PMATH 340

Elementary Number Theory

F. Al-Faisal

Spring 2024

Lectures

To the reader

If you spot any typos/errors or have any feedback that you would like to share, please let me know.

The course. We're going to learn some elementary number theory. *Elementary* in this context doesn't mean *easy*. It's the traditional name for "number theory that doesn't use complex analysis." This distinction exists for historical reasons, which I'll explain at some point.

So, what are we going to cover? A lot of the basics: divisibility, prime numbers, congruences, some famous Diophantine equations, arithmetic functions, continued fractions, and some applications (mostly to cryptography). If we have time, I hope to say a thing or two about elliptic curves at the end.

You should know that number theory has many sub-disciplines: there's algebraic number theory, analytic number theory, combinatorial number theory, computational number theory, Diophantine analysis, arithmetic geometry, and a bunch of others. We will explore a little bit of each during this course.

Problems. The best way to learn number theory is by trying to solve problems. Each lecture will have **Exercises** meant for you to tackle as you make your way through the material. Additionally, there are end-of-lecture problems which have been designed to test your understanding at a deeper level. Some of them (marked with ▶) are especially important, since they will be used in later lectures. Solutions can be found in Appendix A.

Anyway—you should do all of these problems, and more. You can find plenty of problems online and in textbooks. Here are two recommended textbooks:

- Jones and Jones, *Elementary Number Theory*, Springer, 1998.
- Niven, Zuckerman and Montgomery, *An Introduction to the Theory of Numbers* (5th ed.), Wiley, 1991.

You can access the first one digitally for free through the UW Library with your WatIAM credentials. The copyright on the second one has lapsed; a PDF can be found online easily.

Acknowledgements. The material in these notes is fairly standard. I make no claim to originality. The base LaTeX template was created by Michael A. La Croix.

Lecture 1 Introduction

> *"Begin at the beginning," the King said gravely, "and go on till you come to the end: then stop."*
>
> – Lewis Carroll, *Alice in Wonderland*

What is number theory?

We first need to agree on what we mean by "number." The natural starting point is to consider the positive integers

$$\mathbb{Z}_{>0} = \{1, 2, 3, \ldots\}$$[1]

or maybe just the integers altogether

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

This will suffice for a while, but soon we'll find ourselves needing more flexibility. For example, we'll want use the rational numbers $\mathbb{Q}$ and the real numbers $\mathbb{R}$, and also more exotic things like the **Gaussian integers**

$$\mathbb{Z}[i] = \{a + bi \colon a, b \in \mathbb{Z}\}$$

where i satisfies $i^2 = -1$ (a complex number).

Let's fix a choice of number system, like $\mathbb{Z}$. This system should come equipped with basic algebraic operations, like addition and multiplication. These operations generally lead to related natural notions, like divisibility (if we can multiply two things, we can ask whether we can divide two things) and prime numbers (what numbers can't be divided into other numbers?). Number theory can be broadly defined as the study of the structure that arises from operations in number systems. This turns out, as we'll learn throughout this course, to be simultaneously *very interesting* and (potentially) *very difficult*.[2]

To be more concrete, let's look at an example of something number theorists study.

Diophantine equations

If we can add and multiply, we can create polynomials. One of the classical number theoretic problems is that of solving polynomials equations, where we insist that the solutions must belong to our chosen number system. We call these **Diophantine equations**, in honour of the Greek mathematician Diophantus who was the first to seriously study such equations.

For instance, the equation

$$2x = 5$$

which is trivial to solve over $\mathbb{Q}$ or $\mathbb{R}$, becomes slightly more interesting over $\mathbb{Z}$ where it has no solution because 5 is not *divisible* (a number-theoretic concept) by 2. But this is kind of boring. Let's look at more interesting examples.

[1] In these notes, the positive integers will be denoted by $\mathbb{Z}_{>0}$ and the non-negative integers by $\mathbb{Z}_{\geq 0}$. I will not be using the ambiguous $\mathbb{N}$.

[2] It's probably interesting *because* it's difficult.

Example 1.1 **(Some famous Diophantine equations)**

- **Pythagorean triples.** What are the positive integer solutions to the Pythagorean equation
$$x^2 + y^2 = z^2?$$
Solutions $(x, y, z) = (a, b, c)$ with $a, b, c \in \mathbb{Z}_{>0}$ are called **Pythagorean triples**; they represent integer side lengths of a right-angled triangle.

  You might be familiar with the triples $(3, 4, 5)$ and $(5, 12, 13)$. Are there any others? Are there infinitely many?

  The answer to both questions is *yes* for a silly reason: if (a, b, c) is a Pythagorean triple, then so is (na, nb, nc) for any $n \in \mathbb{Z}_{>0}$ (check this!). So, for instance, $(30, 40, 50)$ is a Pythagorean triple, as is $(10, 24, 26)$.

  Are there any other non-silly triples? One way to quantify this is to ask for triples (a, b, c) that don't have a common factor; we call such triples **primitive**. We'll come back and answer this question later in the course.

- **Fermat's Last Theorem.** Generalizing the above, we can ask (as Fermat did) for the positive integer solutions to the equation
$$x^n + y^n = z^n$$
where $n > 2$. Fermat famously claimed that there are no such solutions. But why not? This question went unanswered for approximately 300 years (not for lack of trying!) until a proof was finally provided by Andrew Wiles and Richard Taylor in 1995. They used modern ideas that are so far beyond anything accessible to Fermat. The title of one of the two papers containing the proof is

  Ring-theoretic properties of certain Hecke algebras.

  You're probably wondering, *what the Hecke does this mean?* (No offence to E. Hecke, one of my personal mathematical heroes.)

- **Pell's equation.**[3] Fix a positive integer $n \in \mathbb{Z}_{>0}$ that isn't a perfect square. What are the integer solutions to
$$x^2 - ny^2 = 1?$$
We'll be exploring this equation in detail later on. It's an interesting one because its solution set exhibits some peculiar structure. For instance, starting from one solution, we can generate others. To see what I mean by this, take $n = 2$, so that the equation becomes
$$x^2 - 2y^2 = 1.$$
An easy-to-find solution is given by $(x, y) = (3, 2)$. You can check that if (a, b) is an integer solution, then so is $(3a + 4b, 2a + 3b)$. Starting with $(3, 2)$ and applying this rule repeatedly, we can generate the following sequence of (non-obvious) solutions
$$(3, 2),\ (17, 12),\ (99, 70),\ (577, 408),\ (3363, 2378),\ \ldots,\ (22619537, 15994428),\ \ldots.$$
Here's another curious thing. If (x, y) is a solution to Pell's equation, then x/y can be used to approximate $\sqrt{n}$ if y is large. (To see why, divide both sides of the equation

by y^2. Then the right-side becomes $1/y^2 \approx 0$ for large y.) For example, our work above yields

$$\sqrt{2} \approx \frac{22619537}{15994428}.$$

Go plug this into your favorite calculator and be amazed!

Exercise 1.2 (a) Show that if (a, b, c) is a Pythagorean triple, then so is (na, nb, nc) for any $n \in \mathbb{Z}_{>0}$.

(b) Confirm that if $(x, y) = (a, b)$ is a solution to the Pell equation $x^2 - 2y^2 = 1$, then so is $(x, y) = (3a + 4b, 2a + 3b)$.

Given a Diophantine equation, there are three natural questions that we can ask:

1. Does it have any solutions?
2. If it has solutions, how many are there? (Finitely many? Infinitely many?)
3. Can we find all of the solutions?

But even before that, there is a much more pressing question:

0. *Why do we care about any of this?*

This is a fair question; the *why?* is arguably more important than the *how?*. Let me give two answers. The first is that, occasionally, problems that you might care about can be turned into Diophantine equations. If this happens, then surely you will want to know how to solve the resulting Diophantine equation, and at which point you'll be thankful that number theorists have done the hard work for you.

The second answer is more philosophical. Diophantine equations are extremely simple problems to formulate: all that's involved is addition, multiplication and an equals sign. The fact that we can't solve a given Diophantine equation indicates that there is something that we (humanity as a whole) are missing. What is it? And why is it so difficult to figure out? It's this—the fact that there is *something* about such basic mathematics that we don't quite understand—that make the subject worth investigating.

That said, it's really easy to create a lot of hard-to-solve Diophantine equations. The ones that get number theorists excited are those that reveal some hidden mathematical structure. The three equations given in Example 1.1 do this, as we'll come to learn.

[3] Should probably be called the Brahmagupta–Fermat equation, since Pell had little to do with it.

Solving Diophantine equations

Let's return to the three questions above, but let's be a bit more greedy and ask:

> Is there a general algorithm that can decide, in finite time, whether any given Diophantine equation has a solution?

This question was posed by Hilbert at the International Congress of Mathematicians in 1900. It was part of a famous list of 23 problems that he set for mathematicians in the 20th century. The above was the 10th problem on the list.

I have good news and bad news. The good news: The problem has been solved! The bad news: The answer is **NO!** In 1970, Matiyasevich, building on work of Davis, Putnam and Robinson, proved that there can be no *general algorithm* that is capable of deciding whether a given Diophantine equation has a solution (over $\mathbb{Z}$, at least). The proof uses mathematical logic and computability theory; you can read about it here.

Now, while there is no "super algorithm" that you can apply to any old Diophantine equation, there *can be* (in fact, there *are*) algorithms that work on certain classes of Diophantine equations. The point, however, is that you should expect to invoke a certain amount of ingenuity if you want to solve a random Diophantine equation.

Here are some Diophantine equations that we know (as we'll learn in this course) how to solve over $\mathbb{Z}$:

- $x^2 + x = 1$
- $3x + 5y = 7$
- $x^2 + y^2 = n$
- $x^4 + y^4 = z^2$

The first two are much easier to solve than the last two. The first one involves only a single variable (it's a **univariate** Diophantine equation) and the second one involves only variables of degree 1 (it's a **linear** Diophantine equation). We have good methods for dealing with these types of equations. Non-linear multivariable equations are generally much more difficult to analyze.

Exercise 1.3 Try to see if you can solve any of the above equations. (Don't be upset if you can't. Instead, get excited that you'll soon learn how!)

Univariate Diophantine equations

We'll now explore how to solve any Diophantine equation of the form

$$f(x) = 0$$

where $f(x)$ is a polynomial with integer coefficients. Note that any univariate Diophantine equation is of this form, since we can move everything to the left-side; e.g.

$$x^2 + x = 1 \iff x^2 + x - 1 = 0.$$

We're going to make use of the following extremely fundamental number theoretic concept.

Definition 1.4

Divides, Divisible, Divisor, Factor, $a \mid b$, $a \nmid b$

Let $a, b \in \mathbb{Z}$. We say that a **divides** b, and write $a \mid b$, if there is an integer $c \in \mathbb{Z}$ such that $b = ac$. In this case we also say that b is **divisible** by a, and that a is a **divisor** (or **factor**) of b.

If a doesn't divide b, then we denote this by writing $a \nmid b$.

For example, $2 \mid 10$, $-4 \mid 8$ and $1 \mid n$ for all $n \in \mathbb{Z}$, but $3 \nmid 17$. The divisors of 6 are ± 1, ± 2 and ± 3.

Here is our main result.

Proposition 1.5

If $x = s$ is an integer solution to the equation

$$a_n x^n + \cdots + a_1 x + a_0 = 0,$$

where $a_i \in \mathbb{Z}$ and $n \geq 1$, then $s \mid a_0$.

Proof: Plugging the solution $x = s$ into the equation and rearranging, we arrive at

$$a_0 = s(-a_n s^{n-1} - \cdots - a_1).$$

If s is an integer then so is $-a_n s^{n-1} - \cdots - a_1$. Thus, $s \mid a_0$ by definition. ■

This proposition tells us that if we want to solve a univariate Diophantine equation

$$a_n x^n + \cdots + a_1 x + a_0 = 0$$

then all (!) we have to do is find all the divisors of the constant term a_0 and plug them into the equation one by one. Those that satisfy the equation are solutions—and these are all the solutions.

Example 1.6

To solve $x^2 + x = 1$, or equivalently $x^2 + x - 1 = 0$, we just have to test if the divisors of -1 satisfy the equation. The divisors of -1 are ± 1, and we have

$$1^2 + 1 - 1 = 1 \neq 0 \quad \text{and} \quad (-1)^2 + (-1) - 1 = -1 \neq 0.$$

So neither satisfies the equation. Hence the equation has no integer solutions.

On the other hand, consider the equation

$$x^5 + 2x^4 + x + 2 = 0.$$

Plugging in the divisors of 2, which are ± 1 and ± 2, we find that

- $1^5 + 2(1^4) + 1 + 2 = 6 \neq 0$.
- $(-1)^5 + 2(-1)^4 + (-1) + 2 = 2 \neq 0$.
- $2^5 + 2(2^4) + 2 + 2 = 68 \neq 0$.
- $(-2)^5 + 2(-2)^4 + (-2) + 2 = 0$.

So the only integer solution is $x = -2$.

While we certainly have, in theory, a finite-time algorithm that decides the solvability of any univariate Diophantine equation $f(x) = 0$, in practice this method can be extremely impractical. For one, finding all the divisors of a given integer is a **hard** problem in general (in a certain sense that we will investigate later), especially if the integer in question is large. Second, even if we are able to factor a_0, it may have many divisors, and we'll have to check them one by one. This can take a very long time.

Example 1.7 Consider the equation

$$x^2 + x - 2^{100} = 0.$$

Our algorithm requires us to check each of the 202 divisors of 2^{100}. We can arbitrarily increase the exponent on the constant term, say from 100 to $10^{10^{10}}$, to make this computationally infeasible. So our algorithm is not of much help here, really.

There is however an easy way to see that this equation—even if we change the constant term from 2^{100} to 2^n—has no solution in $\mathbb{Z}$ if $n > 1$. We begin by re-writing the equation as

$$x(x+1) = 2^n.$$

If this has a solution in $\mathbb{Z}$, then the left-side would be a product of two consecutive integers, so one of them must be odd, and it must divide 2^n. The only odd divisors of 2^n are ± 1, so either $x = \pm 1$ or $x + 1 = \pm 1$. If $x = 1$ then $x + 1 = 2$ and so $x(x+1) = 2 \neq 2^n$ if $n > 1$. If $x = -1$ then $x + 1 = 0$ so $x(x+1) = 0 \neq 2^n$. Similarly, we can show that $x + 1 = \pm 1$ leads to no solutions either.

Example 1.8 Suppose our Diophantine equation $f(x) = 0$ has constant term given by the 250-digit integer

$$\begin{aligned} a_0 = {} & 21403246502407449612644230728393335630086147151447550177977549208814180 \\ & 23447140136643345519095804679610992851872470914587687396261921557363047 \\ & 45477052080511905649310668769159001975940569345745223058932597669747168 \\ & 17380693648946998715784949759374979 37. \end{aligned}$$

Now we have the joyous task of attempting to factor this integer, which actually has name—it's RSA-250. Factoring this integer, and other "RSA numbers," was set as a challenge by RSA Laboratories in 1991.

The task of factoring RSA-250 was only accomplished in February 2020—almost 30 years after the challenge had been set. There are several other RSA numbers that remain unfactored to this day. The difficulty of factoring large integers is a principle that underlies the security of many widely used cryptosystems.

So our solution-finding algorithm is a bit naive and can run into various difficulties in practice. It is possible to improve it and make it significantly more efficient, but we won't pursue these ideas here since they'll take us a bit too far afield.[4]

[4] If you're curious, you should look up Sturm's theorem, which can be used to perform a root-search in the real interval $[-|a_0|, |a_0|]$ where the integer roots are guaranteed to lie.

Lecture 1 Problems

1.1. Let $n > 2$ be an integer. Find all integer solutions to the equation $x^n + x^{n-1} - 2x - 2 = 0$.

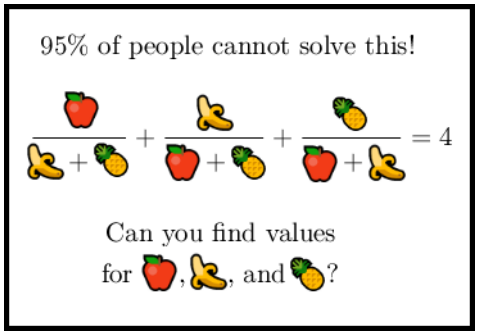
1.2. Find all integer solution (x, y) to the equation $2x^3 + xy - 7 = 0$.

1.3. Show that if equation $ax^2 + bx + c = 0$, where $a, b, c \in \mathbb{Z}$ and $a \neq 0$, has a solution in $\mathbb{Z}$ then $b^2 - 4ac$ must be a perfect square. Does the converse hold?

1.4. Formulate (but do not solve) a two-variable Diophantine equation that models the following problem.

Captain Hook has n cannonballs. He can lay them flat on a table to create a perfect square. He can also stack them vertically to create a perfect square pyramid (i.e. a pyramid whose layers are squares—for example, with 14 cannonballs, he can make a pyramid whose layers from bottom to top consist of 9, 4 and 1 cannonballs arranged into perfect 3×3, 2×2 and 1×1 squares). How many cannonballs does he have?

1.5. Consider the following meme that made the rounds on the internet a few years ago:



Using a, b and c, to represent apples, bananas and pineapples, the problem here is to solve the equation

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} = 4 \qquad (*)$$

presumably over $\mathbb{Z}$ or perhaps $\mathbb{Z}_{>0}$. Our definition of Diophantine equations restricts us to polynomial equations, and the above isn't one.

(a) Show that if we set

$$a = \frac{56 - x + y}{56 - 14x}, \quad b = \frac{56 - x - y}{56 - 14x}, \quad \text{and} \quad c = \frac{-28 - 6x}{28 - 7x}$$

then equation $(*)$ can be transformed into the Diophantine equation

$$y^2 = x^3 + 109x^2 + 224x.$$

(b) Show that if $y = 28$, then the equation in part (a) has an integer solution. Hence determine an integer solution (a, b, c) to equation $(*)$.

Note: The solution you're led to find in part (b) will not consist of positive integers. It's a bit more tricky to find positive solutions to $(*)$—but they exist! The smallest one is:

$$\begin{aligned} a = {} & 154476802108746166441951315019919837485664325669565431 \\ & 700026634898253202035277999 \\ b = {} & 368751317941299998271978115652254748254929799689719709 \\ & 962831374716372246340555 79 \\ c = {} & 437361267792869725786125260237139015281653755816161361 \\ & 86214379933784234677720 36. \end{aligned}$$

Lecture 2 The Remainder Theorem

Let's take a closer look at the notion of divisibility of integers (refer to Definition 1.4).

Proposition 2.1 Let $a, b, c \in \mathbb{Z}$.

(a) If $a \mid b$ and $b \mid c$ then $a \mid c$. (We say that divisibility is **transitive**.)

(b) If $a \mid b$ and $a \mid c$ then $a \mid xb + yc$ for all $x, y \in \mathbb{Z}$.

(c) If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.

Proof: For part (a), observe that $b = ka$ for some $k \in \mathbb{Z}$ since $a \mid b$. Likewise, $c = lb$ for some $l \in \mathbb{Z}$. Therefore, $c = lb = l(ka) = (lk)a$. So $a \mid c$. The proof of (b) is similar and left as an exercise.

For part (c), write $b = ka$ with $k \in \mathbb{Z}$. Then $|b| = |k||a|$. Note that $k \neq 0$ since otherwise we'd get $b = ka = 0$, contrary to the assumption on b. So, since k is a non-zero integer, $|k| \geq 1$, and therefore $|b| = |k||a| \geq |a|$. ■

Exercise 2.2 Prove part (b) of Proposition 2.1. Prove that its converse is true, too.

We usually express Proposition 2.1(b) by saying that if a divides b and c, then a divides every (integer) linear combination of b and c. For example, since $4 \mid 12$ and $4 \mid 40$, we have

$$4 \mid (-6) \cdot 12 + 2 \cdot 40, \text{ that is, } 4 \mid 8.$$

Our next result is extremely fundamental. We'll be making use of it repeatedly.

Theorem 2.3 (The Remainder Theorem)

Let $a, b \in \mathbb{Z}$ with $b > 0$. Then there exist *unique* integers $q, r \in \mathbb{Z}$ such that

$$a = bq + r \quad \text{and} \quad 0 \leq r < b.$$

The integers q and r are called the **quotient** and **remainder**, resp., of a divided by b.

Notice that the remainder r will be 0 if and only if $b \mid a$, in which case the quotient will simply be $q = a/b$.

The proof of the remainder theorem will make use of the *floor* of a real number.

