, assume the former. By unique factorization of $N(\pi)$ in $\mathbb{Z}[i]$, we have $\alpha = \pi \cdot u$ and $\beta = \pi \cdot u'$ for units $u, u' \in \mathbb{Z}[i]$. But then $\pi \in \{\pm\alpha, \pm i\alpha\}$, but we assumed $a, b \neq 0$ in $\pi = a + bi$, so we have reached a contradiction. Hence, $N(\pi)$ is prime. □

## 12.3 Pythagorean Triples

Let's put our hard work to some good use. Let's solve the *Diophantine equation* (ah, there's the magic word) $x^2+y^2 = z^2$ for integers $x, y, z$. One can do this using elementary methods, and in fact it is not too bad, but the $x^2 + y^2$ expression is just *begging* us to consider this equation in the Gaussian integers. So let's do that.

Let us assume $\gcd(x, y, z) = 1$ (otherwise we can just divide through by the common factor). We consider the equation in $\mathbb{Z}[i]$ and use the fact that $\mathbb{Z}[i]$ is a unique factorization domain. We have $(x + iy)(x - iy) = z^2$. The key trick now, which will help us take advantage of unique factorization, is the fact that $x + iy$ and $x - iy$ are coprime.

**Claim 12.7.** If $x, y \in \mathbb{Z}$ are coprime as integers, then $x+iy$ and $x-iy$ are coprime in $\mathbb{Z}[i]$.

*Proof.* Let $\pi$ be an irreducible dividing $x + iy$. Suppose for the sake of contradiction that $\pi \mid x - iy$. Then, $\pi \mid (x + iy) + (x - iy) = 2x = (1 + i)(1 - i)x$, so $\pi$ must divide one of those factors. If $\pi \mid 1 + i$, then $N(\pi) \mid N(1 + i) = 2$; as $\pi$ is a non-unit, we have $N(\pi) = 2$. But we have $\pi \mid x + iy$, and this means $\overline{\pi} \mid \overline{x + iy} = x - iy$; combining gives $2 = N(\pi) \mid N(x + iy) = x^2 + y^2 = z^2$. Thus, $2 \mid z^2$, so $z$ is even.

But then $x^2 + y^2 = z^2 \equiv 0 \pmod 4$, which is only possible if $x^2 \equiv y^2 \equiv 0 \pmod 4$ (if they were both odd, then $x^2 + y^2 \equiv 1 + 1 = 2 \pmod 4$). In particular, that means $x, y$ are both even. This contradicts the assumption that $x, y, z$ are coprime, so $\pi \nmid 1+i$. Likewise, we can make the same argument for $1 - i$ to show $\pi \nmid 1 - i$.

Thus, $\pi \mid (1+i)(1-i)x$ but $\pi \nmid 1+i, 1-i$, so $\pi \mid x$. Likewise, $\pi \mid (x+iy)-(x-iy) = 2iy = i(1+i)(1-i)y$, so $\pi \mid y$ as well. But $x, y$ are coprime as integers, so $\pi$ cannot be a prime integer itself. Lemma 12.6 tells us then that $N(\pi)$ is some prime $p$. Taking the norm of our divisibility conditions, we have $p = N(\pi) \mid N(x) = x^2$ and $p \mid y^2$ similarly. This contradicts again our assumption that $x, y$ are coprime, so indeed $\pi \nmid x - iy$ and thus $x + iy, x - iy$ are coprime. □

Now we will use this claim to classify all Pythagorean triples. Recall we have $x^2 + y^2 = (x + iy)(x - iy) = z^2$, and the two terms in the middle are coprime. We can

factor $z$ into irreducibles in $\mathbb{Z}[i]$: let $z = u \cdot \pi^{a_1} \cdots \pi_r^{a_r}$ where $u$ is a unit and $\pi_j$'s are irreducibles. Then,

$$(x+iy)(x-iy) = (u \cdot \pi_1^{a_1} \cdots \pi_r^{a_r})^2.$$

But since $x+iy$ and $x-iy$ are coprime and their product is a square, they must each be squares! (Up to units, of course.) Thus, $x+iy = w \cdot \beta^2$ for some unit $w$ and $\beta \in \mathbb{Z}[i]$. Write $\beta = a+bi$, and suppose $w = 1$. (It turns out that if you choose some other unit for $w$, then we would get the same answer we are about to obtain, so we will just do the $w = 1$ case here.) Then,

$$x + iy = (a+bi)^2 = (a^2 - b^2) + (2ab)i,$$

so any primitive Pythagorean triple is of the form $(x, y, z) = (a^2 - b^2, 2ab, a^2 + b^2)$ for integers $a, b \in \mathbb{Z}$.

# 13 10/27 - More Diophantine Equations

Today, we'll look at some special (note: easier) cases of Fermat's Last Theorem. The story of Fermat's Last Theorem, first claimed by Fermat in 1637,[11] is truly an incredible one, which was resolved by Andrew Wiles's famous proof in 1995. His proof sparked some of the most important mathematics in modern number theory (most importantly, elliptic curves and modular forms – these two are related by something called the Modularity Conjecture, which Wiles proved), and much of it is being developed to this day. Okay, let me state the theorem, which is a incredibly deceptively simple one.

**Theorem 13.1** (Fermat's Last Theorem)
For $n > 2$, there does not exist integers $x, y, z$ such that $x^n + y^n = z^n$ and $xyz \neq 0$.

Given that the proof of this took more than 350 years to arise, it is clearly out of the scope of this class, but we can investigate this problem for small values of $n$.

## 13.1 Method of Infinite Descent

We will use an argument called the **method of infinite descent**, whose mantra is basically: given a solution, find a smaller solution. This is used for proofs by contradiction, as if we can perform descent on integer solutions, then we will continuously get smaller and smaller integer solutions. But the integers can only get so small in magnitude, so the descent must break at some point. This is best illustrated by an example:

[11] He claimed "I have discovered a truly marvelous proof of this, which this margin is too narrow to contain." In French, of course, which makes it even more pretentious. He was definitely lying here.

**Exercise 13.2.** Let $p$ be a prime. Show that $x^3 + py^3 + p^2z^3 = 0$ has no solutions with $xyz \neq 0$.

*Proof.* Note $p \mid x^3$, so $p \mid x$. Write $x = px'$. Substituting gives $p^3x'^3 + py^3 + p^2z^3 = 0$; clear $p$ to get $p^2x'^3 + y^3 + pz^3 = 0$. Now, we have $p \mid y$. Write $y = py'$ and repeat the same process: we'll get $px'^3 + p^2y'^3 + z^3 = 0$, so $p \mid z$. Repeat to get $x'^3 + py'^3 + p^2z'^3 = 0$. But this means from a solution $(x, y, z)$, we can always generate a smaller solution $(x', y', z')$, and we can do this ad infinitum. But clearly this is not possible for the integers, so there are no solutions. □

**Remark 13.3.** Note we could technically prove this by induction, or a similar argument, by being like "Suppose there are no solutions less than $n$; we will show there are still no solutions less than $n+1$ by performing descent once." For instance, we could make the above proof cleaner by starting with "Assume $(x, y, z)$ is the solution to the equation that minimizes $|x|+|y|+|z|$"; then, we could have stopped once we got $x = px'$ as the first step gave us $y^3 + pz^3 + p^2x'^2 = 0$. So this idea is not really super new. The real upshot, in my opinion, is that saying you proved something by method of descent gives people the impression that you're some cool mad mathematician, which is fun.

## 13.2 Fermat's Last Theorem for $n = 4$

Now we are going to make some reductions to Fermat's Last Theorem. We will first prove a neat theorem by Fermat, which he actually did prove:

**Theorem 13.4** (Fermat's Last Theorem, $n = 4$)
The equation $x^4 + y^4 = z^2$ has no integer solutions $x, y, z$ with $xyz \neq 0$.

Suppose this is true. Then, Fermat's Last Theorem is reduced to the case when $n = p$ is an odd prime. To see why, let $n = mp$ for some odd prime $p$. If $x^p + y^p = z^p$ has no solutions, then so will $(x^m)^p + (y^m)^p = (z^m)^p$, which is just $x^n + y^n = z^n$. The only case not covered here is when $n$ is not divisible by an odd prime, i.e., when $n$ is a power of 2. But this is covered by the $n = 4$ case above, which we will now prove.

*Proof.* We may assume $(x, y, z)$ are pairwise coprime, as if $p$ divided two of them, then the equation forces $p$ to divide the third, and then we could consider the smaller solution $(x/p, y/p, z/p^2)$.

Suppose $(x, y, z)$ is the smallest solution to this equation; by smallest, we mean $|z|$ is minimized. If $(x, y, z)$ satisfies the equation, then $(x^2, y^2, z)$ is a primitive Pythagorean

triple. We found a general form for Pythagorean triples! This means that there exist coprime $k, \ell \in \mathbb{N}$ such that

$$x^2 = k^2 - \ell^2, \qquad y^2 = 2k\ell, \qquad z = k^2 + \ell^2.$$

The first equation gives $x^2 + \ell^2 = k^2$, which is – you guessed it – yet another Pythagorean triple! (So our work on finding Pythagorean triples at the end of last class wasn't all that capricious after all.) This means that there exist coprime $a, b \in \mathbb{N}$ such that

$$x = a^2 - b^2, \qquad \ell = 2ab, \qquad k = a^2 + b^2.$$

Note that there is also the case where $x = 2ab$ and $\ell = a^2 - b^2$, but this cannot hold since from our choice $x^2 = k^2 - \ell^2$ and $y^2 = 2k\ell$ we are declaring $x$ to be odd.

Therefore, $y^2 = 2k\ell = 2ab(a^2 + b^2)$, which means $y$ is even and $(y/2)^2 = ab(a^2 + b^2)$. Observe that for $(a, b) = 1$, we have $(ab, a^2 + b^2) = 1$. So in fact, we have three pairwise coprime elements: $a$, $b$, and $a^2 + b^2$. But their product is a square! So in the spirit of the work that we did for finding irreducible elements in $\mathbb{Z}[i]$ (from last class), this means $a$, $b$, and $a^2 + b^2$ are each perfect squares themselves. Write $a = x'^2$, $b = y'^2$, and $a^2 + b^2 = z'^2$. This means $x'^4 + y'^4 = z'^2$. But

$$z' \leq (z')^2 = a^2 + b^2 = k \leq k^2 < z,$$

so we found a smaller solution $(x', y', z')$, and the proof concludes by an infinite descent argument. □

From our remarks above, now we are just left with the cases of Fermat's Last Theorem when $n = p$ is an odd prime. Again, we won't do this in full generality, but we have the tools to prove it for $n = 3$. We will do this next time.

## 13.3 Sophie Germain's Theorem

Here is another neat Diophantine equation. This is known as the "first case" of Fermat's Last Theorem.

**Theorem 13.5** (Sophie Germain)
Let $p$ be an odd prime such that $2p+1$ is also prime. Then, the equation $x^p + y^p = z^p$ has no integer solutions with $p \nmid xyz$.

**Remark 13.6.** We can list some primes $p$ such that $2p + 1$ is also prime: 3,5,11, 23, ... It is actually an open problem whether or not there are infinitely many such primes, yet another exhibit for why primes are so elusive.

*Proof.* Note that if $x^p + y^p = z^p$, then $x^p + y^p + (-z)^p = 0$, so we will consider the equation $x^p + y^p + z^p = 0$ instead. Like in our previous arguments, we may assume $x, y, z$ are pairwise coprime. We can factor $x^p + y^p = (x+y)(x^{p-1} - x^{p-2}y + \cdots + y^{p-1})$. We will prove that the two factors on the right are relatively prime. First, observe that $p \nmid xyz \implies p \nmid x^p$, which means $p$ cannot divide either factor. If $r \neq p$ is a prime dividing both factors, then $r \mid x + y \implies x \equiv -y \pmod{r}$, in which case

$$\begin{aligned} 0 &\equiv x^{p-1} - x^{p-2}y + \cdots + y^{p-1} \\ &\equiv (-y)^{p-1} - (-y)^{p-2}y + \cdots + y^{p-1} \\ &\equiv py^{p-1} \pmod{r}, \end{aligned}$$

but $r \neq p$, so $y^{p-1} \equiv 0 \pmod{r}$ which is only possible if $r \mid y$. But then this means $r \mid z$ as well, contradicting $y, z$ being relatively prime.

Thus, since the two factors are coprime, and their product is $(-z)^p$ a perfect $p^{\text{th}}$ power, each of the factors must be a perfect $p^{\text{th}}$ power. Write $x + y = A^p$ and $x^{p-1} - x^{p-2}y + \cdots + y^{p-1} = T^p$ for $A, T \in \mathbb{Z}$. But we can do this same process for the equivalent equations $x^p + z^p = (-y)^p$ and $y^p + z^p = (-x)^p$ to get $x + z = B^p$ and $y + z = C^p$ for some $B, C \in \mathbb{Z}$. Letting $q \coloneqq 2p + 1$, which by assumption is prime, we have $p = (q-1)/2$, so $x^p + y^p + z^p = 0$ implies

$$x^{\frac{q-1}{2}} + y^{\frac{q-1}{2}} + z^{\frac{q-1}{2}} \equiv 0 \pmod{q}.$$

If $q \nmid xyz$, then each term in the sum is either $\pm 1$ (recall $x^{\frac{q-1}{2}} = \left(\frac{x}{p}\right)$ which is either $\pm 1$ for $p \nmid x$), so the left hand side cannot possibly be 0 mod $q$. WLOG suppose $q \mid z$ but $q \nmid x, y$. Using $x + y = A^p$, $x + z = B^p$, $y + z = C^p$, we have $B^p + C^p - A^p = 2z$, which means

$$B^{\frac{q-1}{2}} + C^{\frac{q-1}{2}} - A^{\frac{q-1}{2}} \equiv 0 \pmod{q}.$$

By the same argument as in the above paragraph (each term is $\pm 1 \bmod q$), this forces $q \mid ABC$. But looking at the definitions of $A, B, C$ tells us $q \nmid B$ and $q \nmid C$, so we must have $q \mid A$. So now we return to $A$: we have $x + y = A^p \equiv 0 \pmod{q}$, meaning $y \equiv -z \pmod{q}$. We haven't talked about $T$ yet, so let's bring that in now: we see $T^p = x^{p-1} - x^{p-2}y + \cdots + y^{p-1} \equiv py^{p-1} \pmod{q}$. Noting that since $q \mid x$, we have $A^p = x + y \equiv y \pmod{q}$, so $T^p \equiv p(A^p)^{p-1} \pmod{q}$. Rewriting, we obtain

$$\begin{aligned} \frac{q-1}{2} = p &\equiv \left(T \cdot (B^{p-1})^{-1}\right)^p \pmod{q} \\ &= \left(T \cdot (B^{p-1})^{-1}\right)^{\frac{q-1}{2}} \pmod{q} \\ &\equiv \pm 1 \pmod{q}, \end{aligned}$$

so $-1/2 \equiv \pm 1 \pmod{q}$. But $p \geq 3$ implies $q \geq 7$, and this equality cannot hold when $q \geq 7$, so we have reached a contradiction! Sophie Germain's Theorem follows. □

# 14 10/30 - Fermat's Last Theorem for $n = 3$

The title says it all. Recall we did this for $n = 4$ last time, and given the special condition where $2p + 1$ is also a prime, we did this for odd primes $p$ (Theorem 13.5). Now we will do the $n = 3$ case in full generality, without the assumption given in Sophie Germain's result.