Definition 2.4 **Floor**

Let $x \in \mathbb{R}$. The **floor** of x, denoted by $\lfloor x \rfloor$, is the greatest integer less than or equal to x.

For example,

$$\lfloor 3.14 \rfloor = 3, \quad \lfloor 7 \rfloor = 7, \quad \text{and} \quad \lfloor -1.2 \rfloor = -2.$$

We always have

$$x - 1 < \lfloor x \rfloor \leq x \text{ for all } x \in \mathbb{R}. \tag{$*$}$$

Exercise 2.5 Prove (∗). In particular, explain why the inequality $x - 1 < \lfloor x \rfloor$ is strict.

With this in hand, we're ready to give:

Proof of Theorem 2.3 (The Remainder Theorem):

Existence: Let $q = \lfloor a/b \rfloor$ and then set $r = a - bq$. Note that $q, r \in \mathbb{Z}$ and $a = bq + r$, so all that's left is to prove that $0 \leq r < b$. To this end, we have from (∗) (with $x = a/b$)

$$\frac{a}{b} - 1 < q \leq \frac{a}{b}.$$

Multiply through by $-b$ to get

$$b - a > -bq \geq -a.$$

Now add a to all sides to get the desired result. This completes the proof of the existence of q and r.

Uniqueness: Suppose that we also have $a = q'b + r'$ for some $q', r' \in \mathbb{Z}$ with $0 \leq r' < b$. Then

$$qb + r = a = q'b + r' \implies (q - q')b = r' - r.$$

This shows that b divides $r' - r$. However, $r' - r$ is between $-b$ and b, that is, $|r' - r| < b$. So, by Proposition 2.1(c), we must have that $r' - r = 0$. Hence $r = r'$ and then since $qb + r = q'b + r'$ and $b \neq 0$, we also get $q = q'$. ■

The above proof actually tells us how to find q and r: take $q = \lfloor a/b \rfloor$ and $r = a - bq$.

Example 2.6 If $a = 100$ and $b = 7$, then $q = \lfloor 100/7 \rfloor = 14$ and $r = 100 - 7 \cdot 14 = 2$. Indeed, we have

$$100 = 7 \cdot 14 + 2 \quad \text{and} \quad 0 \leq 2 < 7.$$

So the quotient and remainder of 100 divided by 7 are $q = 14$ and $r = 2$, respectively.

Note that we can write 100 in the form $7q + r$ in many ways, e.g. $100 = 7 \cdot 2 + 86$. However, it's only when r satisfies the inequalities $0 \leq r < 7$ that we call it the remainder of 100 divided by 7.

Exercise 2.7 Find the quotient and remainder of -75 divided by 6.

Here are some examples of the remainder theorem in action.

Example 2.8 We can classify integers into their remainders after division by a given integer b.

(a) Taking $b = 2$, we see that every $a \in \mathbb{Z}$ can be written in the form $a = 2q$ or $a = 2q + 1$ for some $q \in \mathbb{Z}$. We call integers of the form $2q$ **even** and those of the form $2q+1$ **odd**.

(b) Taking $b = 3$, we see that every $a \in \mathbb{Z}$ can be written in the form $a = 3q$, $a = 3q + 1$ or $a = 3q + 2$ for some $q \in \mathbb{Z}$.

(c) Taking $b = 10$, we see that every $a \in \mathbb{Z}$ can be written in the form $a = 10q$, $a = 10q+1$, ..., or $a = 10q+9$. In this case we recognize the remainders as being the allowable *ones digits* in the decimal representation of a. For instance,

$$1234 = 10 \cdot 123 + 4.$$

Exercise 2.9 Show that every $a \in \mathbb{Z}$ can be written in the form $a = 3k - 1$, $a = 3k$ or $a = 3k + 1$ for some $k \in \mathbb{Z}$.

Example 2.10 Prove that an odd integer leaves a remainder of 1, 3 or 5 after division by 6.

Solution: The possible remainders are $0, 1, 2, \ldots, 6$. The remainders 0, 2 and 4 give us integers of the form $6k = 2(3k)$, $6k + 2 = 2(3k + 1)$ and $6k + 4 = 2(3k + 2)$, which are even, so they cannot occur. Thus, the only possible remainders for an odd integer are $1, 3$ and 5.

Example 2.11 Prove that if a and b are odd then $a^2 + b^2$ is even but not divisible by 4.

Solution: We can write $a = 2k + 1$ and $b = 2l + 1$ for some $k, l \in \mathbb{Z}$. Then

$$a^2 + b^2 = (4k^2 + 4k + 1) + (4l^2 + 4l + 1) = 4(k^2 + k + l^2 + l) + 2.$$

This shows that $a^2 + b^2$ leaves a remainder of 2 after division by 4, so $4 \nmid a^2 + b^2$. It also follows that $a^2 + b^2$ is even since it's of the form $4m + 2 = 2(2m + 1)$.

Exercise 2.12 Prove that if a and b are arbitrary integers then the remainder of $a^2 + b^2$ after division by 4 is either 0, 1 or 2.

Let's close the lecture by showing how the above considerations can be applied to Diophantine equation

$$x^2 + y^2 = n$$

that was mentioned in Lecture 1. Thanks to Exercise 2.12, we can now assert that this equation doesn't have integer solution if n leaves a remainder of 3 after division by 4.

So, for example, the equations

$$x^2 + y^2 = 3, \quad x^2 + y^2 = 7 \quad \text{and} \quad x^2 + y^2 = 10003$$

do not have any integer solutions.

Note that we could have attempted to solve these equations with a brute-force approach: since $x^2 \leq x^2 + y^2$ we have that $|x| = \sqrt{x^2} \leq \sqrt{x^2 + y^2}$, and so we would need $|x| \leq \sqrt{3}$,

$|x| \leq \sqrt{7}$ and $|x| \leq \sqrt{10003}$ for each equation, respectively. There are not many integers that satisfy the first two inequalities, and we can simply go through them by hand. The last one is a bit more tedious, but you can do it if you were sufficiently motivated (e.g. if I offered a 10% bonus to your final grade).[5] But you should appreciate that our approach via remainders doesn't require this kind of lengthy computation, and it can prevail where such computations will surely fail—for instance, with the equation

$$x^2 + y^2 = 4^{10^{10}} + 3.$$

Exercise 2.13

Warning: Our result concerning $x^2 + y^2 = n$ is not an "if and only if."

Show that $x^2+y^2 = 6$ does not have any integer solutions, even though 6 leaves a remainder of 2 after division by 4.

The complete story of $x^2 + y^2 = n$ will have to wait for another day.

Lecture 2 Problems

2.1. Prove:

(a) $1 \mid a$ for all $a \in \mathbb{Z}$.
(b) $a \mid 0$ for all $a \in \mathbb{Z}$.
(c) $0 \mid a$ if and only if $a = 0$.
(d) If $a \mid b$ and $b \mid a$ then $a = \pm b$.
(e) If $a \mid b$ then $a^n \mid b^n$ for all $n \in \mathbb{Z}_{>0}$.

2.2. Prove or disprove:

(a) If $a \mid b$ and $c \mid d$ then $ac \mid bd$.
(b) If $a \mid b$ and $c \mid d$ then $a + c \mid b + d$.
(c) If $a \mid bc$ then either $a \mid b$ or $a \mid c$.
(d) If $a \mid b^2$ then $a \mid b$.

▶ 2.3. Prove that $3 \mid a^3 - a$ for all $a \in \mathbb{Z}$.

2.4. In our formulation of Theorem 2.3 (The Remainder Theorem), we assumed that $b > 0$. Show that the theorem holds for $b < 0$ too provided we use the inequalities $0 \leq r < |b|$ on the remainder. [**Hint:** Apply the theorem to $-b$.]

▶ 2.5. This problem sketches another proof of the existence part of Theorem 2.3 (The Remainder Theorem) using the **Well-Ordering Principle**:

Every non-empty subset of $\mathbb{Z}_{\geq 0}$ has a smallest element.

This principle is not something that requires proof—it is one of the defining features of $\mathbb{Z}_{\geq 0}$ that is equivalent to the principle of mathematical induction.

(a) Let $S = \{a - nb : n \in \mathbb{Z}\} \cap \mathbb{Z}_{\geq 0}$. Prove that S is nonempty.
(b) By the Well-Ordering Principle, S has a smallest element—call it r. Prove that $r < b$. [**Hint:** Argue by contradiction.]
(c) Explain how The Remainder Theorem follows.

[5] Just to be clear: I am not offering such a bonus.

Lecture 3 The Euclidean Algorithm

One of the most basic facts about positive integers is that they can be factored uniquely into products of primes. The existence of such a factorization is easy to prove. Proving uniqueness, however, is quite tricky. So we're going to postpone the discussion of primes and factorizations until we have the necessary tools. One of these tools—the gcd—will be examined in this lecture.

Definition 3.1

Common Divisor, Greatest Common Divisor, gcd

Let $a, b \in \mathbb{Z}$. A **common divisor** of a and b is an integer d such that $d \mid a$ and $d \mid b$.

If a and b are not both are zero, their **greatest common divisor** is the largest integer that is a common divisor of a and b. We denote it by $\gcd(a, b)$.

We define $\gcd(0, 0)$ to be 0.

For example, $\gcd(8, 20) = 4$ and $\gcd(-10, 15) = 5$. Note that $\gcd(a, b)$ is always non-negative, since if $d < 0$ is a common divisor of a and b, then so is $-d > 0$ and $-d$ is larger than d.

Exercise 3.2 Show that $\gcd(a, 0) = \gcd(0, a) = |a|$ for all $a \in \mathbb{Z}$.

Let's now assume that a and b are both non-zero. How do we actually compute $\gcd(a, b)$? Since $\gcd(a, b) = \gcd(-a, b) = \gcd(a, -b)$ (prove it!), let's also assume that a and b are positive.

The naive way to compute $\gcd(a, b)$ would be to run through the positive divisors of a one at a time and check if they divide b; the largest one that does is $\gcd(a, b)$. This clearly is not very efficient, let alone viable if we're dealing with huge integers. Fortunately, there is a much better method. Let me illustrate with an example.

Suppose we want to compute $\gcd(693, 105)$. We start by applying the remainder theorem:

$$693 = 6 \cdot 105 + 63.$$

We've just written 693 as a linear combination of 105 and 63. So, if d divides 105 and 63, then it will divide 693. Conversely, if d divides both 693 and 105, it will divide $63 = 693 - 6 \cdot 105$. Thus, the common divisors of 693 and 105 coincide with the common divisors of 105 and 63. So

$$\gcd(693, 105) = \gcd(105, 63).$$

So now we have to compute $\gcd(105, 63)$. We repeat the same process above, starting with the remainder theorem:

$$105 = 1 \cdot 63 + 42.$$

By the same reasoning, $\gcd(105, 63) = \gcd(63, 42)$. Repeat this process a couple more times:

$$63 = 1 \cdot 42 + 21$$
$$42 = 2 \cdot 21 + 0.$$

We now have

$$\gcd(693, 105) = \gcd(105, 63) = \gcd(63, 42) = \gcd(42, 21) = \gcd(21, 0).$$

But this last gcd is easy to compute! By Exercise 3.2, it's just 21. So, $\gcd(693, 105) = 21$.

The process we just went through is called the *Euclidean algorithm.* It rests entirely upon the following lemma, which was used repeatedly above.

Lemma 3.3 **(gcd Reduction Lemma)**

Let $a, b \in \mathbb{Z}$ be non-zero integers. Then, for all $q \in \mathbb{Z}$, we have

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

Proof: The key idea was explained above. You should write up the details. ■

Exercise 3.4 Prove Lemma 3.3.

ALGORITHM (The Euclidean Algorithm)

Let $a, b \in \mathbb{Z}_{>0}$ and assume that $a > b$. To compute $\gcd(a, b)$:

- **Step 1:** Repeatedly apply the remainder theorem:

$$\begin{aligned} a &= q_1 b + r_1 && (0 \le r_1 < b) \\ b &= q_2 r_1 + r_2 && (0 \le r_2 < r_1) \\ r_1 &= q_3 r_2 + r_3 && (0 \le r_3 < r_2) \\ r_2 &= q_4 r_3 + r_4 && (0 \le r_3 < r_2) \\ &\vdots && \vdots \quad . \end{aligned}$$

  In the first iteration, the remainder theorem is applied to a and b giving a remainder of r_1. In the second iteration, the remainder theorem is applied to b and r_1 giving a remainder of r_2. In the ith iteration for $i > 2$, the remainder theorem is applied to r_{i-2} and r_{i-1} giving a remainder of r_i.

- **Step 2:** Stop once you reach a zero remainder.

- **Step 3:** Return $\gcd(a, b) = r_{n-1}$, where $r_n = 0$ is the zero remainder reached in Step 2. Here, we take $r_{-1} = b$. (Since if the first iteration produces a remainder of zero, then $b \mid a$ so $\gcd(a, b) = b$.)

We must make two comments. First, this algorithm is guaranteed to terminate (i.e. Step 2 will always occur) because the iterations of the remainder theorem in Step 1 yield a strictly decreasing sequence of non-negative integers

$$r_1 > r_2 > \cdots > r_i > \cdots \ge 0$$

and this sequence cannot continue decreasing indefinitely. Second, Step 3 correctly gives us $\gcd(a, b)$ thanks to Lemma 3.3:

$$\gcd(a, b) = \gcd(b, r_1) = \gcd(r_1, r_2) = \cdots = \gcd(r_{n-1}, r_n) = r_{n-1},$$

where the last equality holds because $r_n = 0$. So the Euclidean algorithm works!

Example 3.5 Determine $\gcd(124, -608)$.

Solution: The Euclidean algorithm requires positive integers $a > b$. No problem:

$$\gcd(124, -608) = \gcd(124, 608) = \gcd(608, 124).$$

So let's take $a = 608$ and $b = 124$, and let's run the algorithm:

$$\begin{aligned} 608 &= 4 \cdot 124 + 112 \\ 124 &= 1 \cdot 112 + 12 \\ 112 &= 9 \cdot 12 + 4 \\ 12 &= 3 \cdot 4 + 0. \end{aligned}$$

We've reached a zero remainder! So what we want is the last non-zero remainder, that is,

$$\gcd(124, -608) = 4.$$

Exercise 3.6 Determine $\gcd(1234, 5678)$.

We can run the Euclidean algorithm in reverse to obtain an interesting result. For instance, in Example 3.5, we have

$$\begin{aligned} \gcd(124, -608) &= 4 \\ &= 112 - 9 \cdot 12 \\ &= 112 - 9(124 - 112) \\ &= 10 \cdot 112 - 9 \cdot 124 \\ &= 10(608 - 4 \cdot 124) - 9 \cdot 124 \\ &= (-10) \cdot (-608) + (-49) \cdot 124. \end{aligned}$$

What we've just managed to do is write $\gcd(124, -608)$ as an integer linear combination of 124 and -608. This works in general.

Proposition 3.7 **(Bézout's Lemma)**

Let $a, b \in \mathbb{Z}$. There exist $x, y \in \mathbb{Z}$ such that

$$\gcd(a, b) = xa + yb.$$

We can give a proof that mimics what we did with $\gcd(124, -608)$ above—namely, by reversing the output of the Euclidean algorithm—but the details will be messy because of all the indices. So I'm going to present a different proof. The drawback is that it's not constructive: it tells us x and y exist, but not how to find them.

Proof of Proposition 3.7 (Bézout's Lemma): We may assume without loss of generality that $a \neq 0$. Indeed, if $a = 0$, then $\gcd(a, b) = \gcd(0, b) = |b| = 0a + (\pm 1)b$, and so the result follows. Likewise, we may assume that $a > 0$ since if $a < 0$, then as $\gcd(a, b) = \gcd(-a, b)$ we can work with $-a > 0$ instead.

Let $S = \{xa + yb \colon x, y \in \mathbb{Z}\}$. Then S contains positive integers since $a = 1a + 0b$ is in S. By the Well-Ordering Principle, there must be a *smallest* positive integer in S; call it d. We can write $d = x_0 a + y_0 b$ for some $x_0, y_0 \in \mathbb{Z}$. We're done if we can show that $d = \gcd(a, b)$.

To start, let's show that d is a common divisor of a and b. By applying the remainder theorem to a and d, we can write

$$a = qd + r, \text{ where } 0 \leq r < d.$$

Note that $r = a - qd = (1 - qx_0)a + (-qy_0)b$ must belong to S. So if $r > 0$ then we've found a positive integer in S that's smaller than d—contradicting the minimality of d! Thus $r = 0$ and therefore $d \mid a$. Similarly, $d \mid b$.

But now if d' is any common divisor of a and b, then it will also be a divisor of the linear combination $d = x_0 a + y_0 b$. So $d' \leq d$ (by Proposition 2.1(c)). Thus, d is the greatest common divisor of a and b, as desired. ■

The proof shows that every common divisor must divide the gcd. This is worth recording.

Corollary 3.8 Let $a, b, c \in \mathbb{Z}$. If $c \mid a$ and $c \mid b$, then $c \mid \gcd(a, b)$.

This is all well and good, but if you actually want to find the x and y in Bézout's Lemma, you can do so by running the Euclidean algorithm backwards.

Example 3.9 Express $\gcd(693, 105)$ as a linear combination of 693 and 105.

Solution: We carried out the Euclidean algorithm on this pair of integers above. Here it is again:

$$\begin{aligned} 693 &= 6 \cdot 105 + 63 \\ 105 &= 1 \cdot 63 + 42 \\ 63 &= 1 \cdot 42 + 21 \\ 42 &= 2 \cdot 21 + 0. \end{aligned}$$

So $\gcd(693, 105) = 21$. Running this backwards, starting from the third equation, we get:

$$\begin{aligned} 21 &= 63 - 1 \cdot 42 \\ &= 63 - 1 \cdot (105 - 1 \cdot 63) \\ &= 2 \cdot 63 - 1 \cdot 105 \\ &= 2(693 - 6 \cdot 105) - 1 \cdot 105 \\ &= 2 \cdot 693 - 13 \cdot 105. \end{aligned}$$

Thus,

$$\gcd(693, 105) = 2 \cdot 693 + (-13) \cdot 105$$

is our desired linear combination.

Exercise 3.10 Express $\gcd(1234, 5678)$ as an integer linear combination of 1234 and 5678.

REMARK (Computational Complexity of the Euclidean Algorithm)

It can be shown that the maximum number of steps required to apply the Euclidean algorithm to positive integers a and b is $O(\log n)$, where $n = \max(a, b)$. Each step involves a division, which roughly takes $O((\log n)^2)$ time. So the total complexity of the Euclidean algorithm is $O((\log n)^3)$. (There are more efficient implementations.)

The naive check-each-divisor approach has complexity $O(n(\log n)^2)$ if a and b are roughly of the same size. This is significantly worse than the Euclidean algorithm when n is large.

Lecture 3 Problems

3.1. Let $a, b, c \in \mathbb{Z}$. Prove that c is a common divisor of a and b if and only if $c \mid \gcd(a, b)$.

3.2. Let $a, b, d, n \in \mathbb{Z}$. Prove:

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

(b) If $d \neq 0$ and $d \mid a$ and $d \mid b$ then $\gcd(\frac{a}{d}, \frac{b}{d}) = \frac{1}{|d|} \gcd(a, b)$.

3.3. The **Fibonacci numbers** $f_0, f_1, f_2, \ldots$ are defined recursively by

$$f_0 = f_1 = 1, \quad f_{n+1} = f_n + f_{n-1} \text{ for } n \geq 1.$$

Prove that $\gcd(f_n, f_{n+1}) = 1$ for all $n \in \mathbb{Z}_{\geq 0}$.

3.4. For $a_1, \ldots, a_n \in \mathbb{Z}$ not all zero, we can define $\gcd(a_1, \ldots, a_n)$ exactly how we did in the case where $n = 2$, namely: it's the largest integer that divides all of $a_1, \ldots, a_n$.

(a) Prove that $\gcd(a_1, a_2, \ldots, a_n) = \gcd(\gcd(a_1, a_2), \ldots, a_n)$.

(b) Compute $\gcd(20, 28, 100, 36)$.

Lecture 4 Coprimality

We'll often find ourselves in the following situation. We have integers $a, b, c \in \mathbb{Z}$ such that $a \mid bc$ and we'd like to conclude that either $a \mid b$ or $a \mid c$. This conclusion is generally false. For example, $4 \mid 2 \cdot 6$ but $4 \nmid 2$ and $4 \nmid 6$. The problem is that "part of 4" can be found in 2 and 6 but not in either of them separately—said differently, $\gcd(4, 2) > 1$ and $\gcd(4, 6) > 1$. This prompts the following definition.

Definition 4.1 **Coprime, Relatively Prime**

Two integers $a, b \in \mathbb{Z}$ are said to be **coprime** (or **relatively prime**) if $\gcd(a, b) = 1$.

The integers $a_1, \ldots, a_n \in \mathbb{Z}$ are said to be **mutually coprime** if $\gcd(a_1, \ldots, a_n) = 1$. They are said to be **pairwise coprime** if $\gcd(a_i, a_j) = 1$ for all $i \neq j$.

For example, 4 and 15 are coprime, but 4 and 6 are not. The integers 2, 3 and 4 are mutually coprime since $\gcd(2, 3, 4) = 1$ but they are not pairwise coprime, since $\gcd(2, 4) \neq 1$.

The following result is an immediate consequence of Proposition 3.7 (Bézout's Lemma).

Proposition 4.2 The integers $a, b \in \mathbb{Z}$ are coprime if and only if there exist integers $x, y \in \mathbb{Z}$ such that

$$ax + by = 1.$$

This result can be immensely helpful when it comes to proving things involving coprime integers. Here is an illustration.

Proposition 4.3 Let $a, b, c \in \mathbb{Z}$.

(a) If $a \mid bc$ and if a and b are coprime, then $a \mid c$.

(b) If $a \mid c$ and $b \mid c$, and if a and b are coprime, then $ab \mid c$.

Proof: For part (a), start by writing $1 = ax + by$ with $x, y \in \mathbb{Z}$. Then multiply through by c to get $c = a(xc) + (bc)y$. Since $a \mid a$ and $a \mid bc$, it follows that $a \mid a(xc) + (bc)y = c$.

The proof of (b) is left for you as an exercise. ■

Exercise 4.4 Prove part (b) of Proposition 4.3.

This proposition explains the issue with trying to go from $4 \mid 2 \cdot 6$ to $4 \mid 2$ or $4 \mid 6$. We cannot make this jump since 4 is not coprime with either 2 or 6.

Example 4.5

Let $a, b \in \mathbb{Z}$ be non-zero integers. Prove that $a/\gcd(a,b)$ and $b/\gcd(a,b)$ are coprime.

Solution: Let $g = \gcd(a,b)$. We can write $g = ak + bl$ with $k, l \in \mathbb{Z}$ by Bézout's lemma. Dividing through by g, we get

$$1 = \frac{a}{g}k + \frac{b}{g}l.$$

Now apply Proposition 4.2.

The previous example should make intuitive sense: when we divide by $\gcd(a,b)$, we remove all common factors from a and b, and so the resulting integers $a/\gcd(a,b)$ and $b/\gcd(a,b)$ must be coprime.

As an application, let's solve a family of Diophantine equations!

4.1 Linear Diophantine Equations

A **linear Diophantine equation** is an equation of the form

$$a_1x_1 + \cdots + a_nx_n = b$$

where $a_1, \ldots, a_n, b \in \mathbb{Z}$ and we want our solutions for the x_i to be in $\mathbb{Z}$ too.

If $n = 1$ the equation takes the form

$$ax_1 = b.$$

This equation has a solution in $\mathbb{Z}$ if and only if $a \mid b$, in which case the solution is $x_1 = \frac{b}{a}$.

So the first interesting case is $n = 2$. It also turns out[6] that we can reduce equations with $n > 2$ variables to ones with two variables, so really the $n = 2$ case is the most interesting one. Fortunately, we know how to solve it completely. To ease notation, let's drop all these subscripts and work with $ax + by = c$.

Theorem 4.6 **(Solvability of Linear Diophantine Equations)**

Suppose that $a, b, c \in \mathbb{Z}$, and consider the Diophantine equation

$$ax + by = c. \tag{$*$}$$

(a) Equation $(*)$ has a solution (x_0, y_0) with $x_0, y_0 \in \mathbb{Z}$ **if and only if** $\gcd(a,b) \mid c$.

(b) If $(x,y) = (x_0, y_0)$ is one particular integer solution to $(*)$, then the general integer solution is given by

$$(x,y) = (x_0, y_0) + n(-b/g, a/g), \tag{$**$}$$

where $n \in \mathbb{Z}$ is arbitrary, and $g = \gcd(a,b)$.

In particular, if $(*)$ has one solution, then it has infinitely many; and if we can find just one solution, then we can find them all.

[6] As you'll explore in the end-of-lecture problems.

Proof: If (∗) has an integer solution (x_0, y_0), then $c = ax_0 + by_0$ will be divisible by any common divisor of a and b—so, in particular, by $g = \gcd(a, b)$. Conversely, if $g \mid c$, then we can write $c = gm$ for some $m \in \mathbb{Z}$. By Bézout's Lemma, we can find $k, l \in \mathbb{Z}$ such that

$$g = ak + bl.$$

If we multiply this equation through by m, we immediately see that (km, lm) is a solution to (∗). This proves part (a).

For part (b), assume that (x_0, y_0) is a solution. I'll leave it to you to check that (∗∗) is a solution to (∗). It remains to show all solutions are of this form for some $n \in \mathbb{Z}$. To this end, suppose that (x_1, y_1) is another integer solution. Then

$$ax_0 + by_0 = ax_1 + by_1.$$

Re-arranging and dividing by g, we get

$$\frac{a}{g}(x_0 - x_1) = \frac{b}{g}(y_1 - y_0). \qquad (\spadesuit)$$

This shows that a/g divides $(b/g)(y_1 - y_0)$ and therefore, since a/g and b/g are coprime, it follows that a/g must divide $y_1 - y_0$. (This is the kind of thing I mentioned at the beginning of the lecture!) Thus, $y_1 - y_0 = (a/g)n$ for some $n \in \mathbb{Z}$ or, equivalently,

$$y_1 = y_0 + \frac{a}{g}n.$$

Substituting this into (♠), we find that

$$x_1 = x_0 - n\frac{b}{g}.$$

Thus, every solution is of the form given in (∗∗), which is what we wanted to prove. ■

REMARK

On page 7, we noted that there are three fundamental questions to ask about any given Diophantine equation. Theorem 4.6 answers all three for the equation $ax + by = c$: It provides us with an easy check for whether a solution exists, and it tells us how many solutions there are and how to find them—indeed, the proof contains an algorithm:

- Determine $g = \gcd(a, b)$. (Use the Euclidean algorithm.)
- If $g \nmid c$, STOP: no solution exists. Otherwise, proceed to next step.
- Determine $k, l \in \mathbb{Z}$ such that $ak + bl = g$. (Reverse Euclidean algorithm.)
- A particular solution is then given by $(x_0, y_0) = (ck/g, cl/g)$. (Proof: Multiply above equation by c/g: $a(ck/g) + b(cl/g) = c$.)
- The general solution is then given by (∗∗).

Let's illustrate how the algorithm works.

Example 4.7 Find all integer solutions to $6x + 22y = 10$.

Solution: First, we find $\gcd(6, 22)$ using the Euclidean algorithm:

$$\begin{aligned} 22 &= 3 \cdot 6 + 4 \\ 6 &= 1 \cdot 4 + 2 \\ 4 &= 2 \cdot 2 + 0. \end{aligned}$$

So $\gcd(6, 22) = 2$ and since $2 \mid 10$, we are guaranteed that the equation has a solution. To find one, we reverse the Euclidean algorithm to express $\gcd(6, 22)$ as an integer linear combination of 6 and 22 (Bézout's Lemma):

$$\begin{aligned} 2 &= 6 - 1 \cdot 4 \\ &= 6 - 1 \cdot (22 - 3 \cdot 6) \\ &= 6 \cdot 4 + 22 \cdot (-1). \end{aligned}$$

Multiplying through by $b/\gcd(a_1, a_2) = 10/2$, this gives

$$10 = 6 \cdot 20 + 22 \cdot (-5).$$

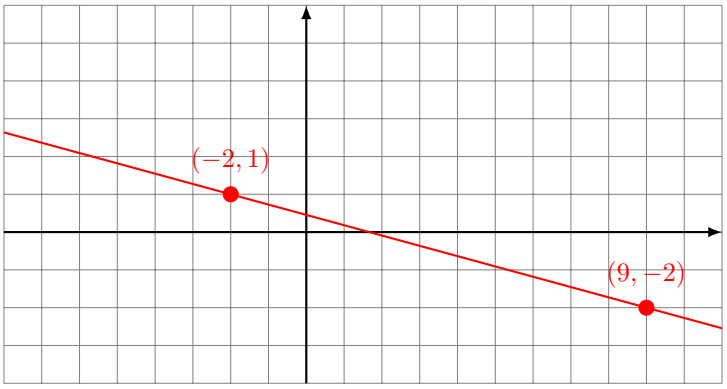
So $(x, y) = (20, -5)$ is a particular solution. The general solution is therefore

$$\begin{aligned} (x, y) &= (20, -5) + n(-22/2, 6/2) \\ &= (20, -5) + n(-11, 3), \quad n \in \mathbb{Z}. \end{aligned}$$

Note that the form of the general solution is somewhat dependent on our choice of particular solution. Had we used the particular solution $(x, y) = (-2, 1)$ instead, our general solution would have looked like

$$(x, y) = (-2, 1) + n(-11, 3), \quad n \in \mathbb{Z}.$$

One thing worth pointing out is that the equation $6x + 22y = 10$ defines a line in $\mathbb{R}^2$. (Indeed, we can re-write the equation into the more familiar form $y = -\frac{6}{22}x + \frac{10}{22}$.) What we are doing here is finding the *lattice points* $(x, y) \in \mathbb{Z}^2$ that lie on this line.



Exercise 4.8 Find all integer solutions to $5x + 7y = 23$.

Example 4.9 **(Postage Stamp Problem)**

Assuming you can only buy 5 cent and 7 cent postage stamps, prove that you cannot mail a letter that costs 23 cents to post.

Solution: The problem here is to show that the Diophantine equation

$$5x + 7y = 23 \tag{$*$}$$

does not have any solutions with $x, y \in \mathbb{Z}_{\geq 0}$ (note: $\mathbb{Z}_{\geq 0}$ and not $\mathbb{Z}$ since we can't purchase a negative amount of stamps).

You solved this equation over $\mathbb{Z}$ in Exercise 4.8. The general solution is given by

$$(x, y) = (69, -46) + n(-7, 5), \quad n \in \mathbb{Z}.$$

Since we want $x \geq 0$ and $y \geq 0$, we are led to the inequalities

$$69 - 7n \geq 0 \quad \text{and} \quad -46 + 5n \geq 0.$$

These give

$$n \leq \frac{69}{7} \quad \text{and} \quad n \geq \frac{46}{5}.$$

That is, we want $n \in \mathbb{Z}$ such that

$$\frac{46}{5} \leq n \leq \frac{69}{7}.$$

But there is no such integer, since $46/5 = 9.2$ and $69/7 \approx 9.9$. So there are no non-negative integer solutions to ($*$).

The number 23 in the previous example is special: it's the largest problematic number, in the sense that you are able to post a letter that costs n cents for all $n > 23$. (See Problem 3 below.)

Lecture 4 Problems

4.1. Let $a, b \in \mathbb{Z}$. Suppose that there are integers $r, s, t, u \in \mathbb{Z}$ such that

$$ra + sb = 2 \quad \text{and} \quad ta + tb = 5.$$

Prove that a and b must be coprime.

4.2. Let $a \in \mathbb{Z}$.

(a) Prove that $2a - 1$ and $2a + 1$ are coprime.

(b) Prove that $a! + 1$ and $(a + 1)! + 1$ are coprime.

4.3. Let $a > 1$ and $b > 1$ be coprime integers.

(a) Prove that there are no solutions to $ax + by = ab - a - b$ with $x, y \in \mathbb{Z}_{\geq 0}$.
[**Hint:** Begin by finding a particular solution $(x_0, y_0) \in \mathbb{Z}^2$ by inspection.]

(b) If $n > ab - a - b$, prove that there are $x, y \in \mathbb{Z}_{\geq 0}$ such that $ax + by = n$.
[**Hint:** Consider the integer solution (x_1, y_1) with smallest non-negative x_1 (why is there such a solution?). Prove that $x_1 \leq b - 1$ and then deduce that $y_1 \geq 0$.]

4.4. In this problem you'll learn how to solve the linear Diophantine equation

$$ax + by + cz = e \tag{♢}$$

where $a, b, c, e \in \mathbb{Z}$. The same techniques will also be able to handle any linear Diophantine equation in more than three variables.

(a) Let $d = \gcd(a, b)$ and let $u = \frac{a}{d}x + \frac{b}{d}y$. Then (♢) is transformed into the two-variable Diophantine equation

$$du + cz = e. \tag{♣}$$

Explain how the solutions to (♣) can be used to generate all solutions to (♢).

(b) Let $g = \gcd(a_1, a_2, a_3)$. Deduce that (♢) has integer solutions if and only if $g \mid b$.

4.5. Solve the Diophantine equation $3x + 12y + 5z = 8$.

Lecture 5 The Fundamental Theorem of Arithmetic

At last, we come to prime numbers! The following definition should be familiar.

Definition 5.1 **Prime Numbers, Composite**

A **prime number** is an integer $p > 1$ whose only positive divisors are 1 and p. An integer that isn't prime is said to be **composite.**

The prime numbers ≤ 100 are

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

We do not consider 1 to be a prime number because doing so would violate the pleasant fact that every positive integer can be written as a product of primes in a unique way. We formally state this *unique factorization* result as a theorem.

Theorem 5.2 **(Fundamental Theorem of Arithmetic)**

Every integer $n > 1$ can be expressed as a product of primes in a unique way (up to re-ordering).

Thus, for example, we view the factorizations

$$12 = 2 \cdot 2 \cdot 3 = 2 \cdot 3 \cdot 2 = 3 \cdot 2 \cdot 2$$

of $n = 12$ as being essentially the same. The uniqueness in the theorem is the assertion that the only primes that appear in the prime factorization of 12 are 2 (which appears twice) and 3 (which appears once). Had we allowed 1 to be a prime, we would have infinitely many distinct factorizations of 12 into a product of primes:

$$12 = 2^2 \cdot 3 \cdot 1 = 2^2 \cdot 3 \cdot 1^2 = \cdots .$$

The key tool that will allow us to prove the uniqueness of prime factorizations is the following lemma, which is essentially a special case of Proposition 4.3(a).

Lemma 5.3 **(Euclid's Lemma)**

Let $a, b \in \mathbb{Z}$. If p is prime and $p \mid ab$ then either $p \mid a$ or $p \mid b$.

Proof: For any $a \in \mathbb{Z}$, $\gcd(a, p)$ is either 1 or p. If it's 1, then apply Proposition 4.3(a). If it's p, then that means $p \mid a$. ■

Exercise 5.4 Let $a_1, \ldots, a_n \in \mathbb{Z}$. Prove that if p is prime and if $p \mid a_1 \cdots a_n$ then $p \mid a_i$ for some i.

Proof of Theorem 5.2 (Fundamental Theorem of Arithmetic):

Existence: If the theorem were false, then there would be a smallest integer $n_0 > 1$ that cannot be expressed as a product of primes. Note that n_0 itself cannot be prime and so we can factor it as $n_0 = ab$ where $a, b \in \mathbb{Z}_{>0}$ are both different from 1 and n_0. In particular, $a < n_0$ and $b < n_0$, so by the minimality assumption on n_0, we can factor $a = \prod p_i$ and $b = \prod q_i$ into a product of primes. But then $n_0 = ab = \prod p_i q_i$ is a product of primes! Contradiction.

Uniqueness: Suppose $n = p_1 \cdots p_k$ and $n = q_1 \cdots q_l$, where the p_i and q_j are primes. Then since p_i divides $n = q_1 \cdots q_l$, and since the q_j are coprime, Euclid's lemma tells us that $p_i \mid q_{j_0}$ for some j_0. This implies that $p_i = q_{j_0}$ since q_{j_0} is prime. Thus, every p_i occurs amongst the q_j. Similarly, by considering $q_j \mid n = p_1 \cdots p_k$, we see that every q_j occurs amongst the p_i. This completes the proof. ■

We usually collect prime factors together when we express n as a product of primes. That is, we write

$$n = p_1^{a_1} \cdots p_k^{a_k}$$

where the p_i are *distinct* primes and the a_i are *positive* integers. This is called the **canonical factorization** of n into a product of primes. It is unique up to re-ordering. If we wish to eliminate the ambiguity of ordering, we can choose to list the primes in ascending order: $p_1 < p_2 < \cdots < p_k$.

For example, the canonical factorization of 1400 is

$$1400 = 2^3 \cdot 5^2 \cdot 7.$$

If $n < -1$ then, by factoring $-n$, we can obtain unique factorization of n into (-1) times a product of primes. For example,

$$-132 = (-1) \cdot 2^2 \cdot 3 \cdot 11.$$

In this way, we see that every integer n, other than $n = 0$ and $n = \pm 1$, can be expressed uniquely in the form

$$n = (-1)^s p_i^{a_i} \cdots p_k^{a_k}$$

where $s \in \{0, 1\}$ is the **sign** of n, the p_i are the **distinct prime divisors** of n, and the a_i are the so-called prime valuations of n. Let's elaborate on this last bit.

Definition 5.5
p-adic Valuation, v_p

Let p be a prime and n be a non-zero integer. The **p-adic valuation** of n, denoted by $v_p(n)$, is defined to be the largest integer a such that $p^a \mid n$. (So $p^{v_p(n)} \mid n$ but $p^{v_p(n)+1} \nmid n$.)

We do not define $v_p(0)$.

For example,

$$v_2(18) = 1 \text{ since } 2^1 \mid 18 \text{ but } 2^2 \nmid 18$$

and

$$v_3(18) = 2 \text{ since } 3^2 \mid 18 \text{ but } 3^3 \nmid 18.$$

For all other primes $p \neq 2, 3$, we have $v_p(18) = 0$.

In general, if

$$n = (-1)^s p_i^{a_i} \cdots p_k^{a_k}$$

is the canonical factorization of n, then $v_p(n) = a_i$ for $p = p_i$, and $v_p(n) = 0$ for all other primes p. In particular, $v_p(n) = 0$ for all $p \nmid n$.

Exercise 5.6 Determine $v_p(100)$ for all primes p.

Example 5.7 **(Legendre's formula for $v_p(n!)$)**

Let $n \in \mathbb{Z}_{>0}$ and let p be prime. Show that

$$v_p(n!) = \left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \left\lfloor \frac{n}{p^3} \right\rfloor + \cdots .$$

(Note that the sum is finite since $\lfloor n/p^k \rfloor = 0$ for all sufficiently large k.)

Solution: Since $n! = 1 \cdot 2 \cdots n$, we need to count the multiples of p that are $\leq n$.

There are $\lfloor n/p \rfloor$ multiples of p. Of these, the multiples of p^2 are only counted once, but they should be counted twice since they contribute another p in $n!$. So we should add $\lfloor n/p^2 \rfloor$. Likewise, to properly account for the multiples of p^3, we should add $\lfloor n/p^3 \rfloor$, and so on.

As an illustration, we have

$$\begin{aligned} v_5(100!) &= \left\lfloor \frac{100}{5} \right\rfloor + \left\lfloor \frac{100}{5^2} \right\rfloor + \left\lfloor \frac{100}{5^3} \right\rfloor + \cdots \\ &= \lfloor 20 \rfloor + \lfloor 4 \rfloor + 0 + \cdots \\ &= 24. \end{aligned}$$

Exercise 5.8 Determine $v_3(1000!)$.

REMARK (Unique Factorization in Other Number Systems)

That we have unique factorization into primes in $\mathbb{Z}$ is something special. It should not be taken for granted since it can fail in other number systems. For example, in

$$\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} \colon a, b \in \mathbb{Z}\} \subseteq \mathbb{C},$$

the number 6 has two distinct factorizations

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

into "primes." This phenomenon is the starting point of *algebraic number theory*.

Exercise 5.9 Find two distinct factorizations of 9 in $\mathbb{Z}[\sqrt{-5}]$. [**Hint:** For a clue, try to reverse-engineer the factorization of 6 given in the previous remark.]

Lecture 5 Problems

5.1. Prove the following "converse" to Lemma 5.3 (Euclid's Lemma): If $q \in \mathbb{Z}_{>0}$ is such that whenever $q \mid ab$ then $q \mid a$ or $q \mid b$, then q must be a prime number.