To work in the $n = 3$ case, we take some inspiration from the equation $a^2 + b^2 = c^2$. (This is the work we did for finding Pythagorean triples, see §12.3.) For this equation, we factored $a^2 + b^2 = (a + bi)(a - bi)$, so the key was working this equation over $\mathbb{Z}[i]$. In the case $a^3 + b^3 = c^3$, we can factor $a^3 + b^3 = (a + b)(a + b\omega)(a + b\omega^2)$, where $\omega = e^{2\pi i/3}$ (so $\omega^3 = 1$, meaning $\omega$ is a primitive third root of unity). Let us state Fermat's Last Theorem for $n = 3$ from this lens:

**Theorem 14.1** (Fermat's Last Theorem for $n = 3$)
Let $u \in \mathbb{Z}[\omega]$ be a unit. Then, the equation $x^3 + y^3 = uz^3$ has no solutions for $x, y, z \in \mathbb{Z}[\omega]$ with $xyz \neq 0$.

Note this is actually stronger than what we need for Fermat's Last Theorem, but hey, we will take stronger statements any day.

## 14.1 Eisenstein Integers

Note that our problem is actually a problem on $\mathbb{Z}[\omega]$ in disguise. The ring $\mathbb{Z}[\omega]$ is often called the **Eisenstein integers**. We will prove some facts about this ring, beginning with something you showed in your first (second?) homework:

**Lemma 14.2**
$\mathbb{Z}[\omega]$ is a Euclidean domain.

Before we start working with $\omega$, let's lay out a few identities we will consistently use. First, note $\overline{\omega} = \omega^2 = \omega^{-1}$ and $\omega^3 - 1 = 0 \implies 1 + \omega + \omega^2 = 0$. This means $\omega + \overline{\omega} = \omega + \omega^{-1} = -1$.

*Proof.* Let $\lambda : \mathbb{Q}[\omega] \to \mathbb{Q}$ send $\alpha \mapsto \alpha \cdot \overline{\alpha}$. Indeed, $A(\alpha) \in \mathbb{Q}$: if $\alpha = a + b\omega$, then

$$\begin{aligned} \alpha\overline{\alpha} &= (a + b\omega)(a + b\overline{\omega}) \\ &= a^2 + b^2 + ab(\omega + \overline{\omega}) \\ &= a^2 - ab + b^2 \in \mathbb{Q}. \end{aligned}$$

If $m, n \in \mathbb{Z}[\omega]$ with $m \neq 0$, then we can write $n/m = q + r$, where $q \in \mathbb{Z}[\omega]$ and

$r = x + y\omega \in \mathbb{Q}[\omega]$ such that $|x|, |y| \le 1/2$. Thus, $n = mq + mr$, with

$$\lambda(mr) = \lambda(r) \cdot \lambda(m) \le \frac{3}{4}\lambda(m) < \lambda(m),$$

so we found a valid Euclidean function $\lambda$. □

Our statement (Theorem 14.1) mentions the units of $\mathbb{Z}[\omega]$; let us state what they are.

**Lemma 14.3**
$\mathbb{Z}[\omega]^\times = \{\pm 1, \pm\omega, \pm\omega^2\}$.

*Proof.* If $\alpha \in \mathbb{Z}[\omega]$ is a unit, then $1 = N(\alpha\alpha^{-1}) = N(\alpha)N(\alpha^{-1})$, so $N(\alpha) = 1$. Likewise, if $N(\alpha) = 1$, then by definition of the norm function, $\alpha$ has an inverse, i.e., it is a unit. Letting $\alpha = a + b\omega$, we have $N(\alpha) = a^2 - ab + b^2 = 1$, which we can rewrite as $(2a - b)^2 + 3b^2 = 4a^2 - 4ab + 4b^2 = 4$. We have two cases here.

If $2a - b = \pm 1$, then $b^2 = 1$, so we get $a = 0$ and $b = \pm 1$. For the other case, if $2a - b = \pm 2$, then $b^2 = 0$, so $a = \pm 1$. These give our four claimed units, as desired. □

## 14.2 Properties of $\lambda = 1 - \omega$

We continue with proving more properties of elements in $\mathbb{Z}[\omega]$ to build up machinery.

**Lemma 14.4**
Let $\lambda = 1 - \omega$. Then,

1. $N(\lambda) = 3$,
2. $\lambda$ is irreducible in $\mathbb{Z}[\omega]$,
3. $(\lambda^2) = (3)$,
4. $\mathbb{Z}[\omega]/\lambda\mathbb{Z}[\omega] \simeq \mathbb{Z}/3\mathbb{Z}$.

*Proof.* Let's see how fast we can work through all of these.

The first one is literally just using the equation $N(a + b\omega) = a^2 - ab + b^2$. Let's move on to (2). If $\lambda = \alpha\beta$, then $3 = N(\lambda) = N(\alpha)N(\beta)$, which is only possible if one of $N(\alpha)$ or $N(\beta)$ is 1, i.e., one of $\alpha, \beta$ is a unit.

For (3), note $\lambda^2 = (1 - \omega)^2 = 1 - 2\omega + \omega^2 = -3\omega$. Taking the ideals generated on each side and noting $-\omega$ is a unit, we get the result.

The hardest part is (4), but it is really not that bad. In fact, you did most of this work in a previous problem set! (Problem Set 2, maybe?) The homework problem

demonstrated that any $\alpha \in \mathbb{Z}[\omega]$ is congruent to either 0 or $\pm 1$ mod $\lambda$. Thus, we can construct a map $\mathbb{Z}[\omega]/\lambda\mathbb{Z}[\omega] \to \mathbb{Z}/3\mathbb{Z}$ where some $\alpha$ on the left maps to its residue mod $\lambda$, which we just said is either $0, \pm 1$. This map is surjective ($\{-1, 0, 1\}$ map to $\{-1, 0, 1\}$ identically), but from (3), we see that $|\mathbb{Z}[\omega]/\lambda\mathbb{Z}[\omega]|$ divides 3. But it is certainly not the trivial group, so it has size 3, which means the map must be an isomorphism. This gives our desired result. □

Another interesting fact related to $\lambda$:

**Fact 14.5.** If $x \in \mathbb{Z}[\omega]$ and $x \equiv 1 \pmod{\lambda}$, then $x^3 \equiv 1 \pmod{\lambda^4}$. Additionally, if $\lambda \nmid x$, then $x^3 \equiv \pm 1 \pmod{\lambda^4}$.

*Proof.* If $x = 1 + \lambda t$, then

$$\begin{aligned} x^3 - 1 &= (x-1)(x-\omega)(x-\omega^2) \\ &= \lambda t(1 - \omega + \lambda t)(1 - \omega^2 + \lambda t) \\ &= \lambda t(\lambda + \lambda t)(\lambda(1+\omega) + \lambda t) \\ &= \lambda^3 t(1+t)(1+\omega+t). \end{aligned}$$

Note that any $t$ must be $\equiv 0, \pm 1 \pmod{\lambda}$). If $\lambda \equiv -1$, the $\lambda \mid 1+t$, so $\lambda^4 \mid x^3 - 1$. If $t \equiv 0$, then $\lambda^4 \mid \lambda^3 t \mid x^3 - 1$. Finally, if $t \equiv 1$, then $1 + \omega + t \equiv 2 + \omega \equiv 3 - \lambda = \lambda(\overline{\lambda} - 1)$, so again we get an extra factor of $\lambda$ and the first conclusion follows.

Note $x \equiv 0, \pm 1 \pmod{\lambda}$. If $x \equiv -1$ (i.e., $-x \equiv 1$), then $(-x)^3 \equiv 1 \pmod{\lambda^4} \implies x^3 \equiv -1$. This finishes the result. □

## 14.3 Proving Theorem 14.1

Now we make our first big step towards proving Theorem 14.1. This also shows why we are caring so much about this $\lambda$ element.

**Lemma 14.6** (Weaker Version of Theorem 14.1)
Theorem 14.1 holds if $\lambda \nmid xyz$.

*Proof.* We know from Lemma 14.4.2 that $\lambda$ is irreducible, so if $\lambda \nmid xyz$, then $\lambda$ does not divide each factor. Look at the equation $x^3 + y^3 = uz^3$ in modulo $\lambda^4$. From Fact 14.5 above, the left hand side is of the form $\pm 1 \pm 1$, which takes on values $\{0, \pm 2\}$. Meanwhile, the right hand side is congruent to $\pm u$, but we found all the units of $\mathbb{Z}[\omega]$ in Lemma 14.3! So the right hand side takes on values $\{\pm 1, \pm\omega, \pm\omega^2\}$. There is no overlap mod $\lambda^4$ (note $(\lambda^4) = (9)$, and it is clear none of these values are congruent mod 9), so there are no solutions. □

We strive to do even better.

**Lemma 14.7**
If $x^3 + y^3 = uz^3$, with $\lambda \nmid xy$ and $\lambda \mid z$, then $\lambda^2 \mid z$, i.e., $\operatorname{ord}_\lambda z \geq 2$.

*Proof.* Again, we use the incredibly useful Fact 14.5. We reduce our equation to mod $\lambda^4$. Again, like above, the left hand side takes on values $\{0, \pm 2\}$. Let $L$ be the value of the left hand side. Then, we have $L \equiv uz^3 \pmod{\lambda^4}$, but $\lambda \mid z$, so we must have $L \equiv 0 \pmod{\lambda}$. This is only true when $L = 0$, so $uz^3 \equiv 0 \pmod{\lambda^4}$. Thus, $\operatorname{ord}_\lambda z^3 = 3 \operatorname{ord}_\lambda z \geq 4$, which means $\operatorname{ord}_\lambda z \geq 2$, as desired. □

We will again use the method of infinite descent. The following result activates the descent step by finding a smaller solution given an initial one.

**Lemma 14.8**
If $x^3 + y^3 = uz^3$ with $\lambda \nmid xy$ and $\operatorname{ord}_\lambda z \geq 2$, then there exist $x_1, y_1, z_1 \in \mathbb{Z}[\omega]$, where $x_1 y_1 z_1 \neq 0$, and a unit $u_1 \in \mathbb{Z}[\omega]^\times$ such that $x_1^3 + y_1^3 = u_1 z_1^3$ with $\lambda \nmid x_1 y_1$ and $\operatorname{ord}_\lambda z_1 = \operatorname{ord}_\lambda z - 1$.

As a consequence, the infinite descent method tells us that there are no solutions when $\lambda \nmid xy$.

*Proof.* If $\alpha, \beta \in \mathbb{Z}[\omega]$ are nonzero, then in general we have the inequality $\operatorname{ord}_\lambda(\alpha + \beta) \geq \min(\operatorname{ord}_\lambda \alpha, \operatorname{ord}_\lambda \beta)$. Equality holds when $\operatorname{ord}_\lambda \alpha \neq \operatorname{ord}_\lambda \beta$. WLOG let $s = \operatorname{ord}_\lambda \alpha < \operatorname{ord}_\lambda \beta$. Then, we can write $\alpha + \beta = \lambda^s(\alpha/\lambda^s + \beta/\lambda^s)$.

We may assume $x, y, z$ have no common factors in $\mathbb{Z}[\omega]$. We can factor our equation as

$$(x + y)(x + \omega y)(x + \omega^2 y) = uz^3.$$

From our assumption $\operatorname{ord}_\lambda z \geq 2$, we have $\operatorname{ord}_\lambda uz^3 \geq 6$, so one of the factors on the left hand side must have order $\geq 2$. But the terms on the left are all "symmetric" (for instance, we could replace $y$ with $\omega y$ and get the same exact expression), so we can assume without loss of generality that $\operatorname{ord}_\lambda(x + y) \geq 2$. Then,

$$(x + y) - (x + \omega y) = (1 - \omega)y = \lambda y,$$

and as $\lambda \nmid y$, this means $\operatorname{ord}_\lambda((x + y) - (x + \omega y)) = 1$. By the work we did in the beginning of the proof, it follows that $\operatorname{ord}_\lambda(x + \omega y) = 1$. Likewise, $\operatorname{ord}_\lambda(x + \omega^2 y) = 1$, so $\operatorname{ord}_\lambda(x + y) = 3 \operatorname{ord}_\lambda z - 2$.

If $\pi \nmid \lambda$ is an irreducible and $\pi \mid x + y, x + \omega y$, then $\pi \mid (x + y) - (x + \omega y) = (1 - \omega)y = \lambda y$, which is only possible if $\pi \mid y$. But then $\pi \mid x$, which contradicts our assumption that $x, y, z$ share no common factors. Likewise, we have that all three of the terms on

the left are pairwise coprime. But their product is a perfect cube up to some unit, so we can write

$$x + y = u_1\alpha^3\lambda^t \tag{3}$$
$$x + \omega y = u_2\beta^3\lambda \tag{4}$$
$$x + \omega^2 y = u_3\gamma^3\lambda, \tag{5}$$

where $t = 3\operatorname{ord}_\lambda z - 2$, $u_i$ are units, and $(\alpha, \beta, \gamma) = 1$. Note that $(x+y)+\omega(x+\omega y)+\omega^2(x+\omega^2 y) = (x+y)(1+\omega+\omega^2) = 0$, so we have

$$\begin{aligned} 0 = (3) + (4)\omega + (5)\omega^2 &= u_1\alpha^3\lambda^t + u_2\beta^3\lambda\omega + u_3\gamma^3\lambda\omega^2 \\ \implies 0 &= u_1\alpha^3\lambda^{t-1} + u_2\beta^3\omega + u_3\gamma^3\omega^2 \\ \implies (-u_2\omega)^{-1}u_1\alpha^3\lambda^{t-1} &= \beta^3 + (u_2\omega)^{-1}u_3\omega^2\gamma^3. \end{aligned}$$

Alright, we are in the thick of it, but we will reach the end of this tunnel soon. We now want to construct our smaller solution, which completes the descent step. Let $z_1 = \alpha\lambda^{(t-1)/3}$ (in particular, $\frac{t-1}{3} = \operatorname{ord}_\lambda z - 1$), $y_1 = \gamma$, and $x_1 = \beta$. Also, let $\varepsilon_2 = (-u_2\omega)^{-1}u_1$ and $\varepsilon_1 = (u_2\omega)^{-1}u_3\omega^2$; note that both $\varepsilon_1, \varepsilon_2$ are units. This leaves us with the equation $\varepsilon_2 z_1^3 = x_1^3 + \varepsilon_1 y_1^3$.

Now we may reduce this equation mod $\lambda^2$. Recall $z_1 = \alpha\lambda^{(t-1)/3}$, so $\operatorname{ord}_\lambda z_1^3 \geq 3\cdot 1 > 2$, so the left hand side is congruent to 0 mod $\lambda^2$. Following Fact 14.5, the right hand side reduces to $\pm 1 \pm \varepsilon_1 \pmod{(\lambda^2)}$, and combining with $(\lambda^2) = (3)$ (Lemma 14.4.3), we get $\varepsilon_1 \equiv \pm 1 \pmod 3$. Then, $\varepsilon_2 z_1^3 = x_1^3 \pm y_1^3$; replacing $y_1$ with $-y_1$ if the sign is negative, we get $\varepsilon_2 z_1^3 = x_1^3 + y_1^3$, as desired. □

As mentioned before the proof, the infinite descent method tells us that there are no solutions when $\lambda \nmid xy$, as we can continue decreasing $\operatorname{ord}_\lambda z$ but the order must stay non-negative.

**Corollary 14.9**

The equation $x^3 + y^3 = uz^3$, where $u \in \mathbb{Z}[\omega]^\times$ and $x, y, z \in \mathbb{Z}[\omega]$ such that $xyz \neq 0$ and $\lambda \nmid xy$, has no solutions.

We can finally prove Theorem 14.1, in a slightly more general setting.

**Proposition 14.10** (implies Fermat's Last Theorem for $n = 3$)

The equation $x^3 + y^3 = uz^3$ has no solutions for $x, y, z \in \mathbb{Z}[\omega]$, where $xyz \neq 0$ and $u \in \mathbb{Z}[\omega]^\times$.