5.2. Let $a, b \in \mathbb{Z}_{>0}$. Prove that if $a^5 \mid b^2$ then $a \mid b$.

5.3. Let a and b be non-zero integers. Prove:

(a) $v_p(ab) = v_p(a) + v_p(b)$ for all primes p.
(b) $v_p(a^k) = kv_p(a)$ for all primes p and $k \in \mathbb{Z}_{>0}$.
(c) $v_p(a+b) \geq \min(v_p(a), v_p(b))$, with equality if $v_p(a) \neq v_p(b)$.
(d) $a \mid b$ if and only if $v_p(a) \leq v_p(b)$ for all primes p.
(e) a is a perfect kth power if and only if $k \mid v_p(a)$ for all primes p.

5.4. Let $a, b \in \mathbb{Z}_{>0}$ have prime factorizations $a = \prod_i p_i^{a_i}$ and $b = \prod_i p_i^{b_i}$, where $a_i \geq 0$ and $b_i \geq 0$ to allow for the same set of primes to occur in both factorizations. Prove that

$$\gcd(a, b) = \prod_i p^{m_i}$$

where $m_i = \min(a_i, b_i)$.

5.5. Let p and q be prime numbers.

(a) Determine the number of divisors of p^k, where $k \in \mathbb{Z}_{>0}$.
(b) Determine the number of divisors of $p^k q^l$, where $k, l \in \mathbb{Z}_{>0}$.

5.6. Let $n, k \in \mathbb{Z}_{>0}$ and let p be prime. Recall the definition of the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

(a) Prove that

$$v_p\left(\binom{n}{k}\right) = \sum_{i=1}^{\infty} \left\lfloor \frac{n}{p^i} \right\rfloor - \left\lfloor \frac{n-k}{p^i} \right\rfloor - \left\lfloor \frac{k}{p^i} \right\rfloor.$$

(b) Determine $v_2\left(\binom{600}{300}\right)$.

▶ 5.7. Let p be a prime number. Prove that

$$p \mid \binom{p}{k} \quad \text{and} \quad p^2 \nmid \binom{p}{k}$$

for all integers k such that $0 < k < p$.

Lecture 6 The Infinitude of Primes

The most basic fact about the distribution of prime numbers is following.

Theorem 6.1 There are infinitely many primes.

I will present two proofs of this theorem.

Proof 1 of Theorem 6.1 (Euclid): Suppose there are only finitely many primes; call them $p_1, p_2, \ldots, p_n$. Let $N = p_1 \cdots p_n + 1$. Then $N > 1$ so N has a prime divisor. It must be one of the p_i since these are all the primes. But N leaves a remainder of 1 after division by p_i, so in particular $p_i \nmid N$. Contradiction! ■

Let me make three comments about Euclid's proof.

1. If we let p_n denote the nth prime (so $p_1 = 2$, $p_2 = 3$, etc.), then Euclid's proof shows that $p_{n+1} \leq p_1 \cdots p_n + 1$. Using this, we can show that
$$p_n \leq 2^{2^{n-1}}.$$
(See the end-of-lecture problems.) This bound, however, is not great. For instance, when $n = 10$, it tells us that $p_{10} = 29$ is no larger than 2^{2^9}, a number which has 155 digits...

2. A common misconception is to think Euclid's proof shows that $p_1 \cdots p_n + 1$ is a prime. This is false. For example, while
$$\begin{aligned} 2 + 1 &= 3 \\ 2 \cdot 3 + 1 &= 7 \\ 2 \cdot 3 \cdot 5 + 1 &= 31 \\ 2 \cdot 3 \cdot 5 \cdot 7 + 1 &= 211 \\ 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 &= 2311 \end{aligned}$$
are primes,
$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509$$
is composite. This prompts the following question.

   Are there infinitely many primes of the form $p_1 \cdots p_n + 1$?

   This is an *open problem*. No one knows what the answer is!

3. This is less of a comment and more of a joke.[7] *Theorem:* There are infinitely many composite numbers. *Proof:* Suppose there are finitely many and then multiply them but don't add 1.

As dumb as that joke was, the underlying idea gives the following interesting result.

[7] It's a math joke, where the standards for humor are lower.

Example 6.2 Prove that there are arbitrarily long sequences of consecutive composite numbers. That is, for each $n > 1$, find n consecutive composite numbers.

Solution: The "joke" teaches us that $(n+1)! = (n+1) \cdot n \cdots 2 \cdot 1$ is divisible by every positive integer $k \leq (n+1)$.

Hence $k + (n+1)!$ will be divisible by k if $k \leq (n+1)$.

Consequently, the n consecutive numbers

$$2 + (n+1)!,\ 3 + (n+1)!,\ \ldots,\ n + (n+1)!$$

are composite.

This shows that the primes, while infinite, can be arbitrarily far apart. So if we try to make a list of primes by going through the integers one-by-one, then it might be quite a while between each discovery.

On the other hand, sometimes primes are fairly close to each other: for instance, $(2, 3)$, $(3, 5)$ and $(5, 7)$ are all primes, as are $(641, 643)$ and $(197597, 197599)$. Primes that differ by ≤ 2 are called **twin primes**. It is an open problem whether there are infinitely many twin primes. In 2013, Yitang Zhang surprised the mathematical world by proving that there are infinitely many primes that differ by ≤ 70 million. The bound was subsequently reduced to 246. Unfortunately, the consensus is that Zhang's techniques cannot be further optimized to reduce the bound to 2, so new ideas are needed to settle the twin prime conjecture.

We'll have a bit more to say about the distribution of primes, but before going down that rabbit hole, let's look at another proof of the infinitude of primes.

Proof 2 of Theorem 6.1 (Euler): The starting point is the geometric series expansion

$$\frac{1}{1-1/p} = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots.$$

Consider what happens when we multiply two of these series for different primes p. For example, for the primes 2 and 3, we want to look at

$$\frac{1}{1-1/2}\frac{1}{1-1/3} = \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots\right)\left(1 + \frac{1}{3} + \frac{1}{3^2} + \frac{1}{3^3} + \cdots\right)$$

If we expand the right-side and re-arrange it (which is OK to do since we're dealing with convergent series of positive numbers), we get

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{9} + \frac{1}{12} + \frac{1}{16} + \frac{1}{18} + \cdots.$$

The denominators that appear are precisely the positive integer n whose only prime divisors are 2 and 3. If we multiply the above by the series for $1/(1-1/p)$ for another prime p, then we will introduce additional terms $1/n$ where the only prime divisors of n are $2, 3$ and p.

Now assume that there are finitely many primes $p_1, \ldots, p_k$. Then if we multiply out

$$\frac{1}{1-1/p_1}\frac{1}{1-1/p_2}\cdots\frac{1}{1-1/p_k}$$

as above, we obtain all numbers of the form $1/n$ where the prime divisors of n involve only the p_i. Since the p_i are *all* the primes, this means we obtain all positive integers n, thanks to the Fundamental Theorem of Arithmetic.

Thus,

$$\frac{1}{1-1/p_1}\frac{1}{1-1/p_2}\cdots\frac{1}{1-1/p_k} = \sum_{n=1}^{\infty}\frac{1}{n}.$$

This, however, is impossible! The series on the right—the famous harmonic series—is divergent. It cannot be equal to the finite product on the left-side. This contradiction proves that there must be infinitely many primes. ■

Euler's proof marks the beginning of *analytic number theory*, where ideas from calculus (analysis) are brought to bear on number theoretic problems. The ideas in the proof can be tweaked to yield many other interesting results, such as the following (which we state without proof).

Proposition 6.3 **(Euler)**

The infinite series

$$\sum_p \frac{1}{p} = \frac{1}{2}+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\frac{1}{11}+\cdots,$$

where p ranges over the prime numbers, is divergent.

This result can be viewed as a strengthening of the fact that there are infinitely many primes. Indeed, since $\sum_n 1/n$ and $\sum_p 1/p$ are both divergent while $\sum_n 1/n^2$ is convergent, in some sense we can say that the primes are "more infinite" than the perfect squares. To learn how to make this idea more precise, take a course in analytic number theory!

6.1 Primes and Arithmetic Progressions

Besides the prime number 2, prime numbers are odd and hence will leave a remainder of 1 or 3 after division by 4. Since there are infinitely many primes, either infinitely many will have remainder 1, or infinitely many will have remainder 3, or both of these scenarios will happen. Which is it?

Proposition 6.4 There are infinitely many primes of the form $4q+3$.

Proof: We're going to mimic Euclid's proof. Suppose there are only finitely many such primes, call them $p_1,\ldots,p_k$, and set $N = 4p_1\cdots p_k - 1$. Then N is odd (so not divisible by 2) and not divisible by any of the p_i. So it must be the case that all the prime divisors prime divisors are of the form $4q+1$.

But this is impossible! The product of numbers of the form $4q+1$ is also of the form $4q+1$:

$$(4q_1+1)(4q_2+1) = 16q_1q_2+4q_1+4q_2+1 = 4(4q_1q_2+q_1+q_2)+1.$$

Since N is not of the form $4q+1$, we've reached a contradiction. ■

There are also infinitely many primes of the form $4q+1$, but the proof is a little beyond us at this point. We'll come back to it later. In fact, much more is true! This next result is a famous theorem of Dirichlet, which we give without proof. (The proof that Dirichlet gave involved an ingenious modification of Euler's already ingenious proof of the infinitude of primes—in particular, it brought in some ideas from complex analysis!)

Theorem 6.5 **(Dirichlet's Theorem on Primes in Arithmetic Progressions)**

If $\gcd(a,b)=1$, then there are infinitely many primes of the form $a+bq$.

The gcd condition is necessary since if $d=\gcd(a,b)>1$ then the numbers $a+bq$ would all be divisible by d. The name of the theorem refers to the fact that the numbers $a+bq$, as q varies, form an arithmetic progression. For instance, we are guaranteed to find infinitely many primes in the arithmetic progression

$$7,\ 10,\ 13,\ 16,\ \ldots,\ 7+3q,\ \ldots.$$

On the flip side, we might ask if we can find arithmetic progressions consisting entirely of primes. It's impossible to find an infinite such progression (why?) so we might ask instead for *really long* ones. The fact that this is possible was only proved in 2004!

Theorem 6.6 **(Green–Tao)**

For all $N\in\mathbb{Z}_{>0}$, there is an arithmetic progression of length N consisting entirely of primes.

For example, when $N=5$, the progression $5+6q$ consists of primes for $0\le q\le 4$:

$$5,\ 11,\ 17,\ 23,\ 29.$$

For larger N, we'll have to go deeper into the primes. Unfortunately, the proof of the Green–Tao theorem is not constructive, and there are no efficient methods for generating prime progressions. At the time of writing, the longest known progression has length $N=27$. Three such progressions are known; the first was discovered in 2019 and is given by

$$224584605939537911+181356965979489 30q,\quad 0\le q\le 26.$$

Exercise 6.7 Find an arithmetic progression of length 6 consisting of primes. [**Hint:** There is one starting at a single-digit prime and ending at 157.]

Lecture 6 Problems

▶ 6.1. Let p_n denote the nth prime number.

(a) Explain how Euclid's proof shows that $p_{n+1}\le p_1\cdots p_n+1$.

(b) Prove by induction that $p_n\le 2^{2^{n-1}}$ for all $n\ge 1$.

6.2. Let $F_n=2^{2^n}+1$ for $n\in\mathbb{Z}_{\ge 0}$. (This is the nth **Fermat number**, named after Fermat who falsely believed F_n to be prime for all n.)

(a) Prove that $F_n - 2 = F_0 F_1 \cdots F_{n-1}$ for all $n \geq 1$.

(b) Prove that $\gcd(F_n, F_m) = 1$ for $n \neq m$.

(c) Deduce that there are infinitely many primes.

6.3. Prove that there are infinitely many primes of the form $3q + 2$.

6.4. Unlike with twin primes, we can prove that there are finitely many "prime triplets": Find all primes p such that $p + 2$ and $p + 4$ are primes.

6.5. Prove that there cannot be an infinite arithmetic progression consisting entirely of primes. That is, prove that if $a, b \in \mathbb{Z}_{>0}$, then there must be a composite number in the arithmetic progression

$$a, a + b, a + 2b, a + 3b, \ldots.$$

Lecture 7 The Distribution of Primes

> *"There are two facts about the distribution of prime numbers of which I hope to convince you [...]. The first is that, despite their simple definition and role as the building blocks of the natural numbers, the prime numbers grow like weeds among the natural numbers, seeming to obey no other law than that of chance, and nobody can predict where the next one will sprout. The second fact is even more astonishing, for it states just the opposite: that the prime numbers exhibit stunning regularity, that there are laws governing their behavior, and that they obey these laws with almost military precision."*
>
> – Don Zagier

This lecture will showcase some results from analytic number theory. It will be relatively light on proofs.

The Sieve of Eratosthenes

Before diving in, let's look at a simple method for tabulating prime numbers. Suppose we want to find all the primes ≤ 50. We start by listing all the integers from 2 to 50.

	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50

Next, we circle 2 indicating that it's prime and then cross off every multiple of 2 since none of them can be prime.

	(2)	3	~~4~~	5	~~6~~	7	~~8~~	9	~~10~~
11	~~12~~	13	~~14~~	15	~~16~~	17	~~18~~	19	~~20~~
21	~~22~~	23	~~24~~	25	~~26~~	27	~~28~~	29	~~30~~
31	~~32~~	33	~~34~~	35	~~36~~	37	~~38~~	39	~~40~~
41	~~42~~	43	~~44~~	45	~~46~~	47	~~48~~	49	~~50~~

The next number that isn't circled or crossed off, which in this case is 3, must then be prime. So we circle it and cross off all its multiples.

	(2)	(3)	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	25	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	35	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	49	~~50~~

We repeat this process a couple more times, first with 5 and then with 7.

	②	③	~~4~~	⑤	~~6~~	⑦	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~

Since $\sqrt{50} \approx 7.07$, we can now stop: all remaining numbers that haven't been crossed off are necessarily prime. This is because we have the following result.

Proposition 7.1 Let $n > 1$. Then either n is prime or else n has a prime divisor p such that $p \leq \sqrt{n}$.

Proof: If n is prime, that's that. If n is composite, then we can factor it as $n = ab$ with $1 < a < n$ and $1 < b < n$. One of a and b must be $\leq \sqrt{n}$, since if they were both $> \sqrt{n}$ we would have

$$n = ab > \sqrt{n}\sqrt{n} = n,$$

which is absurd. So let's say $a < \sqrt{n}$. Since $a > 1$, it must have a prime divisor p and this p will divide n (why?) and must satisfy $p < a < \sqrt{n}$. ■

We can reformulate the previous result as follows.

Corollary 7.2 **(Simple Primality Test)**

Let $n > 1$. If none of the primes $\leq \lfloor \sqrt{n} \rfloor$ divide n, then n must be prime.

Corollary 7.2 guarantees that any composite number ≤ 50 will be divisible by some prime $\leq \lfloor \sqrt{50} \rfloor = 7$. So once we've removed all the composite numbers divisible by the primes ≤ 7 from our list, we've removed *all* composite numbers. Thus, the primes ≤ 50 are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47.$$

This method of finding primes is called the **sieve of Eratosthenes**. We're basically passing numbers through a sieve and catching the ones that are prime, hence the name.

Exercise 7.3 Find all primes ≤ 100.

Example 7.4 Determine if 691 is prime.

Solution: Since $\lfloor \sqrt{691} \rfloor = 26$, we just have to check if 691 is divisible by any of the primes ≤ 23. I'll leave it to you to check that it isn't. So 691 is prime.

Exercise 7.5 Determine if 1891 is prime.

REMARK (Computational Complexity of the Sieve of Eratosthenes)

The sieving technique just described is not very efficient from a computational perspective. Its complexity is *exponential* in the number of bits of the input. The same can be said about Corollary 7.2 (Simple Primality Test). This makes them completely impractical when the numbers under consideration get large.

For instance, to check whether the 33265-digit number $2^{110503}-1$ is prime (it is!), a computer executing one-billion bit operations per second would take approximately 10^{16619} years to run Corollary 7.2.

The Prime Counting Function

The sieving procedure we just learned raises a natural question. Can we determine, beforehand, how many primes there are less than a given constant?

Definition 7.6

Prime Counting Function, $\pi(x)$

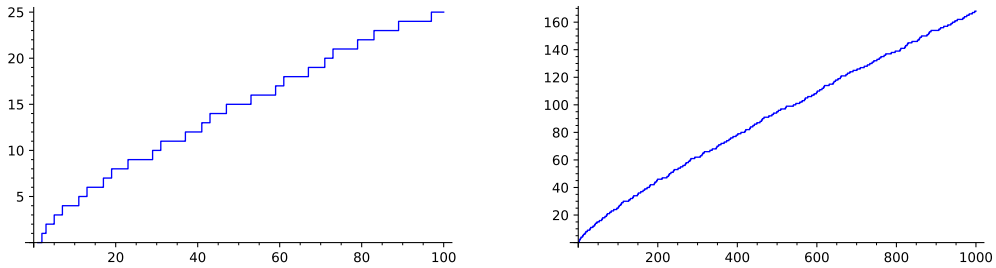
For $x \in \mathbb{R}_{\geq 0}$, the number of primes $\leq x$ is denoted by $\pi(x)$. That is,

$$\pi(x) = \#\{p\colon p \leq x \text{ and } p \text{ is a prime}\}.$$

We call $\pi(x)$ the **prime counting function**.[8]

For example,

$$\begin{aligned}\pi(1) &= \#\{\} = 0\\ \pi(10) &= \#\{2,3,5,7\} = 4\\ \pi(21.34) &= \#\{2,3,5,7,11,13,17,19\} = 8.\end{aligned}$$



(a) $\pi(x)$ for $x \leq 100$ (b) $\pi(x)$ for $x \leq 1000$

Figure 7.1: Graphs of $\pi(x)$

Since there are infinitely many primes, we know that

$$\pi(x) \to \infty \quad \text{as} \quad x \to \infty.$$

What can we say about the growth rate of $\pi(x)$? Here is a simple lower bound.

[8] The notation $\pi(x)$ for this function is not great but unfortunately it's completely standard.

Proposition 7.7 For all $x \geq 2$, $\pi(x) > \log(\log x)$.

Proof: Let p_n denote the nth prime. Last lecture we saw that $p_n \leq 2^{2^{n-1}}$ for $n \geq 1$. This means there are at least n primes $\leq 2^{2^{n-1}}$ and therefore $\pi(2^{2^{n-1}}) \geq n$.

Given any $x \geq 2$, we can find an integer $n \geq 1$ such that $2^{2^{n-1}} \leq x < 2^{2^n}$ (why?). Since $2 < e$, it follows that $x < e^{e^n}$ and therefore

$$\log(\log x) < n.$$

On the other hand, from $2^{2^{n-1}} \leq x$ we get

$$\pi(x) \geq \pi\left(2^{2^{n-1}}\right) \geq n.$$

Now combine both inequalities to get the desired result. ■

This bound is actually quite terrible, as the following table shows.

x	$\pi(x)$	$\log(\log x)$
10^2	25	1.527...
10^3	168	1.932...
10^4	1229	2.220...
10^5	9592	2.443...
10^6	78498	2.625...

So $\pi(10^6) > \log(\log(10^6))$ tells us we can find at least 3 primes under one million. No kidding! Can we do any better? The answer is yes. In fact, we can do *much* better.

7.1 The Prime Number Theorem

In the 1790s, while presumably studying tables of primes, Gauss (at the age of 15) and Legendre independently conjectured that $\pi(x)$ grows like the function $x/\log x$:

$$\pi(x) \sim \frac{x}{\log x} \quad \text{for large } x$$

where $\sim$ means that the ratio of both sides tends to 1 as x increases.