*Proof.* We may assume $(x, y, z) = 1$, else we can divide through by their common factor. We proved this above for when $\lambda \nmid xy$. Suppose $\lambda \mid x$. Then, we must have $\lambda \nmid yz$ for

$(x, y, z) = 1$ to hold. Since $\lambda \nmid y, z$, by Fact 14.5, we have $y^3, z^3 \equiv \pm 1 \pmod{\lambda^4}$, so $u \equiv (yz^{-1})^3 \equiv \pm 1 \pmod{\lambda^4}$. Reducing to mod $\lambda^2$, we have $u \equiv \pm 1 \bmod (\lambda^2) = \pm 1 \bmod (3)$, so $3 \mid u \pm 1$ in $\mathbb{Z}[\omega]$. But we know all the units of $\mathbb{Z}[\omega]^\times$ from Lemma 14.3! Going through all six possibilities, we conclude $u = \pm 1$.

Thus, our equation looks like $x^3 + y^3 = \pm z^3$. We can rewrite this as $x^3 = -y^3 \pm z^3 = (-y)^3 + (\pm z)^3$. But we have $\lambda \nmid yz$, so it now follows from our Corollary above that there are no solutions, as desired. □

# 15 11/03 - Pell's Equations

We continue to study Diophantine equations which look simple but are deceptively difficult to solve. Welcome to Pell's Equations.[12]

Let $d \in \mathbb{N}$ be a square-free integer. A **Pell equation** is of the form $x^2 - dy^2 = 1$, and we are interested in finding solutions $x, y \in \mathbb{N}$. Note we just need to care about square-free $d$, as if $D = d \cdot n^2$, then we have $x^2 - Dy^2 = x^2 - d(ny)^2$, so it suffices to find solutions for the latter Pell equation. The study of finding solutions to these equations is very beautifully related to the theory of continued fractions, as for large enough $x, y$, we have $x^2 \sim dy^2 \implies x/y \sim \sqrt{d}$, and the continued fraction expansion of $\sqrt{d}$ provides the "closest possible" rational approximations to $\sqrt{d}$.

Note that the expression $x^2 - dy^2$ is just begging for us to factor it as $(x + y\sqrt{d})(x - y\sqrt{d})$ in $\mathbb{Z}[\sqrt{d}]$. Indeed, this is where we are going.

**Theorem 15.1** (Solutions to Pell's Equations)

The equation $x^2 - dy^2 = 1$ has infinitely many solutions $x, y \in \mathbb{N}$, all of the form $(x_n, y_n)$ where

$$x_n + y_n\sqrt{d} = (x_1 + y_1\sqrt{d})^n,$$

where $(x_1, y_1)$ is the smallest solution to the equation. (Smallest here can just mean $x_1$ is minimized.)

**Remark 15.2.** Although $\mathbb{Z}[\sqrt{d}]$ does not have any imaginary part ($d \in \mathbb{N}$ so $\sqrt{d} \in \mathbb{R}$), we still have a notion akin to conjugation. In particular, the map $\mathbb{Z}[\sqrt{d}] \to \mathbb{Z}[\sqrt{d}]$ sending $a + b\sqrt{d} \mapsto a - b\sqrt{d}$ is an automorphism (bijective homomorphism)! Denoting $\overline{a + b\sqrt{d}} = a - b\sqrt{d}$, one can check that $\overline{\alpha} + \overline{\beta} = \overline{\alpha + \beta}$ and $\overline{\alpha} \cdot \overline{\beta} = \overline{\alpha \cdot \beta}$.

This is useful because it gives us a nice new perspective of Pell's equation. If we take the norm map as the product of an element with its conjugate, as we do in the complex numbers, then we have $N(x + y\sqrt{d}) = (x + y\sqrt{d})(x - y\sqrt{d}) = x^2 - dy^2$. So we are essentially just finding the elements of $\mathbb{Z}[\sqrt{d}]$ with norm 1.

[12] Note: not due to Pell! Impostor.

This is nice because it verifies that all of our claimed solutions for Pell's equation as described in Theorem 15.1 are indeed solutions! If $N(\alpha) = 1$ (so writing $\alpha = x_1 + y_1\sqrt{d}$, we have $x_1^2 - dy_1^2 = 1$), then by multiplicativity of the norm, we have $N(\alpha^n) = 1$, so $(x_n, y_n)$ is also a solution to the equation. Also, if we find some $\alpha$ such that $N(\alpha) = -1$, then we can recover a solution to Pell's equation by just looking at $\alpha^2$, since $N(\alpha^2) = (-1)^2 = 1$.

**Example 15.3**

Consider $x^2 - 5y^2 = 1$. We have $2^2 - 5 = -1$, so we look at $(2 + \sqrt{5})^2 = 9 + 4\sqrt{5}$. Indeed, $9^2 - 5 \cdot 4^2 = 1$, so we found a solution!

## 15.1 Approximating with Fractions

Let us take a step away from this real quadratic $\mathbb{Z}[\sqrt{d}]$ for a second and indulge in the perspective of continued fractions.

**Lemma 15.4**

Let $\xi \in \mathbb{R}$ be irrational. Then, there are infinitely many $x/y \in \mathbb{Q}$ (where $x, y \in \mathbb{Z}$ and $(x, y) = 1$) such that

$$\left| \frac{x}{y} - \xi \right| < \frac{1}{y^2}.$$

This is quite a powerful statement! The error being bounded by $1/y^2$ is forcing the error to be impressively small.

Before we begin the proof, let us lay out some notation. For $\alpha \in \mathbb{R}$, denote $[\alpha]$ as the largest integer less than or equal to $\alpha$, and denote $\{\alpha\} = \alpha - [\alpha] \in [0, 1)$ as the fractional part of $\alpha$.

*Proof.* Choose some $n \in \mathbb{N}$. We divide up the interval $[0, 1) = [0, 1/n) \cup [1/n, 2/n) \cup \cdots \cup [(n-1)/n, 1)$ into $n$ equal subintervals. Consider the list $0, \{\xi\}, \{2\xi\}, \ldots, \{n\xi\}$. This has $n + 1$ terms, so by Pigeonhole Principle, two of them are in the same subintervals. In other words, $\exists 0 \leq j < k \leq n$ such that $|\{j\xi\} - \{k\xi\}| < 1/n$. Rewriting this via $\{\xi\} = \xi - [\xi]$, we have

$$|j\xi - k\xi + [k\xi] - [j\xi]| < 1/n.$$

Letting $x = [k\xi] - [j\xi]$ and $y = k - j$, we have $|x - y\xi| < 1/n$. Note that both $0 \leq j, k \leq n$, so we have $y = k - j \leq n$, meaning

$$\left| \frac{x}{y} - \xi \right| < \frac{1}{ny} \leq \frac{1}{y^2}.$$

Great, so this gives us one fraction $x/y$ satisfying the inequality. Let us generate infinitely many more! Note that we could run the same argument for any $n$, but we

must be careful in not producing the same fraction $x/y$ over and over again. We will be more careful in our choice of $n$.

Since $\xi$ is irrational, we know $|x/y-\xi| \neq 0$; choose $m \in \mathbb{N}$ such that $|x/y-\xi|^{-1} < m$. Now, run the same argument as above but with $m$ instead of $n$ to get some fraction $x_1/y_1$ such that

$$\left|\frac{x_1}{y_1} - \xi\right| < \frac{1}{my_1} \leq \frac{1}{m} < \left|\frac{x}{y} - \xi\right|,$$

and we know $\frac{x_1}{y_1}$ satisfies the inequality by construction. We can continue this process to construct infinitely many $x_n/y_n \in \mathbb{Q}$ satisfying the inequality, as desired. □

## 15.2 Proving Solutions to Pell's Equation

**Lemma 15.5**

Let $d \in \mathbb{N}$. There exists some $M > 0$ such that $|x^2 - dy^2| < M$ has infinitely many solutions $x, y \in \mathbb{Z}$. (In fact, we can take $M = 2\sqrt{d} + 1$.)

*Proof.* By the previous lemma (15.4) with $\xi = \sqrt{d}$, there are infinitely many $x/y \in \mathbb{Q}$, with $x, y \in \mathbb{N}$ and $(x, y) = 1$, such that $|x/y - \sqrt{d}| < 1/y^2$, or equivalently $|x - y\sqrt{d}| < 1/y$. Noting $x + y\sqrt{d} = (x - y\sqrt{d}) + 2y\sqrt{d}$, so by Triangle Inequality $|x + y\sqrt{d}| \leq |x - y\sqrt{d}| + |2y\sqrt{d}| < 1/y + 2y\sqrt{d}$, we have

$$\begin{aligned} |x^2 - dy^2| &= |(x - y\sqrt{d})(x + y\sqrt{d})| < 1/y(1/y + 2y\sqrt{d}) \\ &= 1/y^2 + 2\sqrt{d} \leq 2\sqrt{d} + 1. \end{aligned}$$

Taking $M = 2\sqrt{d} + 1$ will do the trick. □

Great, this actually gives us enough to tackle Theorem 15.1!

*Proof of Theorem 15.1.* By Lemma 15.5 above, there exists some $m \in \mathbb{Z}$ such that $x^2 - dy^2 = m$ has infinitely many solutions $x, y \in \mathbb{Z}$. Fix such an $m$, and let the solutions be $(x_n, y_n)$ for $n \in \mathbb{N}$. By Pigeonhole Principle again, there must be two solutions $(x_i, y_i)$ and $(x_j, y_j)$ such that $x_1 \equiv x_2 \pmod{m}$ and $y_1 \equiv y_2 \pmod{m}$.

Let $\alpha = x_i - y_i\sqrt{d}$ and $\beta = x_j - y_j\sqrt{d}$. Note $N(\alpha) = x_i^2 - dy_i^2 = m$, and likewise $N(\beta) = m$.

$$\begin{aligned} \alpha \cdot \overline{\beta} &= \alpha(\overline{\alpha} + \overline{\beta} - \overline{\alpha}) \\ &= \alpha \cdot \overline{\alpha} + \alpha(\overline{\beta} - \overline{\alpha}). \end{aligned}$$

Let $A + B\sqrt{d} = \alpha \cdot \overline{\beta}$. Note that $\alpha \cdot \overline{\alpha} = m$ and, by our choice of $i, j$, we have $m \mid \overline{\beta} - \overline{\alpha}$, so $m \mid A$ and $m \mid B$. Thus, we can write $A + B\sqrt{d} = m(u + v\sqrt{d})$ for some $u, v \in \mathbb{Z}$. We have $N(A + B\sqrt{d}) = N(\alpha)N(\overline{\beta}) = m \cdot m = m^2$, so

$$m^2 = N(m(u + v\sqrt{d})) = N(m)N(u + v\sqrt{d}) = m^2 N(u + v\sqrt{d}),$$

so $N(u+v\sqrt{d}) = 1$. Aha, this is promising – we know elements with norm 1 correspond to a solution of Pell's equation! Furthermore, note that $y_i \neq y_j$, so the coefficient of $\sqrt{d}$ in $\overline{\beta} - \overline{\alpha}$ is $mv = y_j - y_i \neq 0$. In particular, $v \neq 0$.

Choose a solution $(x, y)$ of $x^2 - dy^2 = 1$ with $x, y \in \mathbb{N}$ and $x$ as small as possible. Let $\alpha = x + y\sqrt{d}$, and denote $\beta = u + v\sqrt{d}$. (I know we defined $\alpha$ and $\beta$ earlier in the proof, but we are repurposing them here. Sorry!) We observed above that $N(u + v\sqrt{d}) = u^2 - dv^2 = 1$. Since $\alpha$ is our minimal solution, we have $\beta > \alpha$. This means either $\beta$ lies between $\alpha^n$ and $\alpha^{n+1}$ for some $n \in \mathbb{N}$, or it is equal to $\alpha^n$ on the dot for some $n$. The second case is what we want, so we suppose the first case is true, and we hope to reach a contradiction.

Suppose $\alpha^n < \beta < \alpha^{n+1}$. Multiplying by $\overline{\alpha}^n$ on both sides, we have

$$\begin{aligned} 1 = N(\alpha)^n &= \alpha^n \cdot \overline{\alpha}^n \\ &< \beta \cdot \overline{\alpha}^n < \alpha^{n+1} \cdot \overline{\alpha}^n = \alpha. \end{aligned}$$

Let $\gamma = \beta \cdot \overline{\alpha}^n$. We have $N(\gamma) = N(\beta) \cdot N(\overline{\alpha})^n = 1$, but this contradicts the minimality of $\alpha$, so we must have $\beta = \alpha^n$ for some $n$, as desired. □

# 16 11/06 - More on Pell's Equation

Some recaps of what went on in last lecture, and additional remarks. Notably, $x^2 - dy^2$ is the norm of the element $x + y\sqrt{d} \in \mathbb{Z}[\sqrt{d}]$. Such an element is a unit in $\mathbb{Z}[\sqrt{d}]$ if and only if its norm is a unit in $\mathbb{Z}$, i.e. $x^2 - dy^2 = \pm 1$. Furthermore, if $\alpha$ is the smallest solution such that $N(\alpha) = -1$, then $\alpha^2$ is the smallest solution to Pell's equation $x^2 - dy^2 = 1$. The crux behind this is that all solutions to a certain equation are just powers of the smallest solution, so any solution to $x^2 - dy^2 = 1$ must be an even power of $x^2 - dy^2 = -1$.

## 16.1 Motivation for $\frac{1+\sqrt{d}}{2}$

Let us take a very brief interlude. Consider when $d \equiv 1 \pmod 4$, and let $\alpha = \frac{1+\sqrt{d}}{2}$. We can compute the norm to be $N(\alpha) = \alpha \cdot \overline{\alpha} = \frac{1-d}{4}$; since we assumed $d \equiv 1 \pmod 4$, we have $N(\alpha) \in \mathbb{Z}$. Therefore, the *minimal polynomial* of $\alpha$, that is, the polynomial of smallest degree for which $\alpha$ is a root, is

$$\begin{aligned} (X - \alpha)(X - \overline{\alpha}) &= X^2 - (\alpha + \overline{\alpha})X + \alpha \cdot \overline{\alpha} \\ &= X^2 - X + \frac{1-d}{4} \in \mathbb{Z}[X]. \end{aligned}$$

Thus, when $d \equiv 1 \pmod 4$, it may make sense to consider the ring $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ instead of $\mathbb{Z}[\sqrt{d}]$. This is in fact what happens in number theory! To give a bit more

explanation, what we actually want is to find the "integers" in the field $\mathbb{Q}(\sqrt{d})$. In $\mathbb{Q}$, we can recover $\mathbb{Z}$ creatively via Proposition 8.3: the integers are the *algebraic integers* contained in $\mathbb{Q}$. Likewise, we can define the "integers" of $\mathbb{Q}(\sqrt{d})$ as the algebraic integers contained in this field. Turns out, this ring of integers is $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ when $d \equiv 1 \pmod 4$.

## 16.2 Units of Ring of Integers

Let us consider the units of $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$. First, note that $\mathbb{Z}[\sqrt{d}] \subseteq \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$, so we can take the units on both sides to get $\mathbb{Z}[\sqrt{d}] \subseteq \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]^\times$. But all units of $\mathbb{Z}[\sqrt{d}]^\times$, by the solutions to Pell's equation, are of the form $(x_1 + y_1\sqrt{d})^n$ where $(x_1, y_1)$ is the fundamental solution and $n \in \mathbb{Z}$. Note that given any solution $(x_n, y_n)$, we can also have solution $(\pm x_n, \pm y_n)$. The $(x_n, -y_n)$ can be recovered via $x_n - y_n\sqrt{d} = (x_1 - y_1\sqrt{d})^n = (x_1 + y_1\sqrt{d})^{-n}$, and the other two sign changes can be recovered by just multiplying our unit by $-1$. Thus, we have $\mathbb{Z}[\sqrt{d}]^\times \simeq \mathbb{Z} \times \{\pm 1\}$: the $\mathbb{Z}$ comes from the exponent of the fundamental unit, and $\{\pm 1\}$ is just including all possible signs.

We can write any $\beta \in \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ as $\beta = \frac{A}{2} + \frac{B}{2}\sqrt{d}$, where $A, B \in \mathbb{Z}$ and $A \equiv B \pmod 2$. (Write out elements explicitly if you don't believe this, it just follows from construction.) Suppose $\beta \in \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]^\times$. This means $N(\beta) = \frac{A^2 - dB^2}{4} = \pm 1$, so we are really considering solutions to the equation $A^2 - dB^2 = \pm 4$.

Let us take $d = 5$ as our example, as that is indeed the smallest nontrivial $d > 0$ such that $d \equiv 1 \pmod 4$. (Wow quick maths.) We see that $2 + \sqrt{5}$ satisfies $N(2+\sqrt{5}) = -1$. But if the number 2 is too big for you, then what we could do is look for a solution to $A^2 - 5B^2 = \pm 4$, which is quite easy to find here: $A = B = 1$. Indeed, $N(1+\sqrt{5}) = -4$, and so $N\left(\frac{1+\sqrt{5}}{2}\right) = -1$. This is the fundamental unit in $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$. By the same logic as two paragraphs above, we have an isomorphism $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] \simeq \mathbb{Z} \times \{\pm 1\}$.

But wait, we have a copy of $\mathbb{Z} \times \{\pm 1\} \simeq \mathbb{Z}[\sqrt{d}]^\times$ contained in $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right] \simeq \mathbb{Z} \times \{\pm 1\}$! Even more, it is a (nontrivial) subgroup, and as any nontrivial subgroup of $\mathbb{Z}$ has finite index (i.e. it is of the form $n\mathbb{Z}$ for some $n \neq 0$), we have the $\mathbb{Z}[\sqrt{5}]^\times \subseteq \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]^\times$ has finite index. Here, for $d = 5$, we can see that the index is 3, because

$$\left(\frac{1+\sqrt{5}}{2}\right)^3 = \frac{16 + 8\sqrt{5}}{8} = 2 + \sqrt{5},$$

so we raise the fundamental unit of $\mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ to the third power to get the fundamental unit of $\mathbb{Z}[\sqrt{d}]$.

### 16.3 Finding Solutions when $d \equiv 5 \pmod 8$

It turns out that the index is always 3 for $d \equiv 5 \pmod 8$. We require this congruence because if $A^2 - dB^2 = \pm 4$ for $A, B$ odd, then since $A^2, B^2 \equiv 1 \pmod 8$, it follows that $d \equiv 5 \pmod 8$. (Consequently, when $d \equiv 1 \pmod 8$, we do not have such solutions to this equation, so $\mathbb{Z}[\sqrt{d}]^\times = \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]^\times$.) We can check this as follows: if $A^2 - dB^2 = \pm 4$, where $A, B$ are odd, then $\left(\frac{A}{2} + \frac{B}{2}\sqrt{d}\right)^2 \notin \mathbb{Z}[\sqrt{d}]$ but $\left(\frac{A}{2} + \frac{B}{2}\sqrt{d}\right)^3 \in \mathbb{Z}[\sqrt{d}]$. I will verify the latter first: we have

$$\left(\frac{A}{2} + \frac{B}{2}\sqrt{d}\right)^3 = \frac{1}{8}(A^3 + 3AB^2d + (3A^2B + B^3d)\sqrt{d}),$$

and when $d \equiv 5 \pmod 8$, we have $A^3 + 3AB^2d \equiv A^3 - AB^2 \equiv A(A^2 - B^2) \equiv 0 \pmod 8$ and $3A^2B + B^3d \equiv B(3A^2 - 3B^2) \equiv 0 \pmod 8$. One can run the computations for $\left(\frac{A}{2} + \frac{B}{2}\sqrt{d}\right)^2$ and show that it does in fact live in $\mathbb{Z}[\sqrt{d}]$.

This neat fact about the index being 3 gives us power in computations! Let us take the example when $d = 29$.

**Example 16.1** ($d = 29$)

We have $5^2 - 29 \cdot 1^2 = -4$, so $u = \frac{5+\sqrt{29}}{2}$ is a fundamental unit of $\mathbb{Z}\left[\frac{1+\sqrt{29}}{2}\right]$. Then, $\left(\frac{5+\sqrt{29}}{2}\right)^3$ is the smallest solution to the equation $x^2 - 29y^2 = -1$, and so the sixth power is the smallest solution to Pell's equation $x^2 - 29y^2 = 1$. Finding this smallest solution straight from the wild, without this $\frac{1+\sqrt{29}}{2}$ business to guide us, would be very difficult! (It is equal to $9801 + 1820\sqrt{29}$.)

# 17 11/10 - Dirichlet's Theorem, an Introduction

We started this topic at the end of the last lecture, but it is more fitting to start it for a new section. We begin this topic with an absolute banger of a theorem, credited to Dirichlet.

**Theorem 17.1** (Dirichlet)

Suppose $a, m \in \mathbb{N}$ with $(a, m) = 1$. The set of prime $p$ congruent to $a \pmod m$ is infinite. In other words, the arithmetic progression $a, a + m, a + 2m, \ldots$ contains infinitely many primes.

A baby example of this is when $m = 4$ and $a = 3$, which we can prove in a more grounded way a la Euclid.

**Proposition 17.2**

There are infinitely many primes congruent to 3 (mod 4).

*Proof.* Suppose there are finitely many; let $\{p_1, \dots, p_r\}$ be the complete list. Consider $N = 4p_1 \cdots p_r + 3$. None of the $p_i$'s divides $N$, so all of its prime factors must be 1 mod 4. But the product of numbers 1 mod 4 will still be 1 mod 4, but $N \equiv 3 \pmod 4$ by construction, so we reach a contradiction. □

## 17.1 Riemann Zeta Function

We define the **Riemann zeta function** by

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

For this class, we will just concern ourselves with $s \in \mathbb{R}_{>1}$, but its real power comes from taking $s \in \mathbb{C}$. (This requires some knowledge of complex analysis, which is a *really* fun subject (everything is so nice in complex analysis!) but goes beyond the scope of this course. Take Math 113 if you're interested, though!)

We can check the sum converges for $s > 1$. By the integral test, we can bound

$$\begin{aligned} (n+1)^{-s} &< \int_n^{n+1} t^{-s}\,dt < n^{-s} \\ \implies \zeta(s) - 1 &< \int_1^{\infty} t^{-s}\,dt < \zeta(s), \\ \int_1^{\infty} t^{-s}\,dt &= \frac{-t^{-s+1}}{s-1}\bigg|_1^{\infty} = \frac{1}{s-1} < \infty. \end{aligned}$$

Although $\zeta(s)$ diverges for $s = 1$ (it is then the harmonic series, which we've shown in one of the first lectures on the Prime Number Theorem that it diverges), we see that the divergence is "not too bad." (For people who know complex analysis, the following says the pole of $\zeta(s)$ at $s = 1$ is simple with residue 1.)

**Proposition 17.3**

$\lim_{s \to 1^+} (s-1)\zeta(s) = 1$.

*Proof.* From the second line of inequalities above, multiplying by $s-1$ gives us $\zeta(s)(s-1) - (s-1) < 1 < \zeta(s)(s-1)$, so both $1 < \zeta(s)(s-1)$ and $\zeta(s)(s-1) < s$. As $s \to 1^+$, we get the desired. □

I was trying to avoid discussing what happens when $s \in \mathbb{C}$, but Kisin is going full force with this complex discussion, so let me try to provide some explanation. We can define the exponential function $e^z$ for $z \in \mathbb{C}$ by $e^{x+iy} = e^x \cdot e^{iy}$. The first part $e^x$ is just the exponential function for the reals, which we know and love. The second is just $e^{i\theta} = \cos\theta + i\sin\theta$; note, importantly, that $|e^{i\theta}| = 1$, so the magnitude of $e^z$ really comes from just the $e^x = e^{\operatorname{Re} z}$ part. Writing $n = e^{\log n}$, we have, for $s = \alpha + i\beta$,

$$n^{-s} = e^{-s\log n} = e^{-\alpha \log n} \cdot e^{-i\beta \log n} = n^{-\alpha} e^{-i\beta \log n}.$$

By Triangle Inequality, we have $|\zeta(s)| \le \sum_{n=1}^{\infty} |n^{-s}| = \sum_{n=1}^{\infty} |n^{-\operatorname{Re} s}|$, so in fact our check for convergence when $s > 1$ really checked it for when $\operatorname{Re} s > 1$.

This means we can define the Riemann zeta function without tears for the "half"-plane where $\operatorname{Re} s > 1$. The pesky pole (i.e., place where $\zeta(s)$ goes to infinity) at $s = 1$ prevents us from doing better. However, we just demonstrated that this pole is "not bad": in fact, it is perhaps the most nicely behaved pole possible. There is a method in complex analysis called *analytic continuation* which allows to bypass this pole and define $\zeta(s)$ for the whole complex plane minus $s = 1$.

The next question is, then, where are the zeroes of this Riemann zeta function? This is the content of the famous **Riemann Hypothesis**, one of the Millennium Problems. First, one can show that there are zeroes at the negative even integers $s = -2n$ for $n \in \mathbb{N}$. These are not hard to show, and so they are called the "trivial zeroes." Besides these, it is conjectured that the zeroes lie on the vertical line $\operatorname{Re} s = 1/2$, which is weird and seems to come out of nowhere. This has been checked for really large values of $|s|$, so people believe it to be true, but no proof has been provided yet.

Here is one formal consequence of Proposition 17.3 above.

**Corollary 17.4**
$\lim_{s\to 1^+} \frac{\log \zeta(s)}{\log(s-1)^{-1}} = 1.$

*Proof.* Let $\rho(s) = (s-1)\zeta(s)$, so $\log \rho(s) = \log(s-1) + \log \zeta(s)$. Dividing by $\log(s-1)^{-1} = -\log(s-1)$ on both sides, we have

$$\frac{\log \rho(s)}{\log(s-1)^{-1}} = -1 + \frac{\log \zeta(s)}{\log(s-1)^{-1}}.$$

Taking the limit as $s \to 1^+$, we note from Proposition 17.3 that $\lim_{s\to 1^+} \rho(s) = 1$, so

$$\lim_{s\to 1^+} \frac{\rho(s)}{\log(s-1)^{-1}} = 0 = -1 + \frac{\log \zeta(s)}{\log(s-1)^{-1}},$$

and the conclusion follows. □

## 17.2 Euler Factorization

So we've mentioned before that the Riemann Hypothesis is related to the primes in some way. But how? So far, in our definition $\zeta(s) = \sum_{n\geq 1} n^{-s}$, primes appear nowhere. But some really smart guy named Euler (surprise!) made the following observation:

**Proposition 17.5** (Euler Product)
For $\operatorname{Re} s > 1$,

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1} = (1 - 2^{-1})^{-1}(1 - 3^{-1})^{-1} \cdots .$$

Before we begin the formal proof, I think it is useful to just convince yourself, perhaps slightly non-rigorously, that this is true. A good place to start would be a simpler problem such as:

**Exercise 17.6.** What is the sum

$$\sum_{n=3^a 5^b} \frac{1}{n},$$

where the sum is taken over all $n$ with only 3 and 5 in the prime factorization?

Once you get this, it is not too difficult to see how this generalizes when we sum over all $n \in \mathbb{N}$.

*Proof.* Our good old sum of infinite geometric series tells us that

$$(1 - p^{-s})^{-1} = \frac{1}{1 - p^{-s}} = 1 + p^{-s} + p^{-2s} + \cdots .$$

Take some $N \in \mathbb{N}$. We know, trivially, that any $n \leq N$ must factor into primes also at most $N$. Thus, by unique factorization,

$$\prod_{p\leq N} (1 - p^{-s})^{-1} = \prod_{p\leq N} (1 + p^{-s} + p^{-2s} + \cdots) = \sum_{n\leq N} n^{-s} + R_N(s),$$

where $R_N(s) \leq \sum_{n>N} n^{-s}$. Taking the limit on both sides as $N \to \infty$, we conclude

$$\begin{aligned}\prod_{p} (1 - p^{-s})^{-1} &= \lim_{N\to\infty} \prod_{p\leq N} (1 - p^{-s})^{-1} \\ &= \lim_{N\to\infty} \sum_{n\leq N} n^{-s} + R_N(s) \\ &= \zeta(s) + \lim_{N\to\infty} R_N(s) = \zeta(s),\end{aligned}$$

as desired. □

This is perhaps the simplest example of an Euler Product. There are many classes of functions which exhibit an Euler Product similar to this; if it does, then because this factorization is so nice, it suggests that the function has some really nice properties/deeper connections to number theory.

One upshot of expressing the zeta function as a product is that when we take the logarithm, we can split it up based on the terms in the product (we can't do anything with $\log(x+y)$, but we know $\log(xy) = \log x + \log y$). We see this here:

**Proposition 17.7**
For $\operatorname{Re} s > 1$, $\log \zeta(s) = \sum_p p^{-s} + R(s)$ for some function $R(s)$ bounded near $s = 1$.

*Proof.* We will use the fact from calculus

$$-\log(1-x) = x + \frac{x^2}{2} + \cdots = \sum_{n\geq 1} \frac{x^n}{n}.$$

From the Euler factorization, we can write

$$\zeta(s) = \prod_{p\leq N} (1-p^{-s})^{-1} \lambda_N(s)$$

for some function $\lambda_N$ with $\lim_{N\to\infty} \lambda_N(s) = 1$. Taking the logarithm on both sides and using our Taylor series expansion above gives

$$\begin{aligned}
\log \zeta(s) &= \sum_{p\leq N} -\log(1-p^{-s}) + \log \lambda_N(s) \\
&= \sum_{p\leq N} \sum_{m\geq 1} \frac{p^{-ms}}{m} + \log \lambda_N(s) \\
\implies \lim_{N\to\infty} \log \zeta(s) = \log \zeta(s) &= \lim_{N\to\infty} \sum_{p\leq N} \sum_{m\geq 1} \frac{p^{-ms}}{m} + \log \lambda_N(s) \\
&= \sum_p \sum_{m\geq 1} \frac{p^{-ms}}{m} + \log 1 \\
&= \sum_p p^{-s} + \sum_p \sum_{m\geq 2} \frac{p^{-ms}}{m}.
\end{aligned}$$

The double sum at the very end is our $R(s)$. Looking at the last sum separately, we can bound

$$\sum_{m\geq 2} \frac{p^{-ms}}{m} \leq \sum_{m\geq 2} p^{-ms} = p^{-2s}(1 + p^{-s} + \cdots) = p^{-2s}(1-p^{-s})^{-1}.$$

Therefore,

$$\begin{aligned} R(s) &\le \sum_p p^{-2s}(1-p^{-s})^{-1} \le (1-2^{-s})^{-1}\sum_p p^{-2s} \\ &\le (1-2^{-s})^{-1}\zeta(2) \le 2\zeta(2), \end{aligned}$$

which is just a constant, and hence clearly bounded near $s = 1$. □

## 17.3 Dirichlet Density

Our goal is to prove Dirichlet's Theorem (17.1). The way we will do this is, roughly speaking, show that for any $a$ such that $(a, m) = 1$, the "proportion of primes which are congruent to $a$ mod $m$ is nonzero." This implies there are infinitely many such primes, since there are an infinite number of primes in total.

But what do we mean exactly by "proportion" here? How do we define fractions when we are counting over an infinite set? We do this by defining something called the **Dirichlet density**.

**Definition 17.8** (Dirichlet Density)**.** Let $P$ be a set of primes. Then,

$$d(P) = \lim_{s\to 1^+} \frac{\sum_{p\in S} p^{-s}}{\log(s-1)^{-1}}$$

is called the **Dirichlet density of $P$**, if it exists.