Actually, Legendre conjectured that

$$\pi(x) \sim \frac{x}{\log x - 1.08366}$$

and Gauss conjectured that

$$\pi(x) \sim \mathrm{Li}(x),$$

where $\mathrm{Li}(x)$ is the **logarithmic integral function** defined by

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}.$$

It can be shown that $\dfrac{x}{\log x - 1.08366} \sim \dfrac{x}{\log x} \sim \mathrm{Li}(x)$ as $x \to \infty$, so these approximations are asymptotically the same. The table below lists some of these values (rounded to the nearest integer).

x	$\pi(x)$	$x/\log x$	$x/(\log x - 1.08366)$	$\mathrm{Li}(x)$
10^2	25	21	28	29
10^3	168	145	172	177
10^4	1229	1086	1231	1245
10^5	9592	8686	9588	9629
10^6	78498	72382	78543	78627
10^7	664579	620421	665140	664917

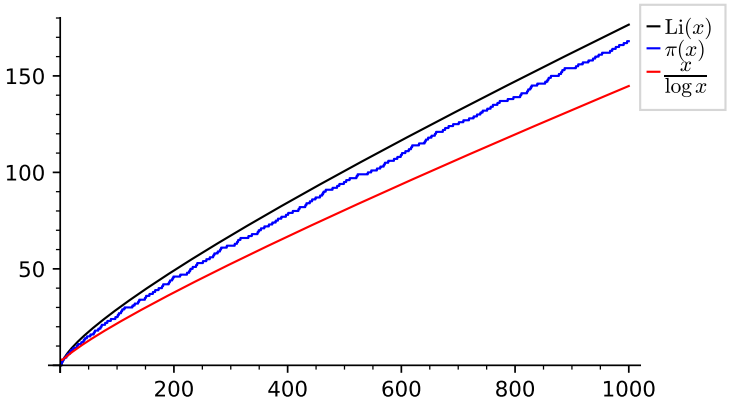


Figure 7.2: Approximations of $\pi(x)$

Here's how you can informally interpret the statement that $\pi(x) \sim x/\log x$. Since there are n positive integers $\leq n$, the approximation $\pi(n) \sim n/\log(n)$ says that approximately $1/\log n$ of the integers $\leq n$ are prime numbers.

Neither Gauss nor Legendre could prove that their empirical observations held in general. The world had to wait 100 years for J. Hadamard and C.-J. de la Vallée Poussin who independently proved, in 1896, what we now call the Prime Number Theorem.

Theorem 7.8 **(Prime Number Theorem [PNT])**

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

If we let p_n denote the nth prime, so that $p_1 = 2$, $p_2 = 3$, etc., then the PNT gives the following result. (In fact, the result below is *equivalent* to the PNT, though we won't prove that here.)

Corollary 7.9 For all $n \geq 1$,

$$p_n \sim n \log n.$$

Proof: By the Prime Number Theorem, we have

$$\pi(p_n) \sim \frac{p_n}{\log p_n}.$$

Since $\pi(p_n) = n$, the above implies that

$$p_n \sim \pi(p_n) \log p_n = n \log p_n.$$

Using this very result, we get that

$$\log p_n \sim \log(n \log p_n) = \log n + \log(\log p_n) \sim \log n.$$

Thus, $p_n \sim n \log p_n \sim n \log n$, as desired. ■

Exercise 7.10 The above proof was a bit sketchy on details. Confirm all steps by using the fact that

$$f(n) \sim g(n) \iff \lim_{n\to\infty} \frac{f(n)}{g(n)} = 1.$$

Although the proof of the PNT lies beyond the scope of this course, let me sketch some of the details. The crucial object here is the **Riemann zeta function**

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

The connection between $\zeta(s)$ and prime numbers was actually first discovered by Euler. If you recall Euler's proof of the infinitude of primes, you'll remember that if we expand the product

$$\prod_p \frac{1}{1-1/p} = \prod_p \left(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\right)$$

then we obtain all numbers of the form $1/n$ where the prime divisors of n are the primes p occurring in the product. By the same token, if we take the product over *all* primes p, we obtain

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

This is called the Euler product expansion of $\zeta(s)$. Of course, here I am ignoring issues of convergence and whatnot. In particular, the zeta function $\zeta(s)$ only converges when $s > 1$.

Riemann's contribution starts with him viewing $\zeta(s)$ as a function of a *complex* variable $s \in \mathbb{C}$, in which case the series converges for all s with $\mathrm{Re}(s) > 1$. Riemann proved that there is a unique extension of $\zeta(s)$ to a complex analytic function defined for all $s \in \mathbb{C}\backslash\{1\}$, which we also denote by $\zeta(s)$. Let me stress that this extension is **not** given by the series $\sum_n 1/n^s$. Rather, Riemann found another complex analytic function that agrees with $\sum_n 1/n^s$ on the domain $\mathrm{Re}(s) > 1$. He then leveraged the Euler product expansion to connect properties of prime numbers to the zeroes of $\zeta(s)$.

The proof of the Prime Number Theorem has two key steps:

Step 1. Prove that $\zeta(s) \neq 0$ for all s with $\mathrm{Re}(s) = 1$.

Step 2. Deduce the Prime Number Theorem from Step 1.

Both steps require a substantial amount a bit of work.

REMARK (The "Elementary" Proof of the PNT)

Historically, proofs of number theoretic statements that used complex analytic techniques were labeled as being non-elementary, with the proof sketched above being the prime example (no pun intended). This distinction seems to be due to G.H. Hardy who believed it "extraordinarily unlikely" for there to be an "elementary" proof of the PNT that didn't use results from complex analysis.

So it came as a huge surprise when Selberg and Erdős discovered, in 1949, an elementary proof of the PNT. Their proof follows the basic outline of the complex analytic proof, except it manages to replace the complex analysis with intricate "elementary" arguments. The proof still uses real analysis, which perhaps is to be expected since the statement of the PNT involves a limit!

7.2 Bonus: Bertrand's Postulate

"Chebyshev said it, but I'll say it again: There's always a prime between n and $2n$."

– N.J. Fine

Here's an interesting result concerning the distribution of primes.

Theorem 7.11 **(Bertrand's Postulate)**

For all $n > 1$, there is a prime p such that $n < p < 2n$.

(In other words, $\pi(x) - \pi(x/2) \geq 1$ for $x > 2$.)

I will sketch Erdős's proof of this fact. The main idea is to examine the prime factorization of the binomial coefficient $\binom{2n}{n}$. Then some magic happens.

Lemma 7.12 Let p be a prime and let $v_p = v_p\left(\binom{2n}{n}\right)$. Assume that $n \geq 3$. Then:

(a) $v_p = 0$ if $p > 2n$ or $\frac{2n}{3} < p \leq n$.

(b) $v_p \leq 1$ if $\sqrt{2n} < p \leq 2n$.

(c) $v_p = 1$ if $n < p \leq 2n$.

(d) $p^{v_p} \leq 2n$.

Proof: Let's use Legendre's formula:

$$v_p = v_p((2n)!) - 2v_p(n!) = \sum_{k=1}^{\infty} \left\lfloor \frac{2n}{p^k} \right\rfloor - 2\left\lfloor \frac{n}{p^k} \right\rfloor. \qquad (*)$$

(a) If $p > 2n$ then all terms on in the sum are 0, hence $v_p = 0$.

If $\frac{2n}{3} < p \leq n$, then $2p > n$ and $3p > 2n$, so p only appears once in the prime factorization of $n!$ and appears twice (once as p and once as $2p$) in the prime factorization of $(2n)!$. Thus, $v_p = v_p((2n)!) - 2v_p(n!) = 2 - 2 \cdot 1 = 0$. Note since $n \geq 3$ we have $p > 2n/3 \geq 2$ so $p \neq 2$. So our previous counts are correct—in particular p appears only once in $2p$ and not twice as it would've if $p = 2$.

(b) If $p > \sqrt{2n}$ then $p^2 > 2n$ so in $(*)$ all the terms where $k > 1$ are zero, leaving us with $\lfloor 2n/p \rfloor - 2 \lfloor n/p \rfloor$. This is either 0 or 1. (See exercise below.)

(c) Since $n \geq 3$, $n \geq \sqrt{2n}$. So by part (b), we have $v_p \leq 1$. On the other hand, $p \mid (2n)!$ since $p \leq 2n$ and $p \nmid n!$ since $p > n$. So p divides $\binom{2n}{n}$. Thus, $v_p > 0$ and so $v_p = 1$.

(d) Let l be such that $p^l \leq 2n < p^{l+1}$. Then all the summands in $(*)$ with $k > l$ vanish, leaving us with l summands each of which is ≤ 1 (by the exercise below). Thus, $v_p \leq l$. So $p^{v_p} \leq p^l \leq 2n$. ■

Exercise 7.13 Prove that the value of $\lfloor 2x \rfloor - 2 \lfloor x \rfloor$ is either 0 or 1.

Our proof will involve breaking up

$$\binom{2n}{n} = \prod_{p \leq 2n} p^{v_p}$$

into pieces and then bounding each piece. The next lemma will be useful for this.

Lemma 7.14 For all $k \in \mathbb{Z}_{>0}$,

$$\prod_{p \leq k} p \leq 4^k,$$

where the product runs over all prime numbers $p \leq k$.

Proof: We'll prove this by induction on k. The cases $k = 1, 2, 3$ are clear by inspection. So now assume that $k \geq 4$ and that the result is true for all integers $< k$.

If k is even then k isn't prime and so $\prod_{p \leq k} p = \prod_{p \leq k-1} p \leq 4^{k-1} \leq 4^k$ by the inductive hypothesis.

If $k = 2m + 1$ is odd then $\binom{2m+1}{m} = \frac{(2m+1)!}{m!(m+1)!}$ is divisible by every prime p satisfying $m + 2 \leq p \leq 2m + 1$. Thus,

$$\prod_{p \leq k} p = \prod_{p \leq m+1} p \prod_{k=m+2}^{2m+1} p \leq 4^{m+1} \binom{2m+1}{m}, \qquad (**)$$

where the first bound is given by the inductive hypothesis. Now, using the binomial theorem, we have

$$2\binom{2m+1}{m} = \binom{2m+1}{m} + \binom{2m+1}{m+1} \leq \sum_{i=0}^{2m+1} \binom{2m+1}{i} = (1+1)^{2m+1} = 2^{2m+1}.$$

So $\binom{2m+1}{m} \leq 2^{2m} = 4^m$. By combining this with $(**)$, we complete the proof. ■

Proof of Theorem 7.11 (Bertrand's Postulate): Assume for the moment that $n \geq 600$ and that the statement is false: there is no prime between n and $2n$. Then, since $\sqrt{2n} \leq \frac{2n}{3}$, Lemma 7.12 (parts (a), (b) and (c)) implies that all the prime divisors p of $\binom{2n}{n}$ satisfy $p \leq \frac{2n}{3}$. Thus,

$$\binom{2n}{n} = \prod_{p \leq 2n} p^{v_p} = \prod_{p \leq \sqrt{2n}} p^{v_p} \prod_{\sqrt{2n} < p \leq 2n/3} p \leq (2n)^{\sqrt{2n}}\, 4^{2n/3},$$

where we used Lemma 7.12 part (d) and Lemma 7.14 to bound the first and second products, respectively.

On the other hand, the middle binomial coefficient $\binom{2n}{n}$ is the largest term in the binomial expansion

$$4^n = 2^{2n} = (1+1)^{2n} = \binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{2n}.$$

Since there are $2n+1$ terms in this sum, and all are $\leq \binom{2n}{n}$, we deduce that

$$4^n \leq (2n+1)\binom{2n}{n}.$$

Combining this with our upper bound on $\binom{2n}{n}$, we arrive at

$$\frac{4^n}{2n+1} \leq (2n)^{\sqrt{2n}} 4^{2n/3}.$$

Since $2n+1 < (2n)^2$ for $n \geq 2$, we can re-write the above as

$$2^{2n/3} < (2n)^{\sqrt{2n}+2}.$$

Using calculus, you can show that this inequality is false for all $n \geq 600$. This contradiction proves Bertrand's postulate for all $n \geq 600$.

The truth of the postulate for $n \leq 600$ can be established by noting that the in the sequence of primes

$$2,\ 3,\ 5,\ 7,\ 13,\ 23,\ 43,\ 83,\ 163,\ 317,\ 631$$

each prime is less than twice the previous one. ■

Lecture 7 Problems

7.1. Use the Prime Number Theorem to estimate the number of 100-digit primes.

7.2. Use the Prime Number Theorem to prove that there are arbitrarily large gaps between consecutive primes. (Note that we'd already proved this in Example 6.2.) [**Hint:** Suppose the largest gap between consecutive primes is g, in the sense that there cannot be more than g consecutive composite integers . Deduce that $\pi(x) \geq x/g$.]

7.3. Prove that $\lim_{n\to\infty} \frac{p_{n+1}}{p_n} = 1$.

7.4. Find all solutions to the Diophantine equation $x! = y^2$ with $x, y \in \mathbb{Z}_{>0}$. [**Hint:** Bertrand's postulate.]

Lecture 8 Congruence Modulo n

One of the complicating features of $\mathbb{Z}$ is that there are infinitely many integers. However, we can sometimes get away with condensing $\mathbb{Z}$ into a finite number system. The idea is to group together numbers that share the same remainder after division by a fixed integer n. We saw an instance of this when we considered the Diophantine equation

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

We proved that the sum of two squares can only leave a remainder of 0, 1 or 2 after division by 4. So if m leaves a remainder of 3, then we can conclude that the above equation has no integer solutions. We're going to put this type of argument into a general framework.

Here is the key definition.

Definition 8.1

Congruent Modulo n, $a \equiv b \pmod{n}$

Let $n \in \mathbb{Z}_{>0}$ be fixed. We say that $a, b \in \mathbb{Z}$ are **congruent modulo** n if they leave the same remainder after division by n. We denote this by writing

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

If a and b are not congruent modulo n, we write $a \not\equiv b \pmod{n}$.

Theorem 2.3 (The Remainder Theorem) implies that

$$a \equiv b \pmod{n} \quad \text{if and only if} \quad n \mid a - b.$$

So, for example:

- $7 \equiv 11 \pmod{4}$ since $4 \mid 7 - 11$;
- $100 \equiv 0 \pmod{25}$ since $25 \mid 100 - 25$; and
- $-16 \equiv 8 \pmod{6}$ since $6 \mid -16 - 8$,

Before proceeding, let me just mention that there are familiar examples of congruence modulo n in everyday life. Analogue clocks operate modulo 12 (e.g. 14 o'clock is the same as 2 o'clock, and $14 \equiv 2 \pmod{12}$); digital clocks operate modulo 24; weekly calendars operate modulo 7 (e.g. if Day 1 is a Monday then Day 8 is also a Monday, and $8 \equiv 1 \pmod{7}$) and yearly ones modulo 12.

The notation $\equiv$ for congruence, which is due to Gauss[9], is meant to signify that congruence behaves in many ways like regular equality $=$.

Proposition 8.2

Let $n \in \mathbb{Z}_{>0}$ and $a, b, c \in \mathbb{Z}$. Then:

(a) Congruence is **symmetric**: $a \equiv a \pmod{n}$.

(b) Congruence is **reflexive**: If $a \equiv b \pmod{n}$ then $b \equiv a \pmod{n}$.

(c) Congruence is **transitive**: If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ then $a \equiv c \pmod{n}$.

[9] C.F. Gauss (1777–1855) was one of the greatest mathematicians of all time. His book *Disquisitiones Arithmeticae* revolutionized number theory and shaped it into what it is today. Most of the main results of PMATH 340 were first clearly stated and proved in the *Disquisitiones*.

Proof: Exercise. ■

Exercise 8.3 Prove Proposition 8.2.

A relationship that is symmetric, reflexive and transitive is said to be an **equivalence relation**. So congruence modulo n is an equivalence relation.[10] In mathematics, when you find yourself in possession of an equivalence relation on a set of objects, the natural thing to do is to partition your set into equivalence classes where each class consists of objects that are equivalent to each other, and no two objects from different classes are equivalent.

Definition 8.4

Congruence Class, Residue Class, Integers Modulo n, $\mathbb{Z}/n\mathbb{Z}$

Let $n \in \mathbb{Z}_{>0}$. The **congruence class** (or **residue class**) **of** $a \in \mathbb{Z}$ **modulo** n is the set

$$[a]_n := \{b \in \mathbb{Z} \colon b \equiv a \pmod{n}\}$$

of all integers congruent to a modulo n. When n is clear from the context, we will drop it from the notation and write $[a]$ instead of $[a]_n$.

The set of all congruence classes modulo n is denoted by $\mathbb{Z}/n\mathbb{Z}$ and is called the set of **integers modulo** n.

For example,

$$\begin{aligned}[1]_3 &= \{b \in \mathbb{Z} \colon b \equiv 1 \pmod{3}\} \\ &= \{b \in \mathbb{Z} \colon b = 3q + 1 \text{ for some } q \in \mathbb{Z}\} \\ &= \{\ldots, -5, -2, 1, 4, 7, \ldots\}.\end{aligned}$$

Note that if $a \equiv b \pmod{n}$, then $[a]_n = [b]_n$ (and conversely). So, for instance,

$$[1]_3 = [4]_3 = [7]_3 = [-2]_3 = \cdots.$$

Example 8.5 Let $n = 2$. The congruence class of 1 modulo 2 consists of all odd numbers, since any two odd integers are congruent modulo 2:

$$[1] = \{\pm 1, \pm 3, \ldots\}.$$

Likewise, since any two even numbers are congruent modulo 2, we have

$$[0] = \{0, \pm 2, \pm 4, \ldots\}.$$

There are no other congruence classes, so

$$\mathbb{Z}/2\mathbb{Z} = \{[0], [1]\}$$

consists of two elements. In some sense we have condensed the infinite set $\mathbb{Z}$ into the finite set $\mathbb{Z}/2\mathbb{Z}$ by classifying integers as even or odd.

[10] Can you think of other equivalence relations you've seen in your studies so far?

Example 8.6 Let's look at congruence classes modulo $n = 4$. There will be one congruence class for each possible remainder after division by 4. For instance, [0] consists of all multiples of 4, [1] consists of all integers that leave a remainder of 1 after division by 4, and so on. Thus, there are 4 congruence classes in total, and we have

$$\mathbb{Z}/4\mathbb{Z} = \{[0], [1], [2], [3]\}.$$

The previous examples can be generalized as follows.

Proposition 8.7 Let $n \in \mathbb{Z}_{>0}$. Then

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

In particular, $\mathbb{Z}/n\mathbb{Z}$ is finite set of size n.

Proof: By the remainder theorem, every integer is congruent to *exactly one* of 0, 1, ... or $n-1$ modulo n. This shows that $[0]_n, [1]_n, \ldots, [n-1]_n$ give all possible congruence classes and that they are all distinct. This proves the proposition. ■

It's important to keep in mind that in $\mathbb{Z}/n\mathbb{Z}$ what matters are the equivalence classes and not the integers chosen to represent the equivalence classes. For instance, we could equally well represent $\mathbb{Z}/4\mathbb{Z}$ by

$$\mathbb{Z}/4\mathbb{Z} = \{[-8]_4, [5]_4, [18]_4, [-1]_4\}.$$

This is the same set as $\{[0]_4, [1]_4, [2]_4, [3]_4\}$ since

$$[-8]_4 = [0]_4, \quad [5]_4 = [1]_4, \quad [18]_4 = [2]_4 \quad \text{and} \quad [-1]_4 = [3]_4.$$

This prompts the following definition.

Definition 8.8 **Complete Set of Representatives**

Let $n \in \mathbb{Z}_{>0}$. An integer b in the congruence class $[a]_n$ is said to be a **representative** of the congruence class.

A **complete set of representatives modulo** n is a set of n distinct integers $\{a_1, \ldots, a_n\}$ such that every integer is congruent modulo n to exactly one a_i.

A representative of $[a]_n$ is any integer b such that $b \equiv a \pmod{n}$. For instance, 1, 5 and -7 are representatives of $[1]_4$ but 6 isn't. The terminology reflects the fact that

$$[1]_4 = [5]_4 = [-7]_4$$

so any one of these integers can be used to *represent* the congruence class.

The sets $\{0, 1, 2, 3\}$ and $\{-8, 5, 18, -1\}$ are complete sets of representatives modulo 4. Proposition 8.7 asserts that $\{0, 1, \ldots, n-1\}$ is a complete set of representatives modulo n.

Exercise 8.9 Find a complete set of representatives modulo 5 that consists of even integers.

The upshot of all this is that we now have a mechanism for passing from the infinite set $\mathbb{Z}$ to the finite set $\mathbb{Z}/n\mathbb{Z}$ and that further every integer has a representative in $\mathbb{Z}/n\mathbb{Z}$. What's more, this passage to $\mathbb{Z}/n\mathbb{Z}$ respects addition, subtraction and multiplication. (Division is a bit more tricky.)