To illustrate what is going on, recall Corollary 17.4, which told us a certain limit is 1. We can rewrite it as

$$\lim_{s\to 1^+} \frac{\log\zeta(s)}{\log(s-1)^{-1}} = \lim_{s\to 1^+} \frac{\sum_p p^{-s}}{\log(s-1)^{-1}} = 1.$$

This is the simplest example of Dirichlet density: this tells us that the density of all primes in, well, all primes is 1.

Dirichlet's Theorem cares about primes congruent to $a \bmod m$, so let us define $P(n, m)$ as the set of primes $p$ such that $p \equiv a \pmod m$. Then, we can reformulate Dirichlet's Theorem as:

**Theorem 17.9** (Dirichlet)
$d(P(n, m)) = 1/\phi(m)$.

In fact, this is stronger than our original statement of Dirichlet's Theorem, as it not only guarantees infinitely many primes in each residue class, but they are distributed equally.

**Example 17.10**

Taking $m = 4$ and noting that the only possible residue classes are 1 and 3 mod 4 (excluding the prime $p = 2$), we have $d(P(1,4)) = d(P(3,4)) = 1/2$, i.e., "half" of the primes are in each residue class.

**Remark 17.11.** A couple of remarks. If $P_1, P_2$ are disjoint sets of primes, then the definition tells us $d(P_1 \cup P_2) = d(P_1) + d(P_2)$, which should agree with our intuition. If $P$ is finite, then it also makes sense that $d(P) = 0$, which we can observe from the definition of the Dirichlet density. (If the set is finite, the numerator is bounded, but the denominator goes to infinity.) Thus, since Dirichlet's Theorem as described above guarantees $d(P(n,m)) = 1/\phi(m) > 0$, it follows that there are infinitely many primes in $P(n,m)$.

## 17.4 Dirichlet *L*-functions

Tackling the theorem directly is daunting, so we will first provide a proof for when $m = 4$. We will use something called **Dirichlet characters** (wow, this Dirichlet guy did a lot of stuff huh).

**Remark 17.12.** In general, whenever you see a "character" (especially in number theory), it is a group homomorphism to some multiplicative group. See below for an example.

We will define a function $\chi : \mathbb{Z} \to \{0, \pm 1\}$ where $2\mathbb{Z} \mapsto 0$ (the evens map to 0) and any odd $a \in \mathbb{Z} \setminus 2\mathbb{Z}$ maps to its residue mod 4. For instance, $\chi(17) = 1$ and $\chi(23) = -1$, while $\chi(122) = 0$. It is not difficult to see that this $\chi$ is multiplicative, that is, $\chi(mn) = \chi(m)\chi(n)$.

Given this character, we can now generalize the Riemann zeta function to what we call **Dirichlet *L*-functions**. In this $m = 4$ case, this is given by

$$L(s,\chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s} = 1 - 3^{-s} + 5^{-s} - 7^{-s} + \cdots .$$

(For any $m$, we can define a Dirichlet character $\chi$, and then the definition of $L(s,\chi)$ would be the same.)

Since $\chi$ takes on nonzero values $\{\pm 1\}$, we have $|\chi(n)n^{-s}| \leq |n^{-s}|$, so Triangle Inequality tells us

$$|L(s,\chi)| = \left| \sum_{n \geq 1} \chi(n) n^{-s} \right| \leq \sum_{n \geq 1} |\chi(n) n^{-s}| \leq \sum_{n \geq 1} |n^{-s}|,$$

so $L(s,\chi)$ converges for $\operatorname{Re} s > 1$.

Even more, this $L$-function follows the story of the Riemann zeta function by admitting an Euler factorization. This is indeed the case because $\chi$ is multiplicative. The factorization looks like

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

## 17.5 Dirichlet's Theorem for $m = 4$

Now, we will tackle Dirichlet's Theorem for $m = 4$, which – note – does not mention this Dirichlet character anywhere. But we will see it is the key tool in the proof.

**Proposition 17.13**
$d(P(1,4)) = d(P(3,4)) = 1/2$.

*Proof.* Let $\zeta^*(s) = \sum_{2\nmid n} n^{-s}$. For any $n = 2^k m$ for $m$ odd, we make the simple observation that $n^{-s} = 2^{-ks}m^{-s}$. Thus, we may factor

$$\zeta(s) = \zeta^*(s)(1 + 2^{-s} + 2^{-2s} + \cdots) = \zeta^*(s)(1-2^{-s})^{-1}.$$

(A more direct way to see this is that, just like how we can construct an Euler product for $\zeta(s)$, we can do the same for $\zeta^*(s)$, except we omit $p = 2$ since we are only summing over odd $n$.)

Akin to Proposition 17.7, with just omitting the $p = 2$ term in the sum, we have

$$\log \zeta^*(s) = \sum_{p\neq 2} p^{-s} + R_2(s),$$

where $R_2(s)$ is bounded near $s = 1$. We can do the same to $L(s,\chi)$, which is similar to $\zeta^*(s)$ but where the coefficients in the sum expansion alternate between $\pm 1$. Again, akin to Proposition 17.7, we have

$$\log L(s,\chi) = \sum_p \chi(p)p^{-s} + R_\chi(s),$$

where $R_\chi(s)$ is bounded near $s = 1$.

Recall the definition of Dirichlet density; for $d(P(1,4))$, the sum on on the numerator is $\sum_{p\equiv 1 \pmod 4} p^{-s}$. How can we get this sum from our two sums above? Well, since

$\chi(p) = 1$ iff $p \equiv 1 \pmod 4$ and the main sum in $\log \zeta^*(s)$ has coefficients all 1, we have

$$\begin{aligned} 2\sum_{p\equiv 1(4)} p^{-s} &= \sum_{p\neq 2} p^{-s} + \sum_p \chi(p)p^{-s} \\ 2\sum_{p\equiv 3(4)} p^{-s} &= \sum_{p\neq 2} p^{-s} - \sum_p \chi(p)p^{-s} \\ \implies \log\zeta^*(s) + \log L(s,\chi) &= 2\sum_{p\equiv 1(4)} p^{-s} + R(s) \qquad (*) \\ \log\zeta^*(s) - \log L(s,\chi) &= 2\sum_{p\equiv 3(4)} p^{-s} + R(s), \end{aligned}$$

where the $R(s)$ error terms are bounded near $s = 1$. (Here, the two $R(s)$'s are different, I am just abusing notation because they don't really matter.)

We can construct crude bounds for $\log L(s,\chi)$. We can group the elements on our sum in two simple ways:

$$\begin{aligned} L(s,\chi) &= 1 - 3^{-s} + 5^{-s} - 7^{-s} + \cdots \\ &= (1-3^{-s}) + (5^{-s} - 7^{-s}) + \cdots > 2/3, \\ L(s,\chi) &= 1 - (3^{-s} - 5^{-s}) - (7^{-s} - 9^{-s}) - \cdots < 1, \end{aligned}$$

so $2/3 < L(s,\chi) < 1$ for $s > 1$. Taking the log, we have $\log 2/3 < \log L(s,\chi) < 0$. In particular, this means $\log L(s,\chi)$ is finite, so we have

$$\lim_{s\to 1^+} \frac{\log\zeta^*(s) + \log L(s,\chi)}{\log(s-1)^{-1}} = \lim_{s\to 1^+} \frac{\zeta^*(s)}{\log(s-1)^{-1}} = 1.$$

But through Equation (*), we see that this is equal to

$$1 = \lim_{s\to 1^+} \frac{2\sum_{p\equiv 1(4)} p^{-s} + R(s)}{\log(s-1)^{-1}} = 2d(P(1,4)),$$

so $d(P(1,4)) = 1/2$ and consequently $d(P(3,4)) = 1/2$ as well. □

# 18 11/13 - Dirichlet Characters

Professor Kisin is disappointed that not more people are showing up to lecture. To encourage attendance, the final exam will have a question where you need to provide your favorite Kisin joke or anecdote. For instance, he gave a really funny anecdote at the beginning of this class, but I am not allowed to share it.

Last time, we proved Dirichlet's theorem for $m = 4$. For the next three to four lectures, we will prove Dirichlet's theorem in full generality. The central characters (pun fully intended) in this story are the Dirichlet characters, so we begin there.

Dirichlet characters are very important, but developing the theory might feel a bit like eating vegetables. Bear with us for a bit.

For $m = 4$, we defined this character $\chi : \mathbb{Z} - 2\mathbb{Z} \to (\mathbb{Z}/4\mathbb{Z})^\times \xrightarrow{\simeq} \{\pm 1\}$ where an element outputs its residue mod 4, and we could extend this to be 0 on the even integers. This is a baby example of the Dirichlet character for a general $m$. Consider a map

$$\overline{\chi} : (\mathbb{Z}/m\mathbb{Z})^\times \to \mathbb{C}^\times$$

which is a group homomorphism (i.e. $\overline{\chi}(ab) = \overline{\chi}(a)\overline{\chi}(b)$, and consequently $\overline{\chi}(1) = 1$). Recall $|(\mathbb{Z}/m\mathbb{Z})^\times| = \phi(m)$, so by multiplicativity, we have

$$\overline{\chi}(a)^{\phi(m)} = \overline{\chi}(a^{\phi(m)}) = \overline{\chi}(1) = 1,$$

so $\overline{\chi}(a)$ must be a $\phi(m)^{\text{th}}$ root of unity. Explicitly, this means

$$\overline{\chi}(a) = e^{2\pi i k/\phi(m)}$$

for some $k \in \mathbb{Z}$.

For instance, let $m = p$ be prime. We know $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic, so $(\mathbb{Z}/p\mathbb{Z})^\times \simeq \mathbb{Z}/(p-1)\mathbb{Z}$. Thus, we can consider $\overline{\chi} : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{C}^\times$ as a function from $\mathbb{Z}/(p-1)\mathbb{Z}$, and the map would be

$$\begin{aligned} \overline{\chi} : \mathbb{Z}/(p-1)\mathbb{Z} &\to \mathbb{C}^\times \\ a &\mapsto e^{2\pi i a k/(p-1)} \end{aligned}$$

for some $k \in \mathbb{Z}$.

Like in the $m = 4$ case, we can extend these characters to be a Dirichlet character mod $m$.

**Definition 18.1** (Dirichlet character)**.** A **Dirichlet character mod** $m$ is a map $\chi : \mathbb{Z} \to \mathbb{C}$ such that for $a \in \mathbb{Z}$,

1. If $(a, m) \neq 1$, then $\chi(a) = 0$;
2. otherwise, there exists some $\overline{\chi} : (\mathbb{Z}/m\mathbb{Z})^\times \to \mathbb{C}^\times$ such that $\chi(a) = \overline{\chi}(a \bmod m)$.

**Example 18.2**

The simplest character is the trivial one, where $\chi$ extends from the trivial character $\overline{\chi} : (\mathbb{Z}/m\mathbb{Z})^\times \to \mathbb{C}^\times$ where $\overline{\chi}(a) = 1$ for all $a \in (\mathbb{Z}/m\mathbb{Z})^\times$.

## 18.1 Dual Group

Great, we have these individual characters. We will slightly generalize first, then consider the set of all Dirichlet characters and see what structures come with this set.

Let $A$ be a finite abelian group. This just means $A$ is finite as a set, and abelian means that $a \cdot b = b \cdot a$ for any $a, b \in A$. Now, consider the set of characters $\widehat{A} = \{\chi : A \to \mathbb{C}^\times \mid \chi(ab) = \chi(a)\chi(b)\}$. IT follows that $\widehat{A}$ is an abelian group, where the group structure is given by $(\chi_1 \cdot \chi_2)(a) \coloneqq \chi_1(a) \cdot \chi_2(a)$ for $\chi_1, \chi_2 \in \widehat{A}$, and $\chi^{-1}(a) \coloneqq \chi(a)^{-1}$.

**Example 18.3**

Consider $A \cong \mathbb{Z}/n\mathbb{Z}$, and let $r$ be a generator of $\mathbb{Z}/n\mathbb{Z}$. (This amounts to just having $(r, n) = 1$, as we've seen repeatedly.) Like we worked out before, $\chi(r) \in \mathbb{C}^\times$ has to satisfy $\chi(r)^n = \chi(r^n) = \chi(1) = 1$, so $\chi(r) = e^{2\pi ik/n}$ for some $k \in \mathbb{Z}$.

Denote $\zeta_m = e^{2\pi ik/n}$. Then, $\chi(r^j) = \zeta_n^j$ by multiplicativity, and this completely determines $\chi$. But all characters mod $n$ must be of this form, so it is really dependent on the choice of $k$ in the exponent. As $k$ ranges across $\mathbb{Z}/n\mathbb{Z}$ (they technically range along all of $\mathbb{Z}$, but $k$ and $k + n$ produce the same character), we conclude $\widehat{A} \simeq \mathbb{Z}/n\mathbb{Z}$, so in fact $A \simeq \widehat{A}$. This may not seem so significant at first, but it is remarkably deep.

It turns out that this isomorphism $A \simeq \widehat{A}$, as demonstrated when $A \simeq \mathbb{Z}/n\mathbb{Z}$ above, is a taste of the more general result.

**Lemma 18.4** (Dual is Isomorphism)

There is a non-canonical isomorphism $A \simeq \widehat{A}$. In particular, $|A| = |\widehat{A}|$.

*Proof.* We proved this already for cyclic groups, as any finite cyclic group is isomorphic to some $\mathbb{Z}/n\mathbb{Z}$.

In general, the Classification of Finite Abelian Groups tells us that any finite abelian group is of the form

$$A \simeq \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z},$$

where the group operation is just addition component-wise.

Since these components are independent of each other, any $\chi \in \widehat{A}$ must be a character on each of its components, and given characters on each components, we can patch it up via multiplication to produce a character on all of $A$. Thus, specifying $\chi \in \widehat{A}$ is equivalent to giving characters $\chi_i \in \widehat{\mathbb{Z}/n_i\mathbb{Z}}$ for $1 \le i \le r$, as $\chi_{\mathbb{Z}/n_i\mathbb{Z}} = \chi_i$ and we can reconstruct $\chi$ from the $\chi_i$'s via $\chi(a_1, \ldots, a_r) = \chi_1(a_1)\chi_2(a_2)\cdots\chi_r(a_r)$. Now, we use our work done for cyclic groups to get

$$\widehat{A} \simeq \prod_{i=1}^{r} \widehat{\mathbb{Z}/n_i\mathbb{Z}} \simeq \prod_{i=1}^{r} \mathbb{Z}/n_i\mathbb{Z} \simeq A.$$

□

This might feel familiar if you've learned some linear algebra: the dual of a vector space $V$ is isomorphic to $V$, but not canonically (it requires the choice of a basis). Similarly, we call $\widehat{A}$ as the **dual** of $A$. But we do know from linear algebra that the dual of the dual of $V$ is canonically isomorphic to $V$. We replicate the same result here.

**Corollary 18.5** (Double Dual is Canonical Isomorphism)

There is a cnonical isomorphism $A \simeq \hat{\hat{A}}$.

*Proof.* If $a \in A$, we wish to produce an element of $\hat{\hat{A}}$, which is a map $\chi : \widehat{A} \to \mathbb{C}^\times$. We have a really choice-free way of doing so: we define $\psi_a : \chi \mapsto \chi(a)$. This gives us our isomorphism $A \xrightarrow{\simeq} \hat{\hat{A}}$ where $a \mapsto \psi_a$.

We check that this is a bijective homomorphism. We start with the latter: we have, for $a, b \in A$,

$$\psi_{ab}(\chi) = \chi(ab) = \chi(a)\chi(b) = \psi_a(\chi)\psi_b(\chi).$$

It now remains to show $a \mapsto \psi_a$ is injective. This suffices, since we know $|A| = |\widehat{A}| = |\hat{\hat{A}}|$, so if we get an injective map $A \to \hat{\hat{A}}$, then it must be surjective as well. Injectivity amounts to proving that for any $1 \neq a \in A$, then $\psi_a$ is not the trivial map, or equivalently there is some $\chi \in \widehat{A}$ such that $\chi(a) \neq 1$.

Via the decomposition $A \simeq \mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_r\mathbb{Z}$ (where $a \mapsto (a_1, \ldots, a_r)$), we can decompose $\chi$ into characters $(\chi_1, \ldots, \chi_r)$. So now we have reduced this problem to the cyclic group case. If $a \neq 1$, then $a_i \neq 1$ for some $i$. Select $\chi_i$'s such that for $j \neq i$, $\chi_j = 1$ is the trivial character, and $\chi_i(a) \neq 1$. Then, $\chi(a) = \chi_1(a_1) \cdots \chi_r(a_r) = \chi_i(a_i) \neq 1$, and we win. □

## 18.2 Orthogonality Relations

We have seen a result before similar in flavor to the one below, when $\chi$ was the Legendre symbol.

**Proposition 18.6**

Let $A$ be a finite abelian group, and let $n = |A|$. If $\chi, \psi \in \widehat{A}$, then

$$\sum_{a \in A} \chi(a)\overline{\psi(a)} = n \cdot \delta_{\chi,\psi} = \begin{cases} n & \text{if } \chi = \psi \\ 0 & \text{if } \chi \neq \psi \end{cases}.$$

Likewise, for $a, b \in A$, then

$$\sum_{\chi \in \widehat{A}} \chi(a)\overline{\chi(b)} = n \cdot \delta_{a,b} = \begin{cases} n & \text{if } a = b \\ 0 & \text{if } a \neq b \end{cases}.$$

**Remark 18.7.** Note that for $a \in A$ and $\chi \in \widehat{A}$, we have $\chi(a)^n = \chi(a^n) = \chi(1) = 1$, so in particular $|\chi(a)| = 1$. In this case, we have $\overline{\chi(a)} = 1/\chi(a)$. We will use this repeatedly.

Before we start the proof, we will prove the following lemma, which we have proven before when the character is the Legendre symbol. In this lemma, 1 represents the trivial character.

**Lemma 18.8**

If $\chi \in \widehat{A}$, then

$$\sum_{a \in A} \chi(a) = \begin{cases} n & \text{if } \chi = 1 \\ 0 & \text{if } \chi \neq 1 \end{cases}.$$

*Proof.* If $\chi = 1$, this is clear from $n = |A|$. If $\chi \neq 1$, then $\chi(b) \neq 1$ for some $b \in A$. We now take advantage of the group structure of $A$:

$$\chi(b) \cdot \sum_{a \in A} \chi(a) = \sum_{a \in A} \chi(ab) = \sum_{a \in A} \chi(a),$$

as the multiplication-by-$b$ map $A \to A$ is an isomorphism. The conclusion follows from $\chi(b) \neq 1$. □

Now we prove Proposition 18.6.

*Proof.* Using $\overline{\psi(a)} = \psi(a)^{-1}$, we have

$$\begin{aligned} \sum_{a \in A} \chi(a)\overline{\psi(a)} &= \sum_{a \in A} \chi(a)\psi(a)^{-1} = \sum_{a \in A} (\chi \cdot \psi^{-1})(a) \\ &= \begin{cases} n & \text{if } \chi \cdot \psi^{-1} = 1 \\ 0 & \text{if } \chi \cdot \psi^{-1} \neq 1 \end{cases}, \end{aligned}$$

where the last equality follows from Lemma 18.8 above. The first statement follows. For the second statement, we can apply the first relation to $\hat{\hat{A}}$ and use the isomorphism $\hat{\hat{A}} \simeq A$ to get our desired result. □

We now bring ourselves back to $(\mathbb{Z}/m\mathbb{Z})^\times$: we will apply Proposition 18.6 to $A = (\mathbb{Z}/m\mathbb{Z})^\times$.

**Corollary 18.9**

If $\chi, \psi$ are Dirichlet characters mod $m$, then

$$\sum_{a=0}^{m-1} \chi(a)\overline{\psi(a)} = \phi(m)\delta_{\chi,\psi}.$$

Likewise, if $a, b \in \mathbb{Z}$, where $(a, m) = (b, m) = 1$, then

$$\sum_{\chi \text{ Dirichlet}} \chi(a)\overline{\chi(b)} = \phi(m)\delta_{a,b}.$$

*Proof.* Uh oh, we have a notation conflict here with the overline bar, but we will close our eyes and push forward. Let $\chi$ and $\psi$ be extensions of characters $\overline{\chi}, \overline{\psi}$ on $(\mathbb{Z}/m\mathbb{Z})^\times$. Then, the Proposition tells us that

$$\sum_{a=0}^{m-1} \chi(a)\overline{\psi(a)} = \sum_{\overline{a} \in (\mathbb{Z}/m\mathbb{Z})^\times} \overline{\chi}(\overline{a})\overline{\overline{\psi}(\overline{a})}.$$

□

## 18.3 Dirichlet *L*-functions

Okay, we are done eating our vegetables. Let's see how this pays off.

Fix $m \in \mathbb{N}$, and let $\chi : \mathbb{Z} \to \mathbb{C}$ be a Dirichlet character. Consider the Dirichlet $L$-function

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

which we showed converges for when $\operatorname{Re} s > 1$. We have the Euler factorization

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

Let us see what happens when $\chi = 1$ is the trivial character. Then, we can recover the Riemann zeta function minus finitely many primes; specifically,

$$L(s, 1) = \prod_{p \nmid m} (1 - p^{-s})^{-1} = \zeta(s) \prod_{p \mid m} (1 - p^{-s}),$$

which is exactly how we defined $\zeta^*$ when $m = 4$.

# 19 11/20 - Dirichlet's Theorem, Part II

**Author's Note 19.1.** There was class on Friday, but since I was not here because of Harvard-Yale and the intersection of people in today's class and Friday's class is exactly one, Kisin decided to repeat this lecture.

Kisin starts this lecture with a recap of the definition of Dirichlet characters (§18), the orthogonality conditions (§18.2), and Dirichlet $L$-functions (§18.3). So in short, read the previous lecture notes!

We will pick up from the last equation line from the last lecture, namely the $L$-function for the trivial character $\chi = 1$. Recall Proposition 17.3, which stated $\lim_{s\to 1^+}(s-1)\zeta(s) = 1$. Then, we can evaluate

$$\begin{aligned}\lim_{s\to 1^+}(s-1)L(s,1) &= \lim_{s\to 1^+}(s-1)\zeta(s)\prod_{p\mid m}(1-p^{-s}) \\ &= 1\cdot\prod_{p\mid m}(1-p^{-1}) \\ &= \phi(m)/m,\end{aligned}$$

where $\phi(m)$ is, as it always has been, the Euler totient function. (The last equality just follows from the formula we gave for $\phi$, see Lemma 4.7.)

We want to consider $\log L(s,\chi)$, in a sensibility akin to Corollary 17.4. This is relevant because, ultimately, we care about the density of certain primes, and so we need an expression that reflects the Dirichlet density. Given our factorization of $L(s,1)$, we can write

$$\lim_{s\to 1^+}\frac{\log L(s,1)}{\log(s-1)^{-1}} = \lim_{s\to 1^+}\frac{\log\zeta(s)}{\log(s-1)^{-1}} + \lim_{s\to 1^+}\frac{\log\prod_{p\mid m}(1-p^{-s})}{\log(s-1)^{-1}}.$$

The product in the numerator of the latter term is a finite product that goes to some nonzero value when $s\to 1^+$, so the latter term vanishes as $s\to 1^+$. (The denominator is unbounded.) Thus, we have

$$\lim_{s\to 1^+}\frac{\log L(s,1)}{\log(s-1)^{-1}} = \lim_{s\to 1^+}\frac{\log\zeta(s)}{\log(s-1)^{-1}} = 1,$$

where the last equality again follows from Corollary 17.4.

We must proceed with caution when consider $\log L(s,\chi)$, though, since these $L$-functions are complex-valued, but there is no single-valued log function on $\mathbb{C}$. To see why, consider $\log z$ as $z$ goes around the unit circle. We know we can parameterize

$z = e^{i\theta}$ as $\theta \in \mathbb{R}$. Start with $z = e^{i\cdot 0} = 1$. As we go around the circle, we reach 1 again when $z = e^{2\pi i}$. But log is a continuous function, so from $e^{i\cdot 0}$ to $e^{2\pi i}$, log goes from 0 to $2\pi i$. But then we have obtained two distinct values for $\log 1$, which should never happen for a well-defined function!

The way to rectify this is to take what we call a **branch cut** in complex analysis. Basically, we have an obstruction on well-definedness when we make a full circle around the origin. To avoid this, we basically remove a ray from the plane (cut out a branch, if you will) so that we can never have a full revolution around the origin. For instance, if we remove the positive real axis $\mathbb{R}^+$, then now we can play the same game without any problems. Technically, log is not defined on $1 \in \mathbb{R}^+$ anymore, but if we were to go from $1 + i\varepsilon$ to $1 - i\varepsilon$, then we would traverse through the range $(0, 2\pi i)$, and the branch cut allows us to make the discontinuous jump from $2\pi i$ to 0 again.

So now we return to $\log L(s, \chi)$. We can recover the Taylor series of $\log(1 - x)$: we know $\log(1-x) = \int \frac{1}{1-x} = \int 1 + x + x^2 + \cdots = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots$. In this vein, denote

$$G(s, \chi) = \sum_p \sum_{k=1}^{\infty} \frac{1}{k} \chi(p)^k p^{-ks}.$$

We claim that this is the logarithm of $L(s, \chi)$.

**Lemma 19.2**

$G(s, \chi)$ converges absolutely for $\operatorname{Re} s > 1$, and $\exp(G(s, \chi)) = L(s, \chi)$ for $\operatorname{Re} s > 1$.

*Proof.* We first address convergence. It is not hard to see $|1/k \cdot \chi(p)p^{-ks}| \leq p^{-ks}$. Therefore,

$$\begin{aligned} |G(s, \chi)| &\leq \sum_p \sum_{k=1}^{\infty} |1/k \chi(p) p^{-ks}| \\ &\leq \sum_p \sum_{k=1}^{\infty} |p^{-ks}| \\ &= \sum_p |p^{-s}(1 - p^{-s})^{-1}| \leq 2 \sum_p |p^{-s}|, \end{aligned}$$

which we know converges for $\operatorname{Re} s > 1$, completing the first part of the proof.

Now we show that $\exp(G(s, \chi)) = L(s, \chi)$. Using the Taylor series of the logarithm $\log(1 - x) = x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots$ that I put above, we have

$$\exp\left( \sum_{k=1}^{\infty} \frac{z^k}{k} \right) = (1 - z)^{-1}.$$

Take $z = \chi(p)p^{-s}$. Then, the above gives

$$\exp\left(\sum_{k=1}^{\infty} \frac{1}{k}\chi(p)^k p^{-ks}\right) = (1-\chi(p)p^{-s})^{-1}.$$

Both sides converge when $|z| < 1$, so we get

$$\begin{aligned}\exp(G(s,\chi)) &= \exp\left(\sum_p \sum_{k=1}^{\infty} \frac{1}{k}\chi(p)^k p^{-ks}\right) \\ &= \prod_p \exp\left(\sum_k \frac{1}{k}\chi(p)^k p^{-ks}\right) \\ &= \prod_p (1-\chi(p)p^{-ks})^{-1} = L(s,\chi),\end{aligned}$$

as desired. □

Now we reach a key step in our proof of Dirichlet's Theorem, which describes the behavior of $G(s,\chi)$. Again, we can think of $G(s,\chi)$ as the logarithm of $L(s,\chi)$ in some sense.

**Proposition 19.3**

Define $G(s,\chi)$ as above.

1. If $\chi \neq 1$, then $G(s,\chi)$ is bounded near $s = 1$.
2. If $\chi = 1$, then $\lim_{s\to 1^+} \frac{G(s,1)}{\log(s-1)^{-1}} = 1$.

We will only prove (2) for now. We will assume (1) to prove Dirichlet's Theorem, then we will go back to prove (1) afterwards.

*Proof of (2).* We have seen before that we can write

$$G(s,1) = \sum_{p\nmid m} p^{-s} + \sum_{k=2}^{\infty} \frac{1}{k}\chi(p)^k p^{-ks},$$

where the latter sum is bounded near $s = 1$ by Lemma 19.2 above. Thus, taking the limit as $s \to 1^+$, we get

$$\lim_{s\to 1^+} \frac{G(s,1)}{\log(s-1)^{-1}} = \lim_{s\to 1^+} \frac{\sum_{p\nmid m} p^{-s}}{\log(s-1)^{-1}} = 1$$

as desired. □

## 19.1 Proof of Dirichlet's Theorem

Now, we are finally ready to prove Dirichlet's Theorem.

**Theorem 19.4**
If $(a, m) = 1$, then $d(P(a, m)) = 1/\phi(m)$.

*Proof.* Like above, we will write

$$G(s, \chi) = \sum_p \chi(p)p^{-s} + \sum_p \sum_{k=2}^{\infty} \frac{1}{k}\chi(p)^k p^{-ks};$$

denote the latter sum as $R_\chi(s)$, which is a bounded "error" term. Then, as $\chi$ ranges over all Dirichlet characters modulo $m$, we compute

$$\begin{aligned}
\sum_\chi \overline{\chi}(a)G(s, \chi) &= \sum_\chi \sum_p \overline{\chi}(a)\chi(p)p^{-s} + \sum_\chi R_\chi(s)\overline{\chi}(a) \\
&= \sum_p p^{-s} \sum_\chi \overline{\chi}(a)\chi(p) + \sum_\chi R_\chi(s)\overline{\chi}(a) \\
&= \sum_p p^{-s}\phi(m)\delta_{a,p} + R_{\chi,a}(s) \\
&= \phi(m) \sum_{p \equiv a(m)} p^{-s} + R_{\chi,a}(s), \qquad (*)
\end{aligned}$$

where the second to last equality follows from the orthogonality relation given in Corollary 18.9 and $R_{\chi,a}(s) \coloneqq \sum_\chi R_\chi(s)\overline{\chi}(a)$. Taking the "Dirichlet density" expression on the left and right, we see that the right hand side corresponds to the Dirichlet density of $P(a, m)$. More explicitly, we have

$$\begin{aligned}
1 &= \lim_{s \to 1^+} \frac{G(s, 1)}{\log(s-1)^{-1}} = \lim_{s \to 1^+} \frac{\sum_\chi \overline{\chi}(a)G(s, \chi)}{\log(s-1)^{-1}} \\
&= \lim_{s \to 1^+} \frac{\phi(m)\sum_{p \equiv a(m)} p^{-s} + R_{\chi,a}(s)}{\log(s-1)^{-1}} \\
&= \phi(m) \lim_{s \to 1^+} \frac{\sum_{p \equiv a(m)} p^{-s}}{\log(s-1)^{-1}} \\
&= \phi(m) \cdot d(P(a, m)),
\end{aligned}$$

and the theorem follows. □

# 20 11/27 - Proving Proposition 19.3

Last time, we proved Dirichlet's Theorem, modulo the first part of Proposition 19.3, which states that $G(s,\chi)$ is bounded near $s=1$ for nontrivial $\chi \neq 1$. Recall we defined $G(s,\chi)$ such that $\exp(G(s,\chi)) = L(s,\chi)$ for $\operatorname{Re} s > 1$.

## 20.1 Reducing to Analytic Continuation

Proving this boundedness fact for $G(s,\chi)$ is quite intense and is commonly not covered in undergraduate courses, but Kisin is truly built different so here we are. We will try to make this journey as easy as possible. We will prove the following:

**Theorem 20.1**

If $\chi \neq 1$, then the $L$-function

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

has analytic continuation to $\operatorname{Re} s > 0$ and $L(1,\chi) \neq 0$.