Proposition 8.10 If $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, then

$$a \pm b \equiv c \pm d \pmod{n} \quad \text{and} \quad ab \equiv cd \pmod{n}.$$

Proof: I'll only prove that $a + b \equiv c + d \pmod{n}$; the others are similarly handled. We must show that $n \mid (a + b) - (c + d)$. Since $n \mid (a - c)$ and $n \mid (b - d)$, it follows that n divides $(a - c) + (b - d) = (a + b) - (c + d)$, as desired. ■

This proposition allows us to define $+, -$ and $\cdot$ on the set $\mathbb{Z}/n\mathbb{Z}$ by

$$[a]_n \pm [b]_n = [a \pm b]_n \quad \text{and} \quad [a]_n \cdot [b]_n = [ab]_n.$$

The subtlety here is that since $[a]_n$ is not uniquely determined by a, we have to be careful whenever we use $[a]_n$ in a formula. The formula should hold regardless of what representative for $[a]_n$ we choose in place of a.

For example, suppose we want to add $[1]_4$ and $[2]_4$. The above formula says this is equal to

$$[1]_4 + [2]_4 = [1 + 2]_4 = [3]_4.$$

However, since $[1]_4 = [13]_4$ and $[2]_4 = [-2]_4$, applying the same formula to these representatives instead gives

$$[1]_4 + [2]_4 = [13]_4 + [-2]_4 = [13 - 2]_4 = [11]_4.$$

The apparent contradiction disappears since $[3]_4 = [11]_4$. This is what Proposition 8.10 guarantees: it says that $[a]_n + [b]_n = [a + b]_n$ is *well-defined* independently of the choice of representatives.

Exercise 8.11 Convince yourself of this. Suppose that $[a]_n = [c]_n$ and $[b]_n = [d]_n$. Prove that $[a]_n + [b]_n = [c]_n + [d]_n$.

To see that this is not something to be taken for granted, consider the following example.

Example 8.12 Suppose we were to define exponentiation in $\mathbb{Z}/n\mathbb{Z}$ via the formula

$$[a]^{[b]} = [a^b].$$

There's an immediate problem here: if b is a negative integer then $[a^b]_n$ would be undefined since a^b need not be an integer. So let's restrict this definition to representatives b that are positive integers.

This is still problematic. For instance, let $n = 4$ and consider

$$[2]^{[5]} = [2^5] = [32] = [0].$$

On the other hand, since $[5] = [1]$, $[2]^{[5]}$ should also be equal to

$$[2]^{[1]} = [2^1] = [2].$$

However, $[2]_4 \neq [0]_4$, so we have reached a contradiction!

The conclusion is that our proposed formula for $[a]^{[b]}$ is *not well-defined*.

REMARK (Notation: $\mathbb{Z}/n\mathbb{Z}$ vs. $\mathbb{Z}_n$)

In other sources (e.g. MATH 135), the set of integers modulo n is denoted by $\mathbb{Z}_n$. The notation $\mathbb{Z}/n\mathbb{Z}$ is more common in number theory for two reasons:

1. It's less ambiguous. ($\mathbb{Z}_p$ is used by number theorists to denote the *p-adic integers*.)
2. It reflects the ring-theoretic construction of $\mathbb{Z}/n\mathbb{Z}$ as the quotient of $\mathbb{Z}$ by the ideal $n\mathbb{Z}$. Because of this, the notation $\mathbb{Z}/n\mathbb{Z}$ generalizes to other contexts more consistently.

That said, you're welcome to use $\mathbb{Z}_n$ if you want.

Lecture 8 Problems

8.1. Prove carefully the assertion made in the lecture that

$$a \equiv b \pmod{n} \quad \text{if and only if} \quad n \mid a - b.$$

8.2. If tomorrow is a Tuesday, what day of the week is it 365 days from now?

8.3. (a) Show that $S = \{3, 6, 17, -5, 14, 34\}$ is a complete set of representatives modulo 6.

(b) Evaluate the following in $\mathbb{Z}/6\mathbb{Z}$. Express your final answer as a congruence class $[a]_6$ with $a \in S$ from part (a).

(i) $[3]_6 + [15]_6$.

(ii) $[1]_6 - [5]_6$.

(iii) $[4]_6 \cdot [-4]_6$.

8.4. Either find a complete set of representatives modulo 4 consisting of perfect squares or prove that no such set can exist.

Lecture 9 Modular Arithmetic

In the previous lecture, we introduced the set $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n. This is a finite set of size n consisting of all congruence classes modulo n. We can list the elements of $\mathbb{Z}/n\mathbb{Z}$ by first choosing a complete set of representatives modulo n. A convenient choice is the set of remainders after division by n, which gives us

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

We learned that we can do arithmetic in $\mathbb{Z}/n\mathbb{Z}$ with congruence classes:

$$\begin{aligned} [a] + [b] &= [a+b] \\ [a] - [b] &= [a-b] \\ [a][b] &= [a][b]. \end{aligned}$$

Division is also possible in some cases, but we will postpone its discussion to later.

These operations obey all the familiar rules of arithmetic, such as

$$\begin{aligned} [a] + [b] &= [b] + [a] \\ [a] + ([b] + [c]) &= ([a] + [b]) + [c] \\ [a]([b] + [c]) &= [ab + ac] \\ [a]^k &= [a^k] \text{ where } k \in \mathbb{Z}_{>0}. \end{aligned}$$

(**Warning:** In the last one, k is a positive integer and not a congruence class. You cannot replace it with k' even if $k \equiv k' \pmod{n}$. See Example 8.12.)

Thus, we've created a new number system. The way to think about it is like a "condensed" version of $\mathbb{Z}$ where each integer a is replaced by a representative modulo n. This can be very convenient when the representatives are chosen in a clever way, since in $\mathbb{Z}/n\mathbb{Z}$ we can replace every instance of a with any a' such that $a \equiv a' \pmod{n}$.

Example 9.1 To calculate $33 \cdot 106 + 63^3$ modulo 4 (by "calculate" let's agree to mean: find the smallest non-negative integer congruent to this one), we first observe that

$$[33] = [1], \quad [106] = [2] \quad \text{and} \quad [63] = [-1].$$

Thus,

$$33 \cdot 106 + 63^3 \equiv 1 \cdot 2 + (-1)^3 \equiv 2 - 1 = 1 \pmod{4}.$$

We could have also done:

$$33 \cdot 106 + 63^3 \equiv (-3)(2) + 3^3 \equiv -6 + 27 \equiv 21 \equiv 1 \pmod{4}.$$

You're free to make whatever replacements modulo 4 you like, though some will be more efficient than others.

Exercise 9.2 Calculate $1! + 2! + 3! + \cdots + 100!$ modulo 5.

The next examples illustrate how modular arithmetic can help prove results concerning arithmetic in $\mathbb{Z}$.

Example 9.3 In the problem set for Lecture 2, you were asked to prove that $3 \mid a^3 - a$ for all $a \in \mathbb{Z}$. Use modular arithmetic to prove the stronger statement that $6 \mid a^3 - a$ for all $a \in \mathbb{Z}$.

Solution: What we want to prove here is that

$$a^3 - a \equiv 0 \pmod 6$$

for all $a \in \mathbb{Z}$. Since a is congruent to one of $0, 1, \ldots, 5$, we just have to plug these in one at a time and confirm that the result holds. Let's do it:

$$\begin{aligned}
0^3 - 0 &\equiv 0 \pmod 6 \\
1^3 - 1 &\equiv 0 \pmod 6 \\
2^3 - 2 \equiv 6 &\equiv 0 \pmod 6 \\
3^3 - 3 \equiv 24 &\equiv 0 \pmod 6 \\
4^3 - 4 \equiv (-2)^3 - (-2) &\equiv 0 \pmod 6 \\
5^3 - 5 \equiv (-1)^3 - (-1) &\equiv 0 \pmod 6
\end{aligned}$$

This proves the result! We've managed to turn our problem into a finite and doable (if slightly tedious) computational task.

(By the way, we'll be able to give a quicker proof of this particular result very soon.)

Exercise 9.4 Prove that $30 \mid a^5 - a$ for all $a \in \mathbb{Z}$.

Example 9.5 **(A test for divisibility by 9)**

Let's prove that $a \in \mathbb{Z}_{>0}$ is divisible by 9 if and only if the sum of its decimal digits is divisible by 9. (For example, 378 is divisible by 9 since 3+7+8=18 is.)

If a is expressed in decimal form as $a_k \cdots a_1 a_0$, where a_0 is the units digits, a_1 is the tens digit, and so on, then that means

$$a = a_k 10^k + \cdots + a_1 10 + a_0.$$

(For example, $378 = 3 \cdot 10^2 + 7 \cdot 10 + 8$.) Since $10^k \equiv 1^k \equiv 1 \pmod 9$ for all $k \in \mathbb{Z}_{>0}$, it follows that

$$a \equiv a_k + \cdots + a_1 + a_0 \pmod 9.$$

Thus, a is congruent to the sum of its digits modulo 9. Since an integer is divisible by 9 if and only if it is congruent to 0 modulo 9, the proof is complete.

Exercise 9.6 **(A test for divisibility by 11)**

Prove that $a \in \mathbb{Z}_{>0}$ is divisible by 11 if and only if the alternating sum of its decimal digits is divisible by 11. (For example, 4818 is divisible by 11 since $4 - 8 + 1 - 8 = -11$ is.)

Modular arithmetic can sometimes be used to show that Diophantine equations have no integer solutions. The idea is that if $f(x_1, x_2, \ldots, x_n)$ is a polynomial with integer coefficients, and if the equation

$$f(x_1, x_2, \ldots, x_n) = 0$$

has an integer solution, say $(x_1, \ldots, x_n) = (a_1, \ldots, a_n)$, then

$$f(a_1, \ldots, a_n) \equiv 0 \pmod{n}$$

for all $n \in \mathbb{Z}_{>0}$, since $f(a_1, \ldots, a_n)$ is in fact equal to 0. So if we can find an n for which the congruence

$$f(x_1, \ldots, x_n) \equiv 0 \pmod{n}$$

doesn't have any solutions, then that means the original Diophantine equation doesn't have any solutions either. (**Warning:** The converse is false. If $f(x_1, \ldots, x_n) = 0$ has a solution modulo n, that doesn't mean it has a solution in $\mathbb{Z}$. See the Remark on page 53.) If n is small, it's a finite and manageable task to check whether there are no solutions modulo n.

Example 9.7 Show that the Diophantine equation

$$x^2 - 5y^2 = 13$$

does not have any integer solutions.

Solution: Let's consider the equation modulo 5. It reduces to

$$x^2 \equiv 3 \pmod{5}.$$

There are only finitely many possibilities for x modulo 5: 0, 1, 2, 3, 4. In turn, these give $x^2 \equiv 0,\ 1,\ 4,\ 4,\ 1 \pmod{5}$. So x^2 is never congruent to 3 modulo 5.

Thus, the equation $x^2 - 5y^2 = 13$ has no solutions in the integers modulo 5, and therefore no solutions in the integers.

In the previous example, it was perhaps obvious to try reducing the equation modulo $n = 5$. Sometimes you have to be creative in choosing n.

Example 9.8 Show that the Diophantine equation

$$x^2 + y^2 + z^2 = 7$$

does not have any integer solutions.

Solution: Let's consider the equation modulo 8. The key observation is that squares are congruent to 0, 1 or 4 modulo 8:

$a \pmod 8$	0	1	2	3	4	5	6	7
$a^2 \pmod 8$	0	1	4	1	0	1	4	1

So the sum of three squares can only be congruent to 0, 1, 2, 3, 4, 5 or 6 modulo 8.

Thus, the congruence

$$x^2 + y^2 + z^2 \equiv 7 \pmod 8$$

does not have any solutions, and so the Diophantine equation

$$x^2 + y^2 + z^2 = 7$$

does not have any integer solutions. Our proof actually shows that the Diophantine equation $x^2 + y^2 + z^2 = m$ has no integer solutions whenever $m \equiv 7 \pmod 8$.

Example 9.9

Show that the Diophantine equation

$$x^3 + 117y^3 = 5$$

does not have any integer solutions.

Solution: The natural thing to try here is to reduce modulo 117, in which case the equation becomes

$$x^3 \equiv 5 \pmod{117}.$$

We *could* try to check all 117 congruence classes by hand (or ask a computer to do this for us!). We would find there are none.

It would be easier to notice that $9 \mid 117$ (why does 9 divide 117?), so let's reduce modulo 9:

$$x^3 \equiv 5 \pmod 9.$$

You can now quickly check that cubes are congruent modulo 9 to 0 or ± 1. Thus, there are no solutions to $x^3 + 117y^3 = 5$ modulo 9, and therefore no solutions in the integers.

This equation is part of an amusing story. It appeared in a paper where the authors used some fairly complicated algebraic number theory to show that it has no integer solutions. Shortly afterwards, another author published the quick mod 9 solution described above.

REMARK (The Local-to-Global Principle)

We've observed that any integer solution to the Diophantine equation $f(x_1, \ldots, x_n) = 0$ will be a solution modulo n for every n.

Solutions in $\mathbb{Z}$ are referred to as **global solutions**, while solutions in $\mathbb{Z}/n\mathbb{Z}$ are referred to as **local solutions**. So our observation can be rephrased as: *a global solution is a local solution.*

Conversely, not every local solution comes from a global solution. For instance, $x^2+y^2 = -1$ clearly has no solutions in $\mathbb{Z}$ (why?) but it has the "local" solution $(x, y) = (1, 1)$ modulo 3. However, there is no solution modulo 4 as you can check (in fact, you proved this in a previous lecture). So the lack of a local solution modulo 4 can perhaps explain why there is no global solution.

A natural question, then, is the following. If a Diophantine equation has solutions modulo n for all n, must it have a solution in $\mathbb{Z}$? In more provocative language: if there are local solutions locally everywhere, must there exist a global solution? Ponder this for a while.

Lecture 9 Problems

9.1. Let $f(x)$ be a polynomial with integer coefficients. Prove that if $a \equiv b \pmod{n}$ then $f(a) \equiv f(b) \pmod{n}$.

9.2. Prove or disprove: If $[a]_n[b]_n = [0]_n$ then either $[a]_n = [0]_n$ or $[b]_n = [0]_n$.

▶ 9.3. Let p be prime. Prove that $(a+b)^p \equiv a^p + b^p \pmod{p}$ for all $a, b \in \mathbb{Z}$. [Hint: There is helpful problem at the end of Lecture 5.]

9.4. Prove that the product of $k \geq 1$ consecutive integers is divisible by k. (A stronger result is possible: the product will be divisible by $k!$ not just by k.)

9.5. Show that the Diophantine equation $x^2 - 11 = 12y^5$ does not have any integer solutions.

9.6. Show that the Diophantine equation $x^3 + y^3 + z^3 = 22$ does not have any integer solutions.

9.7. (a) Show that $1! + 2! + \cdots + 100!$ is not a perfect square.

(b) Find all solutions to the Diophantine equation $1! + 2! + \cdots + x! = y^2$ with $x, y \in \mathbb{Z}_{>0}$.

Lecture 10 Division Modulo n

Let $[a]$ and $[b]$ be congruence classes modulo n. We would like (if possible) to define their quotient $[x] := [b]/[a]$ to be a congruence class $[x]$ modulo n too. For this to behave as expected, we would need to have

$$[a][x] = [b], \quad \text{or equivalently,} \quad ax \equiv b \pmod{n}.$$

Thus, we're led to consider the **linear congruence** $ax \equiv b \pmod{n}$. However, this congruence need not have any solutions; moreover, when it does have solutions, they need not be unique! These are difficulties that need to be addressed if we want to be able to divide by b modulo n.

Example 10.1

The linear congruence

$$2x \equiv 1 \pmod{4}$$

has no solutions, as can be confirmed by substituting $x = 0,\ 1,\ 2$ and 3 into the left-side.

Alternatively, this congruence is equivalent to the Diophantine equation

$$2x = 1 + 4y \iff 2x - 4y = 1.$$

Since $\gcd(2, 4) \nmid 1$, it follows from Theorem 4.6 that there are no integer solutions to this Diophantine equation, and so there are no solutions to the congruence.

In any case, we conclude that $[1]/[2]$ is meaningless modulo 4.

Example 10.2

Consider the linear congruence

$$2x \equiv 2 \pmod{4}.$$

By inspection, this has two solutions given by $x \equiv 1,\ 3 \pmod{4}$. If we had naively tried to "cancel the 2s" we would have missed the second solution.

Less obvious is the congruence

$$4x \equiv 2 \pmod{6}.$$

Now there is no obvious cancellation, but there are still two solutions: $x \equiv 2,\ 5 \pmod{6}$.

The conclusion we draw from both of these congruences is that $[a]/[b]$ can sometimes be ambiguous. What do we mean by $[2]/[4]$ modulo 6? Is it $[2]$? Is it $[5]$? The answer is it is *neither*: because of this ambiguity, we leave it undefined.

The next theorem, which should remind you of Theorem 4.6, tells us how to deal with linear congruences.

Theorem 10.3 **(Solvability of Linear Congruences)**

Let $a, b, n \in \mathbb{Z}$ with $n \geq 1$, and consider the linear congruence

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

(a) The congruence $(*)$ has a solution modulo n if and only if $\gcd(a, n) \mid b$.

(b) If $x \equiv x_0$ is a particular solution to $(*)$, then the full set of solutions modulo n is given by the g distinct congruence classes

$$\left[x_0 + k\frac{n}{g}\right], \quad 0 \leq k \leq g-1,$$

where $g = \gcd(a, n)$.

Proof: The whole thing follows from Theorem 4.6. Indeed, the congruence $ax \equiv b \pmod{n}$ is equivalent to the linear Diophantine equation

$$ax = b + ny \iff ax - ny = b$$

which, according to 4.6(a), has a solution if and only if $\gcd(a, n) \mid n$. This proves part (a). Further, by 4.6(b), if (x_0, y_0) is a particular solution, then the general solution is given by

$$(x, y) = (x_0 + k(n/g)), y_0 + k(a/g)), \quad k \in \mathbb{Z}.$$

So the solutions for x modulo n are all of the form

$$x \equiv x_0 + k\frac{n}{g} \pmod{n}, \quad k \in \mathbb{Z}.$$

It remains to determine which of these are distinct modulo n. This is a good exercise for you to work out. ■

Exercise 10.4 Complete the proof of Theorem 10.3(b) by showing that the congruence classes

$$\left[x_0 + k\frac{n}{g}\right], \quad 0 \leq k \leq g-1,$$

are all distinct, and that further any congruence class $[x_0 + l(n/g)]$, where $l \in \mathbb{Z}$, is congruent to one of the above congruence classes.

One important takeaway from the proof of Theorem 10.3 is that the linear congruence $ax \equiv b \pmod{n}$ is equivalent to the linear Diophantine equation $ax - ny = b$. So we can find solutions to the congruence by solving this latter equation, which we know how to do (e.g. by inspection or by using the Euclidean algorithm). Let's illustrate.

Example 10.5 Solve the following linear congruences (if possible).

(a) $6x \equiv 4 \pmod{15}$.

(b) $6x \equiv 9 \pmod{15}$.

Solution:

(a) Since $\gcd(6, 15) \nmid 4$, this congruence has no solution.

(b) Since $\gcd(6, 15) \mid 9$, this congruence has a solution. We can find a solution by inspection rather quickly by noticing that $9 \equiv -6 \pmod{15}$, so the congruence is equivalent to

$$6x \equiv -6 \pmod{15}$$

where it's clear that $x \equiv -1 \pmod{15}$ is a solution. Since $g = \gcd(6, 15) = 3$, the full solution set modulo 15 is

$$\left\{\left[-1 + k\frac{15}{3}\right] : 0 \leq k \leq 2\right\} = \{[-1],\ [4],\ [9]\}.$$

Exercise 10.6 Solve the linear congruence $15x \equiv 25 \pmod{35}$.

Theorem 10.3(b) shows that if the linear congruence $ax \equiv b \pmod{n}$ has a solution, then it has exactly $g = \gcd(a, n)$ solutions modulo n. However, all of these solutions are congruent to each other modulo n/g! So we obtain the following result.

Corollary 10.7 The linear congruence

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

has a solution if and only if $g = \gcd(a, n) \mid b$, in which case there is a unique solution modulo n/g. Furthermore, this aforementioned solution satisfies the congruence

$$\frac{a}{g}x \equiv \frac{b}{g} \pmod{\frac{n}{g}}.$$

Proof: Only the final assertion still requires proof. It follows from the fact that a solution to the congruence gives a solution to the Diophantine equation $ax = b + ny$ and therefore, by dividing through by g, to the Diophantine equation $(a/g)x = (b/g) + (n/g)y$, hence to the congruence $(a/g)x \equiv (b/g) \pmod{n/g}$. (Note that $a/g,\ b/g$ and n/g are all integers.) ■

Example 10.8 Returning to the congruence $6x \equiv 9 \pmod{15}$, where $g = \gcd(6, 15) = 3$, the preceding corollary tells us that there is a unique solution modulo $15/3 = 5$, and that this solution satisfies

$$2x \equiv 3 \pmod{5}.$$

At this point we can confirm that our solutions $x \equiv -1,\ 4,\ 9 \pmod{15}$ all reduce to 4 (mod 5), and that $x \equiv 4$ satisfies the above congruence, as predicted by the corollary.