We will show that this theorem implies our desired Proposition 19.3. But first, we should define exactly what it means for a function to be analytic. This is a term from complex analysis.

**Definition 20.2** (Analytic Functions)**.** If $\Omega \subseteq \mathbb{C}$ is open, then $f : \Omega \to \mathbb{C}$ is **analytic** if for all $z_0 \in \Omega$, there exists some $D \subseteq \Omega$ open neighborhood containing $z_0$ such that on $D$,

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n$$

and the sum is a convergent power series.

Hidden beneath this definition are many incredible facts, which is kind of the reason why complex analysis is such a beautiful subject. It is useful to think of analytic as the complex analysis notion of differentiable. The magic is that, unlike in real analysis, if a function is differentiable once, it is differentiable infinitely many times. This means we can write $f$ as an infinite power series, which are basically as good as one can get.

Assuming analytic continuation (Theorem 20.1), which we can now interpret as meaning $L(s,\chi)$ can be extended to all of the right half-plane $\operatorname{Re} s > 0$ as an analytic function, we will prove Proposition 19.3.

*Proof of 19.3.* As $L(1,\chi) \neq 0$, there exists a small disc $D \subseteq \mathbb{C}$ centered at 1 such that

$L(s,\chi)|_D$ does not take the value 0. Choose a neighborhood $D'$ of $L(1,\chi)$ such that $L(s,\chi)(D) \subseteq D'$.

Now, choose a branch of the complex-valued logarithm defined on $D'$, and let $G_1(s,\chi) = \log(L(s,\chi))$ for $s \in D$. Then, on $D \cap \{s \mid \operatorname{Re} s > 1\}$, we have $\exp(G(s,\chi)) = L(s,\chi) = \exp(G_1(s,\chi))$. But exp is invariant under addition by $2\pi i$, so on $D \cap \{s \mid \operatorname{Re} s > 1\}$, we have

$$G(s,\chi) - G_1(s,\chi) = 2\pi i n$$

for some $n \in \mathbb{Z}$. But since $L(s,\chi)$ is bounded by $D'$, $G_1(s,\chi)$ is bounded on $D \ni 1$, which implies $G(s,\chi)$ is bounded around $s = 1$, as desired. □

So now we have reduced our task to proving analytic continuation a la Theorem 20.1. This is not much of a reduction in the sense that this is still Really Hard, but we have now gotten to the core of the proof for Dirichlet's Theorem to fall.

## 20.2 Analytic Continuation for Riemann Zeta

Although we are concerned when $\chi \neq 1$, it turns out we have a nice way of analytically continuing, to some extent, $L(s,\chi)$ when $\chi = 1$. Note that this is just the Riemann zeta function $\zeta(s)$.

**Proposition 20.3**

$\zeta(s) - \frac{1}{s-1}$ can be analytically continued to $\{s \in \mathbb{C} \mid \operatorname{Re} s > 0\}$.

Before we prove this, we will prove this lemma, another result from complex analysis.

**Lemma 20.4**

Let $\{a_n\}, \{b_n\} \subseteq \mathbb{C}$ be sequences such that $\sum_{n=1}^{\infty} a_n b_n$ converges. Let $A_n = a_1 + a_2 + \cdots + a_n$. Suppose $A_n b_n \to 0$ as $n \to \infty$. Then,

$$\sum_{n=1}^{\infty} a_n b_n = \sum_{n=1}^{\infty} A_n (b_n - b_{n+1})$$

and the right hand side converges.

If you sit down with this a little bit, this looks to be true: you can cancel a lot of terms on the right to just reduce to $a_n b_n$ terms, which remain on the left. The real content is that the sum is convergent.

*Proof.* Let $S_N = \sum_{n=1}^{N} a_n b_n$. (Also for formality, let $A_0 = 0$.) Then, we can write

$$\begin{aligned} S_N &= \sum_{n=1}^{N} (A_n - A_{n-1}) b_n = \sum_{n=1}^{N} A_n b_n - \sum_{n=1}^{N} A_{n+1} b_n \\ &= \sum_{n=1}^{N} A_n b_n - \sum_{n=1}^{N-1} A_n b_{n+1} = A_N b_N + \sum_{n=1}^{N-1} A_n (b_n - b_{n+1}). \end{aligned}$$

Thus, taking the limit as $N \to \infty$, we get

$$\sum_{n=1}^{\infty} a_n b_n = \lim_{N \to \infty} S_N = \sum_{n=1}^{\infty} A_n (b_n - b_{n+1}).$$

(This is the "you can cancel a lot of terms on the right" I was talking about.) □

Now we prove Proposition 20.3. This is our first time really proving a function can be analytically continued, so it may be useful to lay out the general principle first. A priori, $\zeta(s)$ is defined for $\operatorname{Re} s > 1$. To extend to $\operatorname{Re} s > 0$, we will take a point $z$ with $\operatorname{Re} z > 1$, then find a ball around $z$ which goes beyond $\{\operatorname{Re} s > 1\}$, then show that our function at hand can be analytically defined on this ball.

*Proof of 20.3.* Applying the above lemma with $a_n = 1$ and $b_n = n^{-s}$, we see that as $\sum_{n=1}^{\infty}$ is exactly $\zeta(s)$, it follows that we can write

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

Let $\{x\} = x - \lfloor x \rfloor \in [0, 1)$ denote the fractional part of $x$. Note that we can write

$$n^{-s} - (n+1)^{-s} = s \int_n^{n+1} x^{-s-1}\, dx,$$

so

$$\begin{aligned}
\zeta(s) &= \sum_{n=1}^{\infty} n(n^{-s} - (n+1)^{-s}) \\
&= \sum_{n=1}^{\infty} n \cdot s \int_n^{n+1} x^{-s-1}\,dx \\
&= s\sum_{n=1}^{\infty} \int_n^{n+1} \lfloor x \rfloor x^{-s-1}\,dx \\
&= s\int_1^{\infty} \lfloor x \rfloor x^{-s-1}\,dx \\
&= s\int_1^{\infty} x \cdot x^{-s-1}\,dx - s\int_1^{\infty} \{x\}x^{-s-1}\,dx \\
&= s \cdot \frac{x^{1-s}}{1-s}\bigg|_1^{\infty} - s\int_1^{\infty} \{x\}x^{-s-1}\,dx \\
&= 1 + \frac{1}{s-1} - s\int_1^{\infty} \{x\}x^{-s-1}\,dx.
\end{aligned}$$

We see that the first term has a pole at $s = 1$ (we are dividing by $s - 1$). But 'tis merely a scratch, since it is a simple pole which goes away if we subtract $\frac{1}{s-1}$. (In fact, we are just left with 1.)

We should check that the integral for the second term indeed converges and is analytic for $\operatorname{Re} s > 0$ (i.e., all of our problems lie in the first term). But we know $\{x\} \in [0, 1)$, so $|\{x\}| < 1$, meaning

$$\left|\int_1^{\infty} \{x\}x^{-s-1}\,dx\right| \le \int_1^{\infty} |x^{-s-1}|\,dx = \int_1^{\infty} x^{-1-\operatorname{Re} s}\,dx,$$

which converges for $\operatorname{Re} s > 0$ by just integrating like a high schooler would: for $s \in \mathbb{R}^+$, we have (check this!!)

$$s\int_1^{\infty} x^{-s-1}\,dx = -x^{-s}\big|_1^{\infty} = 1$$

To show it is analytic on the right half-plane, it suffices to show it is analytic at $s = 1$, since that is the only pole of both $\zeta(s)$ and $1 + \frac{1}{s-1}$. We can manipulate

$$\begin{aligned}
\int_1^{\infty} \{x\}x^{-s-1}\,dx &= \int_1^{\infty} \{x\}x^{-2}x^{1-s}\,dx = \int_1^{\infty} \{x\}x^{-2}e^{\log x(1-s)} \\
&= \int_1^{\infty} \{x\}x^{-2} \sum_{n=0}^{\infty} (1-s)^n \frac{(\log x)^n}{n!} \\
&= \sum_{n=1}^{\infty} \left(\int_1^{\infty} \{x\} \cdot \frac{x^{-2}(\log x)^n}{n!}dx\right)(1-s)^n.
\end{aligned}$$

Letting $a_n$ be the integral in the sum above, we see that we have just expressed $\int_1^\infty \{x\}x^{-s-1}\,dx$ as a power series $\sum_{n=1}^\infty a_n(1-s)^n$ around $s=1$, as desired. $\square$

Some additional remarks on analytic functions. If $h_1 \not\equiv 0$ is a function analytic at $s=1$, then we can write $h_1(s) = \sum_{n=0}^\infty a_n(s-1)^n$ as a power series around $s=1$. We denote $\operatorname{ord}_{s=1} h_1 = \min\{i : a_i \neq 0\}$. Let $j = \operatorname{ord}_{s=1} h_1$. Then, in the power series, we can factor out a $(s-1)^j$ and get $h_1(s) = (s-1)^j \cdot g_1(s)$ for some analytic $g_1$ such that $g_1(1) \neq 0$. (By construction of $g_1$, the constant term of $g_1$ as a power series around $s=1$ is nonzero.)

If $h_2$ is another function analytic at $s=1$ with $k = \operatorname{ord}_{s=1} h_2(s)$, then we can write $h_2(s) = (s-1)^k g_2(s)$ similarly. Then,

$$\frac{h_2(s)}{h_1(s)} = \frac{(s-1)^k g_2(s)}{(s-1)^j g_1(s)} = (s-1)^{k-j}\frac{g_2(s)}{g_1(s)}$$

$$\implies \lim_{s\to 1^+} \frac{h_2(s)}{h_1(s)} = \begin{cases} 0 & k > j \\ c \neq 0 & k = j \\ \infty & k < j \end{cases}.$$

# 21 12/01 - Proving Theorem 20.1

Last time, we showed $\zeta(s) - \frac{1}{s-1}$ can be analytically continued to $\operatorname{Re} s > 0$. We now want to prove that $L(s,\chi)$ has analytic continuation (Theorem 20.1). We will first prove the following useful lemma.

**Lemma 21.1**

Let $\chi \neq 1$ be a Dirichlet character modulo $m$. Then, for $N \gg 0$,

$$\left|\sum_{n=0}^{N} \chi(n)\right| \leq \phi(m).$$

*Proof.* We will denote $\chi_0 = 1$ also as the trivial character, as it is less awkward to write $\chi_0(n)$ than $1(n)$. (This is the notation Kisin has maintained throughout the course, anyways.) As $\chi \neq \chi_0$, by the Orthogonality Relations (Corollary 18.9), we have

$$0 = \sum_{n=0}^{m-1} \chi(n)\overline{\chi_0(n)} = \sum_{n=0}^{m-1} \chi(n).$$

Write $N = m \cdot q + r$ for $0 \le r \le m-1$. Then,

$$\sum_{n=0}^{N-1} \chi(n) = q\left(\sum_{n=0}^{m-1} \chi(n)\right) + \sum_{n=0}^{r-1} \chi(n)$$
$$\implies \left|\sum_{n=0}^{N-1} \chi(n)\right| = \left|\sum_{n=0}^{r-1} \chi(n)\right| \le \sum_{n=0}^{m-1} |\chi(n)| = \phi(m),$$

where the second equality follows from $\sum_{n=0}^{m-1} \chi(n) = 0$. □

## 21.1 Proof of Analytic Continuation of $L(s, \chi)$

Now, incredibly, we can prove the first part of Theorem 20.1.

**Proposition 21.2**
If $\chi \neq 1$, then $L(s, \chi)$ has an analytic continuation to $\operatorname{Re} s > 0$.

*Proof.* Let $S(x) = \sum_{0 \le n \le x} \chi(n)$. We know $L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s}$. We will now invoke Lemma 20.4. Letting $a_n = \chi(n)$ and $b_n = n^{-s}$, we have $\sum_{n=1}^{\infty} a_n b_n = L(s, \chi)$ is convergent and $A_n = a_1 + \cdots + a_n = S(n)$. We can check, by the above lemma, that $|A_n b_n| = |S(n)/n^s| \le \phi(m) n^{-s} \to 0$ for $\operatorname{Re} s > 0$. Then, by Lemma 20.4, we have

$$\begin{aligned} L(s, \chi) = \sum_{n=1}^{\infty} \chi(n) n^{-s} &= \sum_{n=1}^{\infty} S(n)(n^{-s} - (n+1)^{-s}) \\ &= \sum_{n=1}^{\infty} S(n)\left(s \int_n^{n+1} x^{-s-1}\, dx\right) \\ &= s \int_1^{\infty} S(x) x^{-s-1}\, dx \qquad (S(x) = S(\lceil x \rceil)) \end{aligned}$$

Again, by Lemma 21.1, which tells us $|S(x)| \le \phi(m)$ (in particular, this is bounded), we have that $\int_1^{\infty} S(x) x^{-s-1}\, dx$ converges absolutely. (This is a generalization of the very last part of the proof for Proposition 20.3, replacing $\{x\}$ with $S(x)$ (or in general, any bounded function).) □

## 21.2 Evaluating $L(1, \chi)$

Now we work towards the second part of Theorem 20.1, which states $L(1, \chi) \neq 0$ for $\chi \neq 1$.

**Proposition 21.3**
Let $F(s) = \prod_{\chi \bmod m} L(s, \chi)$. For $s \in \mathbb{R}$ such that $s > 1$, we have $F(s) > 1$.

*Proof.* Recall $G(s,\chi)$, which satisfies $\exp(G(s,\chi)) = L(s,\chi)$ for $\operatorname{Re} s > 1$, can be written as

$$G(s,\chi) = \sum_p \sum_{k=1}^{\infty} \frac{1}{k}\chi(p^k)p^{-ks}.$$

Through some hard work, we can obtain

$$\begin{aligned}\sum_{\chi \bmod m} G(s,\chi) &= \sum_{\chi \bmod m} \sum_p \sum_{k=1}^{\infty} \frac{1}{k}\chi(p^k)p^{-ks} \\ &= \sum_p \sum_{k=1}^{\infty} \frac{1}{k}p^{-ks} \sum_{\chi \bmod m} \chi(p^k) \\ &= \phi(m) \sum_{\substack{p,k \\ p^k \equiv 1 (m)}} \frac{1}{k}p^{-ks} > 0,\end{aligned}$$

where the last equality follows because $\sum_{\chi \bmod m} \chi(p^k) = 0$ unless $p^k \equiv 1 \pmod m$, in which case the sum is $\phi(m)$. (This can be seen from any of the orthogonality relations given in §18.2.) This implies

$$F(s) = \prod_{\chi \bmod m} L(s,\chi) = \exp\left(\sum_{\chi \bmod m} G(s,\chi)\right) > 1$$

as desired. □

**Proposition 21.4**

$L(1,\chi) = 0$ for at most one nontrivial Dirichlet character $\chi \neq 1$.

We know $\zeta(s) - \frac{1}{s-1}$ has analytic continuation at $s = 1$, so we can write $\zeta(s) = \frac{1}{s-1} + g(s) = (s-1)^{-1}(1 + (s-1)g(s))$ for some $g(s)$ which is analytic at $s = 1$.

To prove this Proposition, we will recall the very end of last lecture regarding the order of $s = 1$ as a zero and writing $h_2(s)/h_1(s)$ in terms of $(s-1)$ for analytic $h_1, h_2$. To recap, if $j_1 = \operatorname{ord}_{s=1} h_1(s)$ and $j_2 = \operatorname{ord}_{s=1} h_2(s)$, then

$$\lim_{s \to 1^+} \frac{h_2(s)}{h_1(s)} = \begin{cases} 0 & j_1 > j_2 \\ c \neq 0 & j_1 = j_2 \\ \infty & j_1 < j_2 \end{cases}.$$

*Proof.* We can write

$$F(s) = \prod_{\chi \bmod m} L(s,\chi) = L(s,1) \prod_{\chi \neq 1} L(s,\chi).$$

We know $L(s,1) = \zeta(s)\prod_{p|m}(1-p^{-s})$; the finite product on the right behaves perfectly well at $s=1$, so $L(s,1)$ just has a pole at $s=1$ of order 1, like $\zeta(s)$. Thus, we write $L(s,1) = 1/h_2(s)$ with $\operatorname{ord}_{s=1} h_2(s) = 1$.

If $\chi \neq 1$ such that $L(1,\chi) = 0$, then by definition $\operatorname{ord}_{s=1} L(s,\chi) \geq 1$. If there were two such $\chi \neq 1$ such that $L(1,\chi) = 0$, then the order of $\prod_{\chi\neq 1} L(s,\chi)$ at $s=1$ would be $\geq 2$. But then this would imply $\operatorname{ord}_{s=1} F(s) \geq -\operatorname{ord}_{s=1} h_2(s) + 2 = 1$, meaning $\lim_{s\to 1^+} F(s) = 0$, which we know from the above proposition is false. The conclusion follows. □

**Corollary 21.5**

If $\chi \neq 1$ such that $\chi(\mathbb{Z}) \not\subseteq \mathbb{R}$, then $L(1,\chi) \neq 0$.

*Proof.* Suppose $L(1,\chi) = 0$, so $\operatorname{ord}_{s=1} L(s,\chi) > 0$; we can write $L(s,\chi) = (s-1)g(s)$ for some $g(s)$ analytic at $s=1$. Note that we can interpret $\chi(\mathbb{Z}) \not\subseteq \mathbb{R}$ as saying $\chi \neq \overline{\chi}$, as $\alpha = \overline{\alpha} \iff \alpha \in \mathbb{R}$. So it makes to look at $\overline{\chi}$. For $s \in \mathbb{R}$, $s > 1$, we have

$$L(s,\overline{\chi}) = \sum_{n=1}^{\infty} \overline{\chi(n)} n^{-s} = \sum_{n=1}^{\infty} \chi(n) n^{-s}$$
$$\implies \text{I didn't write this down in time rip}$$

[get this from somebody and fill in later] □

So now we will demonstrate that $L(1,\chi)$ is indeed what we expect. To recap, we established that $L(s,\chi)$, which a priori was not defined at $s=1$, can be analytically continued to $s=1$. But we don't know the value of $L(1,\chi)$ – we just know it's defined! We now check that it indeed agrees with the sum with $s=1$ plugged in. Recall

$$L(s,\chi) = \sum_{n=1}^{\infty} S(n)(n^{-s} - (n+1)^{-s}),$$

where $S(n) = \sum_{a=1}^{n} \chi(a)$ and $|S(n)| \leq \phi(m)$ by Lemma 21.1.

# 22 12/04 - Last Lecture

And just like that, our semester comes to a close. Not before we end with a bang, though! (By bang, we mean completely finishing the proof of Dirichlet's Theorem by showing $L(1,\chi) \neq 0$ given just $\chi \neq 1$.)

Last time, we showed that this is the case when $\chi(\mathbb{Z}) \not\subseteq \mathbb{R}$. Today, we will assume $\chi(\mathbb{Z}) \subset \{-1,0,1\}$. We will show $L(1,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \neq 0$. Let me state this as an actual result.

**Proposition 22.1**

If $\chi$ is a Dirichlet character modulo $m$ and $\chi(\mathbb{Z}) \subset \{-1, 0, 1\}$, then $L(1, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} \neq 0$.

*Proof.* Let $c_n = \sum_{d|n} \chi(d)$. (We will see in a bit why we are considering this sum.) Suppose $(n, m) = 1$. If $d \mid nm$, write $d = d_1 d_2$ such that $d_1 \mid n$, $d_2 \mid m$. Then,

$$c_{nm} = \sum_{d|nm} \chi(d) = \left( \sum_{d_1|n} \chi(d_1) \right) \left( \sum_{d_2|m} \chi(d_2) \right) = c_n c_m.$$

So it suffices to consider when $n = p^a$ is some prime power. We can explicitly compute, depending on the value of $\chi(p)$,

$$c_{p^a} = 1 + \chi(p) + \chi(p^2) + \cdots + \chi(p^a) = \begin{cases} 1 & p \mid m \\ a+1 & \chi(p) = 1 \\ 0 \text{ or } 1 & \chi(p) = -1 \end{cases};$$

in particular, $c_{p^a} \geq 0$ always and $\geq 1$ if $a$ is even. This shows that $\sum_{n=1}^{\infty} c_n$ is unbounded. Now, let

$$f(t) = \sum_{n=1}^{\infty} \chi(n) \frac{t^n}{1 - t^n}$$

where $t \in (0, 1)$. Note that for $n \geq 1$, we have $1 - t^n \geq 1 - t$, so

$$\left| \chi(n) \frac{t^n}{1 - t^n} \right| \leq \frac{t^n}{1 - t} = \frac{1}{1 - t} \cdot t^n,$$

and since then summing over all $n \geq 1$ gives a sum of geometric series, we see that $f(t)$ converges absolutely for all $t \in (0, 1)$.

Furthermore, for each term in the sum, we can expand

$$\begin{aligned} \chi(d) \frac{t^d}{1 - t^d} &= \chi(d) t^d (1 + t^d + t^{2d} + \cdots) \\ &= \chi(d)(t^d + t^{2d} + t^{3d} + \cdots). \end{aligned}$$

Now, summing over all such terms, we get

$$f(t) = \sum_{n=1}^{\infty} \left( \sum_{d|n} \chi(d) \right) t^n = \sum_{n=1}^{\infty} c_n t^n,$$

so aha, this is why our $c_n$ sums are useful. But we showed that the sum of $c_n$ is unbounded! This means $\lim_{t \to 1^-} f(t) = \infty$. We will rely on this to reach a contradiction, so keep this in mind.

Suppose for the sake of contradiction that $L(1,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n} = 0$. We can cleverly use this as follows:

$$\begin{aligned} -f(t) &= \left(\sum_{n=1}^{\infty} \frac{\chi(n)}{n}\right)(1-t)^{-1} - f(t) \\ &= \sum_{n=1}^{\infty} \chi(n)\left(\frac{1}{n(1-t)} - \frac{t^n}{1-t^n}\right) \\ &=: \sum_{n=1}^{\infty} \chi(n) b_n \end{aligned}$$

where $b_n = b_n(t) = \frac{1}{n(1-t)} - \frac{t^n}{1-t^n}$. We can compute $b_1 = \frac{1}{1-t} = \frac{t}{1-t} = 1$. Furthermore, $\lim_{n\to\infty} b_n(t) = \lim_{n\to\infty}\left(\frac{1}{n(1-t)} - \frac{t^n}{1-t^n}\right) = \lim_{n\to\infty} -\frac{t^n}{1-t^n} = 0$.

We claim that the $b_n$'s form a non-increasing sequence, that is, $b_1 \geq b_2 \geq \cdots$ for each $t \in (0,1)$. We will prove this shortly. Assuming this claim, though, we have

$$-f(t) = \sum_{n=1}^{\infty} \chi(n) b_n = \sum_{n=1}^{\infty} S(n)(b_n - b_{n+1}$$

where like in previous lectures, $S(n) = \sum_{j=1}^{n} \chi(j)$. The last equality follows from Lemma 20.4, which is possible to invoke because $|S(n)| \leq \phi(m)$ (by Lemma 21.1) and $b_n \to 0$, so $S(n)b_n \to 0$. Using the simple bound $|S(n)(b_n - b_{n+1})| \leq \phi(m)|b_n - b_{n+1}| = \phi(m)(b_n - b_{n+1})$, we have

$$\begin{aligned} |f(t)| &= \left|\sum_{n=1}^{\infty} S(n)(b - n - b_{n+1})\right| \\ &\leq \sum_{n=1}^{\infty} |S(n)(b_n - b_{n+1})| \\ &\leq \sum_{n=1}^{\infty} \phi(m)|b_n - b_{n+1}| \\ &= \sum_{n=1}^{\infty} \phi(m)(b_n - b_{n+1}) \\ &= \phi(m)b_1 = \phi(m), \qquad \text{(using claim)} \end{aligned}$$

where the last line follows because we assumed the $b_n$'s form a non-increasing sequence. But this means $f(t)$ is bounded above by a constant for all $t$. This contradicts our conclusion that $\lim_{t\to 1^-} f(t) = \infty$, and the proposition follows.

It remains to prove the claim that $b_1 \geq b_2 \geq \cdots$. Bear with me as we proceed with

some computations:

$$\begin{aligned}(1-t)(b_n - b_{n+1}) &= \frac{1}{n} - \frac{t^n}{1+t+\cdots+t^{n-1}} - \frac{1}{n+1} + \frac{t^{n+1}}{1+t+\cdots+t^n} \\ &= \frac{1}{n(n+1)} + \frac{t^n(1+t+\cdots+t^n) - t^{n+1}(1+t+\cdots+t^{n-1})}{(1+t+\cdots+t^n)(1+t+\cdots+t^{n-1})} \\ &= \frac{1}{n(n+1)} - \frac{t^n}{(1+t+\cdots+t^n)(1+t+\cdots+t^{n-1})}.\end{aligned}$$

Now, we invoke the AM-GM inequality,[13] which states that AM $\geq$ GM. Using this here, we can conclude

$$\begin{aligned}\frac{1+t+t^2+\cdots+t^{n-1}}{n} &\geq \left(t^{\frac{n(n-1)}{2}}\right)^{1/n} = t^{\frac{n-1}{2}} \\ \implies 1+t+t^2+\cdots+t^{n-1} &\geq n \cdot t^{\frac{n-1}{2}} \geq n \cdot t^{n/2} \\ \implies 1+t+\cdots+t^n &\geq (n+1)t^{n/2} \\ \implies (1-t)(b_n - b_{n+1}) &\geq \frac{1}{n(n+1)} - \frac{t^n}{n(n+1)t^n} = 0,\end{aligned}$$

so indeed $b_n - b_{n+1} \geq 0$. Hooray! □

At last, all parts of the proof of Dirichlet's theorem are covered. Free at last, free at last...

## 22.1 So... what is an *L*-function?

Yeah, so what exactly are these things? Clearly they contain a lot of information: the pole of $\zeta(s)$ at $s = 1$ implies the infinitude of primes, and the convergence of $L(s, \chi)$ and the nonvanishing of $L(1, \chi)$ is enough to prove Dirichlet's Theorem. But here is a bird's eye view of $L$-functions, which dives into the most cutting edge of modern number theory.

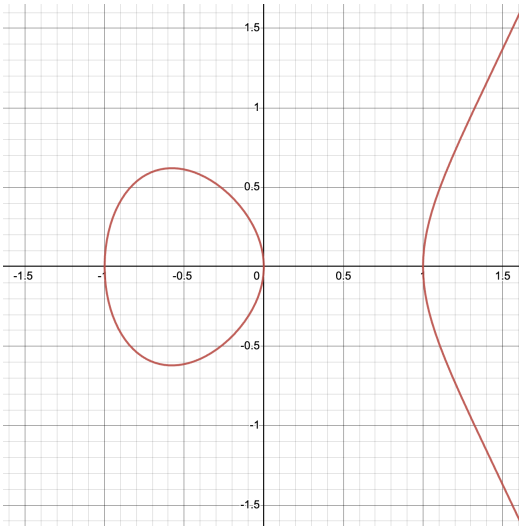
In summary, $L$-functions are *attached to algebro-geometric objects* (objects from algebraic geometry). This is best illustrated by example: consider the curve defined by $y^2 = x(x-1)(x-\lambda)$ for some $\lambda \in \mathbb{Q}$, $\lambda \neq 0, 1$. (For the adults out there, this is an example of an **elliptic curve**.) Here's what Desmos gives me for $\lambda = -1$.

Below is an image of the curve over $\mathbb{R}$. What does it look like over $\mathbb{C}$? This is a bit difficult to imagine, since to visualize any map where our two coordinates are in $\mathbb{C}$, we would need four dimensions, which is too big for a mortal like me. But it turns out, by some really cool theory of elliptic curves of $\mathbb{C}$, that this is a complex torus! It is, in

[13] AM stands for arithmetic mean, which is your normal "sum then divide by number of terms" mean, while GM is the geometric mean, where instead of summing you multiply, and instead of dividing by the number of terms, you take the $n^{\text{th}}$ root where $n$ is the number of terms. For instance, the geometric mean of 2, 3, and 36 is $\sqrt[3]{2 \cdot 3 \cdot 36} = 6$.

more sophisticated terms, a complex Riemann surface of genus 1 (meaning it has one hole). We can realize the below graph as a vertical cross-section of the complex torus, where the rightmost component of the graph can be realized as a circle passing through the point at infinity.

But $\mathbb{R}$ and $\mathbb{C}$ are not special: we can consider this over any field. $\mathbb{Q}$ is of utmost importance because it is related to the integers in an obvious way, e.g. if we solve $a^n + b^n = 1$ over $\mathbb{Q}$, then we have solved $a^n + b^n = c^n$ over $\mathbb{Z}$.




This curve has even more structure than, well, just being a curve. There is a way to add two points on the curve to get another point. This endows the points on the curve with an operation, and so the points, one can show, form a group! In fact, it is an abelian group. Denote $E$ as the elliptic curve, and let $E(K)$ be the points of $E$ over $K$. For instance, $E(\mathbb{C})$ are the points on $E$ where both coordinates are in $\mathbb{C}$, and this forms a torus. Here is an incredible result:

**Theorem 22.2** (Mordwell's Theorem)
$E(\mathbb{Q})$ is a finitely generated group.

Now, how is this related to $L$-functions? Well, I'm going to construct an $L$-function for you. Let $a_p = p + 1 - |E(\mathbb{F}_p)|$ (yes, this seems a bit out of nowhere), and define $L_p(X) = 1 - a_p X + pX^2$ for "good" primes (ignore this for now, it is a technicality). Then, consider the $L$-function

$$L(E, s) \coloneqq \prod_p L_p(p^{-s})^{-1} = \prod_p (1 - a_p p^{-s} + p^{1-2s}.$$

This converges for $\operatorname{Re} s > 3/2$. But even better, this exhibits very nice properties akin to what we proved for our $L$-functions attached to Dirichlet characters:

**Theorem 22.3** (Wiles, Taylor-Wiles, ...)
$L(E, s)$ has analytic continuation to all of $\mathbb{C}$, and $L(E, s) = L(E, 2 - s)$.

Let's relate them more explicitly. Oh wait, we can't actually, because nothing's been proven yet. But at least we can state some really important conjectures. Here's one of the Millenium Problems:

(Birch and Swinnerton-Dyer Conjecture) Let $r$ be the rank of $E(\mathbb{Q})$ as a $\mathbb{Z}$-module (this is well-defined because by Mordell's Theorem, $E(\mathbb{Q})$ is finitely generated). Then, $r = \text{ord}_{s=1} L(E, s)$. In general, for an algebraic variety $X/\mathbb{Q}$,

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

for some polynomials $L_{X,p}$ arising from point counts of $X(\mathbb{F}_p)$.

Here's another one:

$L(X, s)$ has analytic continuation and satisfies a functional equation. The special values of this $L$-function correspond to cycles on $X$, and there is an equivalent of the Riemann Hypothesis for these $L$-functions.

What we did in class was a baby example of this: Dirichlet $L$-functions correspond to 0-dimensional algebraic varieties, in particular the ones defined by the curve $z^m = 1$, which give birth to the number field $\mathbb{Q}(\zeta_m)$. So yeah, geometry is tied with these $L$-functions, which mostly live in the realm of analysis but are completely tied with number theory. All of this is really beautiful, and I would really encourage you all to take a look at some of these things at some point in your academic journey!