However, had we not already not determined the solutions modulo 15, we could have used this simplified congruence modulo 5 to find them. Let's see how this works in a separate example.

Example 10.9

Solve the congruence $12x \equiv 18 \pmod{27}$.

Solution: Since $g = \gcd(12, 27) = 3$ and 3 divides 18, we know that there is a unique solution to this congruence modulo $27/3 = 9$, and that this solution satisfies

$$\frac{12}{3}x \equiv \frac{18}{3} \pmod{\frac{27}{3}} \iff 4x \equiv 6 \pmod{9}.$$

This congruence is easy to solve by inspection since there are only 9 values of x to check. We end up discovering that $x \equiv 6 \pmod{9}$ is the desired solution.

Note that $x \equiv 6$ is guaranteed to be a solution to the original congruence modulo 27. So now we can obtain the full set of solutions modulo 27:

$$\left\{ \left[6 + \frac{27}{3}k \right] : 0 \leq k \leq 2 \right\} = \{[6],\ [15],\ [24]\}.$$

Exercise 10.10

Express your solution to the congruence $15x \equiv 25 \pmod{35}$ from Exercise 10.6 as a single congruence class modulo 7.

A special case of Corollary 10.7 deserves singling out: the congruence $ax \equiv 1 \pmod{n}$ has a solution if and only if $\gcd(a, n) = 1$, that is, if and only if a and n are coprime. In this case, there is a unique a solution modulo n, and we give it a name.

Definition 10.11
Inverse Modulo n

Let $n \in \mathbb{Z}_{>0}$. If $a \in \mathbb{Z}$ is coprime to n, then the unique congruence class $[x]_n$ satisfying $[a]_n[x]_n = [1]_n$ is called the **inverse of** $[a]$ **modulo** n. We denote it by $[a]_n^{-1}$.

If a is not coprime to n, then we say that $[a]_n^{-1}$ is undefined.

When n is clear from context, we will write $[a]^{-1}$ in place of $[a]_n^{-1}$.

One thing worth pointing out here is that if $[a]_n = [b]_n$ then a is coprime to n if and only if b is coprime to n. That is, being coprime to n is a feature of the congruence class $[a]_n$ that is independent of the choice of representative of the class.

Example 10.12

Since $\gcd(3, 10) = 1$, $[3]$ has an inverse modulo 10. To find it, we solve $3x \equiv 1 \pmod{10}$ to get that $[3]^{-1} = [7]$.

On the other hand, since $\gcd(2, 10) \neq 1$, $[2]$ has no inverse modulo 10.

Here is a table of the congruence classes modulo 10 and their inverses, when they exist. (A dash $-$ indicates that $[x]^{-1}$ is undefined.)

$[x]$	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[6]$	$[7]$	$[8]$	$[9]$
$[x]^{-1}$	$-$	$[1]$	$-$	$[7]$	$-$	$-$	$-$	$[3]$	$-$	$[9]$

Exercise 10.13 Construct a similar table for the congruence classes and their inverses modulo 12.

If we have $[a]_n^{-1}$, we can easily solve any linear congruence $ax \equiv b \pmod{n}$ by multiplying both sides by this inverse.

Example 10.14 To solve the congruence $3x \equiv 8 \pmod{10}$, let's multiply through by $[3]^{-1} = [7]$:

$$\begin{aligned} \underbrace{(7 \cdot 3)}_{\equiv 1} x &\equiv 7 \cdot 8 \pmod{10} \\ x &\equiv 56 \pmod{10} \\ x &\equiv 6 \pmod{10}. \end{aligned}$$

Of course, this method is only useful if you can determine $[a]_n^{-1}$ quickly (if it even exists). If you have to run the Euclidean algorithm to find $[a]_n^{-1}$ then you might as well just run it to solve $ax \equiv b \pmod{n}$ directly.

Lecture 10 Problems

10.1. Prove that if $[a]_n = [b]_n$ then $\gcd(a, n) = \gcd(b, n)$ and, in particular, a is coprime to n if and only if b is coprime to n.

10.2. Find all values of $n \in \mathbb{Z}_{>0}$ such that $[2]^{-1}$ exists in $\mathbb{Z}/n\mathbb{Z}$. In the cases where $[2]^{-1}$ exists, determine it explicitly as a congruence class in $\mathbb{Z}/n\mathbb{Z}$. (Your answer will depend on n.)

10.3. The **Chinese Remainder Theorem (CRT)** states that if $n_1, n_2 \ldots, n_k \in \mathbb{Z}_{>0}$ are pairwise coprime integers, then the system of linear congruences

$$\begin{aligned} x &\equiv a_1 \pmod{n_1} \\ x &\equiv a_2 \pmod{n_2} \\ &\vdots \\ x &\equiv a_k \pmod{n_k} \end{aligned}$$

has a unique solution modulo $n_1 n_2 \cdots n_k$.

(a) Prove the CRT in the case where $k = 2$ by writing $x = a_1 + n_1 y$ (with $y \in \mathbb{Z}$) and substituting this into the congruence $x \equiv a_2 \pmod{n_2}$. (You should obtain a linear congruence for y.)

(b) Prove the CRT in the case where $k \geq 2$ by using part (a). [**Hint:** First deal with the congruences modulo n_1 and n_2 to find a solution modulo $n_1 n_2$, then pair this off with the congruence modulo n_3, etc.]

10.4. Solve *Sunzi's problem* (the origin of the Chinese Remainder Theorem): Find all solutions to the system of linear congruences

$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7}. \end{aligned}$$

Lecture 11 The Group of Units Modulo n

Here's a summary of what we discovered last lecture. A congruence class $[a]_n \in \mathbb{Z}/n\mathbb{Z}$ will have a unique inverse modulo n if and only if a and n are coprime. This inverse is a congruence class $[b]_n \in \mathbb{Z}/n\mathbb{Z}$ that is characterized by the fact that

$$[a]_n[b]_n = [1]_n.$$

Since it's unique (if it exists), we can and will denote it unambiguously by $[a]_n^{-1}$.

For example, $[2]_5^{-1} = [3]_5$ since $[2]_5[3]_5 = [1]_5$. Likewise, $[3]_5^{-1} = [2]_5$. In general, to find $[a]_n^{-1}$, we must solve the congruence

$$ax \equiv 1 \pmod{n}.$$

Exercise 11.1 As a refresher, find the inverse of 7 modulo 11.

Here's some helpful terminology.

Definition 11.2

Invertible Modulo n, Inverse Modulo n, Units Modulo n, $(\mathbb{Z}/n\mathbb{Z})^\times$

Let $n \in \mathbb{Z}_{>0}$. A congruence class $[a]_n \in \mathbb{Z}/n\mathbb{Z}$ is called a **unit** (or said to be **invertible**) if there is a congruence class $[b]_n \in \mathbb{Z}/n\mathbb{Z}$ such that $[a]_n[b]_n = 1$. In this case, $[b]_n$ is uniquely determined by $[a]_n$ and is called the **inverse** of $[a]_n$ and denoted by $[a]_n^{-1}$.

The set of all units in $\mathbb{Z}/n\mathbb{Z}$ will be denoted by $(\mathbb{Z}/n\mathbb{Z})^\times$.

REMARK

The terminology for units stems from the fact that in $\mathbb{Z}$, the only integers a that have inverses $1/a$ in $\mathbb{Z}$ are the "units" ± 1. The notation $(\mathbb{Z}/n\mathbb{Z})^\times$ comes from ring theory, where $R^\times$ denotes the set of units in a ring R (and $\mathbb{Z}/n\mathbb{Z}$ is an example of a ring). The $\times$ is meant to signify that we are dealing with elements that have a *multiplicative* inverse.

For example,

$$\begin{aligned}(\mathbb{Z}/4\mathbb{Z})^\times &= \{[1],[3]\}\\ (\mathbb{Z}/5\mathbb{Z})^\times &= \{[1],[2],[3],[4]\}\\ (\mathbb{Z}/6\mathbb{Z})^\times &= \{[1],[5]\}\\ (\mathbb{Z}/8\mathbb{Z})^\times &= \{[1],[3],[5],[7]\}.\end{aligned}$$

Here are the basic properties of units modulo n.

Proposition 11.3 Let $n \in \mathbb{Z}_{>0}$.

(a) $[a]_n \in \mathbb{Z}/n\mathbb{Z}$ is a unit if and only if $\gcd(a, n) = 1$.

(b) $[1]_n$ a unit for all n.

(c) If $[a]_n, [b]_n \in \mathbb{Z}/n\mathbb{Z}$ are units, then so is $[a]_n[b]_n$.

(d) If $[a]_n \in \mathbb{Z}/n\mathbb{Z}$ is a unit, then $[a]_n^{-1}$ is a unit.

Proof: Part (a) was proved last lecture and is restated here for convenience. We can use it to immediately prove parts (b)–(d). However, let's give different proofs. We will use the fact that $[a]$ is a unit if and only if there is a $[b]$ such that $[a][b] = 1$.

For (b), since $[1][1] = [1]$, we see that $[1]$ is a unit with $[1]^{-1} = [1]$. Similarly, since $([a][b])([b]^{-1}[a]^{-1}) = [a][1][a]^{-1} = [1]$, we have that $[a][b]$ is a unit with inverse $[b]^{-1}[a]^{-1}$. Finally, since $[a]^{-1}[a] = [a][a]^{-1} = [1]$, it follows that $[a]^{-1}$ is a unit with inverse $[a]$. ■

Parts (b)–(d) of Proposition 11.3 say that the set $(\mathbb{Z}/n\mathbb{Z})^\times$ of units modulo n forms a *group* under multiplication. A group, much like a vector space, is an abstract mathematical object consisting of a set together with an operation that satisfies certain axioms. Here is the formal definition.

Definition 11.4
Group, Order

A **group** is a set G together with an operation $\star$ defined so that for all $g, h \in G$, $g \star h \in G$. Furthermore, the operation $\star$ must satisfy the following properties, known as the **group axioms**.

1. [Associativity] For all $g, h, k \in G$, we have $g \star (h \star k) = (g \star h) \star k$.
2. [Identity Element] There is a unique element $e \in G$ such that $e \star g = g \star e = g$ for all $g \in G$. This element e is called the **identity element** of G.
3. [Inverses] For all $g \in G$, there exists a unique element $h \in G$ such that $g \star h = h \star g = e$. Such an element h is called the **inverse** of g and is denoted by g^{-1}.

A group G is said to be **commutative** if $\star$ satisfies the following additional property.

4. [Commutativity] For all $g, h \in G$, $g \star h = h \star g$.

If the set G is finite, then the size of G is called the **order** of G and will be denoted by $|G|$ or $\#G$.

We won't delve too deeply in the theory of groups since there are separate courses for that. The above definition was introduced primarily to put things in perspective. One payoff is that some very simple ideas from group theory will allow us to give cleaner statements and proofs of several results down the line.

First, here are some examples of commutative groups:

- $G = \mathbb{Z}$ with $\star$ being the usual addition of integers. In this case, the associativity axioms reads: $a + (b + c) = (a + b) + c$ for all $a, b, c \in \mathbb{Z}$. The identity element $e \in \mathbb{Z}$ must satisfy $e + a = a + e = a$ for all $a \in \mathbb{Z}$ — so $e = 0$. Finally, the inverse of $a \in \mathbb{Z}$ is an integer $b \in \mathbb{Z}$ that satisfies $a + b = e = 0$. So the invese of a is $-a$.
- $G = \mathbb{Z}/n\mathbb{Z}$ is a group with $\star$ being addition of congruence classes. I'll let you confirm that the identity element is $[0]$ and the additive inverse of $[a]$ is $[-a]$.
- $G = (\mathbb{Z}/n\mathbb{Z})^\times$ is a group with $\star$ being multiplication of congruence classes. The identity element this time is $[1]$ and the inverse of $[a]$ is $[a]^{-1}$. (Look back at Proposition 11.3.)

For an example of a non-commutative group, we can take:

- $G = \{n \times n \text{ invertible matrices with entries in } \mathbb{R}\}$ with $\star$ being matrix multiplication. The identity element is the identity matrix I_n and the inverse of A is the matrix inverse A^{-1}. If $n \geq 2$, we can find invertible matrices A and B such that $AB \neq BA$ (prove this!), so G is not commutative in this case.

There are plenty (and I mean *plenty*) of other examples of groups, but the above should suffice for now. It is also useful to have non-examples.

Exercise 11.5 Show that neither $\mathbb{Z}$ nor $\mathbb{Z}/n\mathbb{Z}$ are groups if we take $\star$ to be multiplication instead of addition modulo n. Likewise, show that $(\mathbb{Z}/n\mathbb{Z})^\times$ is not a group if $\star$ is addition modulo n.

Going forward, whenever we view $\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$ as groups, we do so with the understanding that the group operation is addition. Likewise with $(\mathbb{Z}/n\mathbb{Z})^\times$ and multiplication.

One way $\mathbb{Z}/n\mathbb{Z}$ and $(\mathbb{Z}/n\mathbb{Z})^\times$ differ from $\mathbb{Z}$ is that they are finite groups. Finite groups are subject to Lagrange's theorem. A special case of this theorem is given below. To state it, we need the following notation. If G is a group, we define g^n for $g \in G$ and $n \in \mathbb{Z}_{>0}$ by

$$g^n = \underbrace{g \star g \star \cdots \star g}_{n \text{ times}}.$$

We also define $g^0 = e$. If the group operation is addition, then we write $ng = g + \cdots + g$ instead of g^n. If $n < 0$ is a negative integer, then we define g^n to be $(g^{-1})^{|n|}$ (and likewise $ng = |n|(-g)$, where $-g$ is the inverse of g). For example, $g^{-2} = (g^{-1})^2 = g^{-1} \star g^{-1}$.

Theorem 11.6 **(Lagrange's Theorem—Special Case)**

Let G be a finite *commutative* group of order m. Then $g^m = e$ for all $g \in G$.

Proof: Suppose that $G = \{g_1, \ldots, g_m\}$, where the g_i are the m distinct elements of G. Consider the set $S = \{g \star g_1, \ldots, g \star g_m\}$. Since $g \star g_i \in G$ for all i, we have that $S \subseteq G$. I claim that $S = G$. To prove this, it suffices to show that S contains m elements. To prove *that*, it suffices to show that $g \star g_i \neq g \star g_j$ if $i \neq j$. But this is easy:

$$\begin{aligned} \text{if } g \star g_i = g \star g_j \quad &\text{then} \quad g^{-1} \star (g \star g_i) = g^{-1} \star (g \star g_j) \\ &\text{hence} \quad (g^{-1} \star g) \star g_i = (g^{-1} \star g) \star g_j \text{ by associativity} \\ &\text{so } e \star g_i = e \star g_j \end{aligned}$$

and therefore $g_i = g_j$. So $i = j$ since the g_i's are distinct. So if $i \neq j$ then it must follow that $g \star g_i \neq g \star g_j$.

Thus, $S = G$. So if we multiply all the elements in S we get the same result as if we had multiplied all the elements in G:

$$(g \star g_1) \star (g \star g_2) \cdots (g \star g_m) = g_1 \star g_2 \star \cdots \star g_m.$$

Since G is commutative, we can re-order the left-side to obtain

$$g^m \star (g_1 \star g_2 \star \cdots \star g_m) = g_1 \star g_2 \star \cdots \star g_m.$$

Now $h = g_1 \star \cdots \star g_m$ is an element in G, so it has an inverse h^{-1}. Multiplying the above on the right by h^{-1}, we finally arrive at $g^m = e$, as desired. ■

REMARK

This version of Lagrange's theorem also holds for non-commutative groups, but the proof requires different ideas.

Let's end the lecture by noting that Lagrange's theorem is completely obvious if $G = \mathbb{Z}/n\mathbb{Z}$. In this case G has order n and so the theorem claims that

$$n[a] = [0] \text{ for all } [a] \in \mathbb{Z}/n\mathbb{Z}.$$

And indeed,

$$n[a] = [a] + \cdots + [a] = [na] = [0]$$

since $na \equiv 0 \pmod{n}$.

The theorem is much more interesting in the case of $G = (\mathbb{Z}/n\mathbb{Z})^\times$. We'll take this up in detail next time, but here's an example.

Example 11.7 Consider $G = (\mathbb{Z}/5\mathbb{Z})^\times = \{[1], [2], [3], [4]\}$, which has order 4. Lagrange's theorem asserts, in this case, that

$$[a]^4 = [1] \text{ for all } [a] \in (\mathbb{Z}/5\mathbb{Z})^\times.$$

Let's confirm:

$$\begin{aligned}
[1]^4 &= [1] \\
[2]^4 &= [16] = [1] \qquad (\text{since } 16 \equiv 1 \pmod{5}) \\
[3]^4 &= [81] = [1] \qquad (\text{since } 81 \equiv 1 \pmod{5}) \\
[4]^4 &= [256] = [1] \qquad (\text{since } 256 \equiv 1 \pmod{5}).
\end{aligned}$$

Lecture 11 Problems

11.1. Look back at your math courses from previous terms or any math courses that you're taking this term. Can you recognize if groups made an appearance in any of them?

11.2. Let G be the set of invertible 2×2 matrices with entries in $\mathbb{Z}/3\mathbb{Z}$.

(a) Prove that G is a group under matrix multiplication. (That is, if we define $A \star B = AB$, then $\star$ satisfies the group axioms.)

(b) Determine the order of G. [**Hint:** A square matrix with entries in $\mathbb{Z}/3\mathbb{Z}$ is invertible if and only if its columns are linearly independent.]

11.3. Let $S = \{a_1, \ldots, a_n\}$ be a complete set of representatives modulo n. Prove that if $[u]$ is a unit in $\mathbb{Z}/n\mathbb{Z}$, then $\{ua_1, \ldots, ua_n\}$ is a complete set of representatives modulo n.

11.4. Let G be a finite commutative group. Prove that $g^{-1} = g^{|G|-1}$ for all $g \in G$.

▶ 11.5. Let p and q be distinct primes. Determine the orders of:

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

(b) $(\mathbb{Z}/p^2\mathbb{Z})^\times$.

(c) $(\mathbb{Z}/(pq)\mathbb{Z})^\times$.

Lecture 12 The Theorems of Fermat and Euler

Last time we stated and proved Lagrange's theorem which asserts that if G is a finite commutative group of order m, then $g^m = e$ for all $g \in G$.

We would like to investigate what this looks like if $G = (\mathbb{Z}/n\mathbb{Z})^\times$ is the group of units modulo n. The first thing we have to figure out is: What is the order of G?

Definition 12.1
Euler's φ function

Let $n \in \mathbb{Z}_{>0}$. The order of $(\mathbb{Z}/n\mathbb{Z})^\times$ is denoted by $\varphi(n)$. The function $\varphi\colon \mathbb{Z}_{>0} \to \mathbb{Z}$ defined in this manner is called **Euler's φ function**.

For example:

- $\varphi(2) = 1$ since $(\mathbb{Z}/2\mathbb{Z})^\times = \{[1]\}$.
- $\varphi(5) = 4$ since $(\mathbb{Z}/5\mathbb{Z})^\times = \{[1], [2], [3], [4]\}$.
- $\varphi(6) = 2$ since $(\mathbb{Z}/6\mathbb{Z})^\times = \{[1], [5]\}$.

In general, since $\mathbb{Z}/n\mathbb{Z} = \{[0], [1], \ldots, [n]\}$ and $(\mathbb{Z}/n\mathbb{Z})^\times = \{[a] \in \mathbb{Z}/n\mathbb{Z}\colon \gcd(a, n) = 1\}$, we obtain the following alternative characterization of $\varphi(n)$.

Proposition 12.2

For $n \in \mathbb{Z}_{>0}$,

$$\varphi(n) = \#\{a \in \mathbb{Z}\colon 0 \leq a \leq n-1,\ \gcd(a, n) = 1\}.$$

In other words, $\varphi(n)$ is equal to the number of positive integers $\leq n-1$ that are coprime to n. A special case is worth singling out:

Corollary 12.3

If p is prime, then $\varphi(p) = p - 1$.

Proof: All of the $p-1$ integers 1, ..., and $p-1$ are coprime to p. ■

With this in hand, we obtain the following two theorems as immediate corollaries of Lagrange's theorem.

Theorem 12.4

(Euler's Theorem)

If $a \in \mathbb{Z}$ is coprime to n, then

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

Proof: If a is coprime to n, then $[a] \in (\mathbb{Z}/n\mathbb{Z})^\times$, and so $[a]^{\varphi(n)} = [1]$ by Lagrange. ■

Theorem 12.5 **(Fermat's Little Theorem)**

Let p be prime. For all $a \in \mathbb{Z}$, we have

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

If $p \nmid a$, then

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

Proof 1: If $p \mid a$ then $a \equiv 0$ and $a^p \equiv 0$, so the theorem holds. If $p \nmid a$, then the theorem is the case $n = p$ of Euler's theorem. ■

Let's give a completely different proof of Fermat's little theorem!

Proof 2: First off, the second statement follows from the first since if $p \nmid a$ then a is invertible modulo p, so we can divide the first congruence through by a to obtain the second congruence. So it suffices to prove the first congruence.

In the Lecture 9 problem set, you proved that $(a+b)^p \equiv a^p + b^p \pmod{p}$ for all $a, b \in \mathbb{Z}$. In particular, then, $(a+1)^p \equiv a^p + 1 \pmod{p}$. With this in hand, we can prove that $a^p \equiv a$ for all non-negative $a \in \mathbb{Z}$ by induction on a.

Indeed, the base case $a = 0$ is obvious. For the inductive step, assume that $a^p \equiv a$. Then $(a+1)^p \equiv a^p + 1 \equiv a + 1$. Done!

It remains to prove the theorem for negative $a \in \mathbb{Z}$. Exercise! ■

Exercise 12.6 Complete the second proof of Fermat's Little Theorem by showing that if the theorem holds for all non-negative integers, then it also holds for all negative integers. [**Hint:** If $a \in \mathbb{Z}$ is negative, apply the theorem to $-a > 0$.]

These two theorems of Fermat and Euler can help us compute powers modulo n. (As we'll soon see, calculating $a^m \pmod{n}$ has important applications to cryptography.)

Example 12.7 Find the last digit of 3^{1234}.

Solution:

We want to find the remainder of 3^{1234} modulo 10.

Since 3 is coprime to 10, Euler's theorem asserts that $3^{\varphi(10)} \equiv 1 \pmod{10}$. Note that $\varphi(10) = 4$, since there are 4 positive integers < 10 that are coprime to 10 (what are they?). Thus, $3^4 \equiv 1 \pmod{10}$.

Now, $1234 = 4 \cdot 308 + 2$. Therefore,

$$3^{1234} = 3^{4 \cdot 308 + 2} = (3^4)^{308} 3^2 \equiv 1^{617} \cdot 9 \equiv 9 \pmod{10}.$$

So the last digit of 3^{1234} is 9.

Example 12.8 Calculate 7^{365} modulo 24.

Solution: Since there are 8 positive integers < 24 that are coprime to 24, we have $\varphi(24) = 8$. Therefore, $7^8 \equiv 1 \pmod{24}$ by Euler's theorem.

Now, as $365 = 8 \cdot 45 + 5$, we find that

$$7^{365} = (7^8)^{45}7^5 = 1^{45} \cdot 7^5 \pmod{24}.$$

To calculate 7^5 modulo 24, we can proceed as follows. First,

$$7^2 = 49 \equiv 1 \pmod{24}$$

and then

$$7^4 = (7^2)^2 \equiv 1 \pmod{24}$$

so

$$7^5 = 7^4 \cdot 7 \equiv 1 \cdot 7 \equiv 7 \pmod{24}.$$

Thus, $7^{365} \equiv 7 \pmod{24}$.

I want to make two comments about the preceding examples:

1. We calculated $\varphi(n)$ by counting integers $< n$ coprime to n. This isn't very efficient when n is large. We'll come back to this problem later.
2. What we did when we calculated 7^5 modulo 24 can be generalized to give a relatively efficient algorithm for calculating $a^k \pmod{n}$ by repeated squaring. Let's explore this now.

Exponentiation via Repeated Squaring

The basic idea is as follows. To compute a^k for $k \in \mathbb{Z}_{>0}$, we begin by finding the **binary representation** of k:

$$k = 2^0a_0 + 2a_1 + 2^2a_2 + \cdots + 2^la_l, \text{ where } a_i \in \{0, 1\}.$$

This is analogous to how we write numbers in decimal form, except with powers of 2 instead of powers of 10. To find the binary representation of k, we simply repeatedly divide by 2 and use the remainder theorem:

$$k = 2q + r = 2(2q' + r') = 2^2q' + 2r' + r = \cdots.$$

At each step, the remainder is either 0 or 1, and the process eventually terminates because we're getting smaller and smaller positive integers.

Example 12.9 Find the binary representation of 97.

Solution: We have

$$\begin{aligned}
97 &= 2 \cdot 48 + 1 \\
&= 2(2 \cdot 24 + 0) + 1 \\
&= 2^2 \cdot 24 + 2 \cdot 0 + 1 \\
&= 2^2(2 \cdot 12 + 0) + 2 \cdot 0 + 1 \\
&= 2^3 \cdot 12 + 2^2 \cdot 0 + 2 \cdot 0 + 1 \\
&= 2^3 \cdot (2 \cdot 6 + 0) + 2^2 \cdot 0 + 2 \cdot 0 + 1 \\
&= 2^4 \cdot 6 + 2^3 \cdot 0 + 2^2 \cdot 0 + 2 \cdot 0 + 1 \\
&= 2^4(2 \cdot 3 + 0) + 2^3 \cdot 0 + 2^2 \cdot 0 + 2 \cdot 0 + 1 \\
&= 2^5 \cdot 3 + 2^4 \cdot 0 + 2^3 \cdot 0 + 2^2 \cdot 0 + 2 \cdot 0 + 1 \\
&= 2^5(2 \cdot 1 + 1) + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^6 + 2^5 + 2^4 \cdot 0 + 2^3 \cdot 0 + 2^2 \cdot 0 + 2 \cdot 0 + 1.
\end{aligned}$$

Exercise 12.10 Find the binary representation of 173.

Once we have $k = \sum_{i=0}^{l} 2^i a_i$ in binary, it follows that

$$a^k = a^{2^0 a_0} a^{2^1 a_1} a^{2^2 a_2} \cdots a^{2^l a_l}.$$

Since a_i is either 0 or 1, the problem now is reduced to calculating a^{2^r}. This we can perform recursively by squaring, since $a^{2^r} = (a^{2^{r-1}})^2$.

Example 12.11 Calculate $5^{97} \pmod{19}$.

Solution 1: Using the binary representation

$$97 = 2^6 + 2^5 + 1$$

(where the powers of 2 with 0 coefficient have been omitted), we have

$$5^{97} = 5^{2^6} 5^{2^5} 5.$$

Now,

$$5^2 \equiv 25 \equiv 6 \pmod{19}$$

hence

$$5^{2^2} = (5^2)^2 \equiv 6^2 \equiv 36 \equiv -2 \pmod{19}$$

hence

$$5^{2^3} = (5^{2^2})^2 \equiv (-2)^2 \equiv 4 \pmod{19}$$

hence

$$5^{2^4} = (5^{2^3})^2 \equiv 4^2 \equiv -3 \pmod{19}$$

hence

$$5^{2^5} = (5^{2^4})^2 \equiv (-3)^2 \equiv 9 \pmod{19}$$

hence

$$5^{2^6} = (5^{2^5})^2 \equiv 9^2 \equiv 81 \equiv 5 \pmod{19}.$$

So, finally,

$$5^{97} \equiv 5^{2^6} 5^{2^5} 5 \equiv 5 \cdot 9 \cdot 5 \equiv 5^2 \cdot 9 \equiv 54 \equiv 16 \pmod{19}.$$

Solution 2: Since 19 is prime, Fermat's Little Theorem tells us that $5^{18} \equiv 1 \pmod{19}$. Since $97 = 5 \cdot 18 + 7$, we have

$$5^{97} = 5^{5 \cdot 18 + 7} = (5^{18})^5 5^7 \equiv 1^5 5^7 \equiv 5^7 \pmod{19}.$$

Now we just have to compute 5^7 modulo 19. We can do this using the repeated squaring algorithm as in Solution 1. In binary, $7 = 1 + 2 + 2^2$, so

$$5^7 = 5 \cdot 5^2 \cdot (5^2)^2 \equiv 5 \cdot 6 \cdot 6^2 \equiv 5 \cdot 6 \cdot (-2) \equiv 16 \pmod{19}.$$

Exercise 12.12 Calculate $3^{155} \pmod{13}$.

Lecture 12 Problems

▶ 12.1. Show that $2^{340} \equiv 1 \pmod{341}$. [**Warning:** $341 = 11 \times 31$ is not prime.]

▶ 12.2. Let $a, n \in \mathbb{Z}_{>0}$ be coprime integers. Prove that if $k \equiv l \pmod{\varphi(n)}$ then $a^k \equiv a^l \pmod{n}$.

12.3. Let $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$, where p is a prime. Prove that $[a]^{-1} = [a]^{p-2}$.

12.4. Fermat's Little Theorem says that $a^{p-1} \equiv 1 \pmod{p}$ if a is a unit modulo p and $a^p \equiv a \pmod{p}$ for all a. Euler's theorem says $a^{\varphi(n)} \equiv 1 \pmod{n}$ if a is a unit modulo n. Prove/disprove: $a^{\varphi(n)+1} \equiv a \pmod{n}$ for all $a \in \mathbb{Z}$.

Lecture 13 Intro to Mathematical Cryptography

Note: This is the first of two lectures on cryptography. If you want to learn more, I highly recommend the book *An Introduction to Mathematical Cryptography* by Silverman, Pipher and Hoffstein.

Alice wants to send Bob a message m. They can only communicate through a public channel. Alice doesn't want any eavesdropper, such as her nemesis Eve, to know the content of m. So Alice wants to devise a system that allows her communicate secret messages through a public channel.[11]

To achieve this, Alice is looking for a way to **encrypt** her message m in such a way that only Bob knows how to **decrypt** the encrypted message back to m. Alice wants to devise a **cryptosystem**.

We're going to explore a few cryptosystems based on modular arithmetic. For this to make sense, we're going to assume that our messages are integers. This is okay because we can **encode** all text into numbers in a variety of ways. For example, we can simply set

$$A \leftrightarrow 00,\ B \leftrightarrow 01,\ C \leftrightarrow 02,\ \ldots,\ Z \leftrightarrow 25.$$

(We can also similarly deal with punctuation, etc.) Under this scheme, the message

`H E L L O`

is encoded as

07 04 11 11 14.

There are more sophisticated encoding schemes but we won't dwell on this. Let's look at a couple of basic cryptosystems.

Example 13.1 **(Shift ciphers)**

We will encrypt our message letter by letter. So we will assume that m is an integer between 0 and 25. (We can therefore pretend that m lives in $\mathbb{Z}/26\mathbb{Z}$.) For example, $m = 11$ represents the letter `L`. We will choose a secret encryption key k, which will also be an integer $k \in \{0, 1, \ldots, 25\}$. Then we define our encryption function by

$$e(m) \equiv m + k \pmod{26}.$$

This function shifts the alphabet to the right by k letters. For example, if $k = 14$, then the message

H	E	L	L	O
07	04	11	11	14

gets encrypted letter by letter into

21 18 25 25 02

[11] To make this a bit more relatable: Imagine that you are Alice, that Bob is an online vendor, and that m is your credit card information. The public channel here is the internet.

Notice that $e(14) \equiv 14 + 14 \equiv 28 \equiv 2 \pmod{26}$. So in shifting beyond the letter `Z`, we loop back and start again at `A`. If we now consult our encoding table $A \leftrightarrow 00$, $B \leftrightarrow 01$, ..., we see that our encrypted message is

`V S Z Z C`

The decryption function uses the same key k and is defined by

$$d(c) \equiv c - k \pmod{26}.$$

Note that

$$d(e(m)) \equiv d(m + k) \equiv (m + k) - k \equiv m \pmod{26}$$

so that the decryption function undoes the encryption on $e(m)$ and returns the original message m.

For instance, the letter `V` which is represented by 21 is decrypted into $d(21) = 21 - 14 = 7 \pmod{26}$. Since 07 represents the letter `H`, we obtain the first letter of our original message `HELLO`.

Exercise 13.2

Encrypt the message

`MODULAR ARITHMETIC`

using a shift cipher using the key $k = 7$.

Definition 13.3

Plaintext, Ciphertext

A message in original form (pre-encryption) is referred to as **plaintext**. After encryption, the result is referred to as **ciphertext**.

In the preceding example, `HELLO` is the plaintext and `VSZZO` is the ciphertext. We might also consider their encodings 07 04 11 11 14 and 21 18 25 25 02 as being the plaintext and ciphertext, respectively.

You might wonder about the difference between encoding and encryption. It is simply a matter of intent. When we encode, we assume everyone knows about the encoding process and how to reverse (decode) it. On the other hand, with encryption and decryption we are more secretive. Only Alice and Bob should be able to decrypt any encrypted ciphertexts.

The shift cipher is very easy to break. Eve can simply try each possible $k \in \{0, 1, \ldots, 25\}$ to find the secret key. This is called a *brute-force* or *exhaustive search* attack. Since there are only 26 possible keys in total, it is a very manageable task (especially for a computer) to go through them one by one.

More sophisticated attacks are possible. For instance, in English text, the most frequently occurring letters are `E`, `T`, `A`, `O` and `N` (listed in descending order of frequency). So if the shift cipher was used to encrypt a long piece of plaintext, then the frequency of the letters in the ciphertext can help Eve discover k.

Exercise 13.4 A shift cipher with secret key k produced the ciphertext

`VCUJMZ BPMWZG QA MFKMTTMVB.`

Discover the secret key k and use it to obtain the original plaintext message.

Let's take a look at another encryption scheme.

Example 13.5 **(Affine ciphers)**

Again, we will encrypt our plaintext message one letter at a time. This time, our key k will consist of two pieces: $k = (k_1, k_2)$ where $k_1, k_2 \in \{0, 1, \ldots, 25\}$ as in the shift cipher, but now we insist that $\gcd(k_1, 26) = 1$ so that k_1 is a unit modulo 26. Then we define our encryption function by

$$e(m) = k_1 m + k_2 \pmod{26}$$

and our decryption function by

$$d(c) = k_1^{-1}(c - k_2) \pmod{26}.$$

Note that k_1^{-1} exists mod 26 since k_1 is a unit mod 26. Also,

$$d(e(m)) \equiv d(k_1 m + k_2) = k_1^{-1}(k_1 m + k_2 - k_2) \equiv k_1^{-1} k_1 m \equiv m \pmod{26}$$

that is, $d(e(m)) = m$ so that d does in fact decrypt $e(m)$ back into m.

If we use the key $k = (5, 12)$ then our encryption process will look like this:

H	E	L	L	O
07	04	11	11	14
↓	↓	↓	↓	↓
21	06	15	15	04
V	G	P	P	E

To be able to decrypt messages encrypted with (k_1, k_2), we need to first determine k_1^{-1} modulo 26. Fortunately, we have the tools for that. For instance, if $k_1 = 5$, then to find k_1^{-1} we must solve $5x \equiv 1 \pmod{26}$. It is easy to see by inspection that $x \equiv -5$ solves this congruence, so $k_1^{-1} \equiv -5 \pmod{26}$. (Had we not been able to find this by inspection, we could have used the Euclidean algorithm.) Thus, to decrypt the letter `V`, which is encoded as 21, using the key $k = (5, 12)$, we compute

$$d(21) \equiv 5^{-1}(21 - 12) \equiv (-5)(9) = 7 \pmod{26}.$$

So the first letter in our plaintext is the letter corresponding to 07—namely, `H`.

Exercise 13.6 (a) Encrypt the message

`UNIT MOD N`

using an affine cipher with key $k = (3, 9)$.

(b) An affine cipher with key $k = (3, 9)$ produced the ciphertext

`HL HWUVIOHMQV.`

What was the original plaintext message?

Before giving our next example, let's pause to introduce some terminology.

Definition 13.7
Cryptosystem

A **cryptosystem** consists of a triple $(\mathcal{M}, \mathcal{C}, \mathcal{K})$, where

- $\mathcal{M}$ is the **plaintext space** consisting of the set of all possible plaintext messages;
- $\mathcal{C}$ is the **ciphertext space** consisting of the set of all possible ciphertexts; and
- $\mathcal{K}$ is the **key space** consisting of all possible keys.

Furthermore, for each **key** $k \in \mathcal{K}$, there must exist an **encryption function** $e_k \colon \mathcal{M} \to \mathcal{C}$ and a **decryption function** $d_k \colon \mathcal{C} \to \mathcal{M}$ that satisfy

$$d_k(e_k(m)) = m \quad \text{for all } m \in \mathcal{M}.$$

For example, for the shift cipher cryptosystem, we have

$$\mathcal{M} = \mathcal{C} = \mathcal{K} = \mathbb{Z}/26\mathbb{Z}.$$

For a key $k \in \mathcal{K}$, the encryption and decryption functions are given by

$$e_k(m) \equiv m + k \pmod{26} \quad \text{and} \quad d_k(c) = c - k \pmod{26}.$$

For the affine cipher, $\mathcal{M} = \mathcal{C} = \mathbb{Z}/26\mathbb{Z}$ again, but the key space is

$$\mathcal{K} = (\mathbb{Z}/26\mathbb{Z})^\times \times \mathbb{Z}/26\mathbb{Z} = \{(k_1, k_2) \colon k_1 \in (\mathbb{Z}/26\mathbb{Z})^\times \text{ and } k_2 \in \mathbb{Z}/26\mathbb{Z}\}.$$

The encryption and decryption functions are given by

$$e_k(m) = k_1 m + k_2 \pmod{26} \quad \text{and} \quad d(c) = k_1^{-1}(c - k_2) \pmod{26}.$$

For a cryptosystem to be secure, we should have conditions on the acceptable encryption and decryption functions. The following is a bare minimum list of requirements.

1. We must be able to compute $e_k(m)$ efficiently for each $k \in \mathcal{K}$ and $m \in \mathcal{M}$.
2. We must be able to compute $d_k(c)$ efficiently for each $k \in \mathcal{K}$ and $c \in \mathcal{C}$.
3. Given a ciphertext $c \in \mathcal{C}$ encrypted using $k \in \mathcal{K}$, it must be very difficult to determine the corresponding plaintext message $d_k(c)$ without knowledge of k.

We might also require further assurances, such as security against **known-plaintext attacks**:

4. Given a list of pairs (m_1, c_1), $\ldots$, (m_n, c_n) of plaintext messages and their corresponding ciphertexts encrypted using the same key $k \in \mathcal{K}$, it must be very difficult to decrypt a ciphertext $c \in \mathcal{C}$ not in this list without knowledge of k.

Exercise 13.8 Which of the above requirements does the shift cipher cryptosystem satisfy? How about the affine cipher cryptosystem?

If you attempted the previous exercise, you would have reached the conclusion that the shift and affine ciphers are inadequate because their key spaces are too small and hence are susceptible to brute-force attacks. Our next example involves a much larger key space.

Example 13.9 (Pohlig–Hellman exponentiation cipher)

This cipher encrypts blocks of size b. For the sake of illustration, let's take $b = 2$. Suppose we want to encrypt the plaintext message

`MARGIN TOO NARROW.`

We begin by removing all spaces and punctuation, then splitting up the letters into blocks of size $b = 2$ and finally encoding the individual letters in each block:

MA	RG	IN	TO	ON	AR	RO	WX
1200	1706	0813	1914	1413	0017	1714	2223

(The last block has been padded with an `X` to make it have the correct size.)

The key space consists of keys $k = (p, e)$, where

- p is a prime that is $>$ the largest number that can occupy a block of size b. For example, when $b = 2$, we want $p > 2525$.
- $e \in \{1, 2, \ldots, p-2\}$ is an integer coprime to $p - 1$.

Given such a $k = (p, e)$, the encryption function is

$$e_k(m) = m^e \pmod{p}.$$

The decryption function involves the inverse d of e modulo $p - 1$ and is given by

$$d_k(c) = m^d \pmod{p}.$$

To prove that $d_k(e_k(m)) = m$, we first note that

$$d_k(e_k(m)) \equiv m^{de} \pmod{p}.$$

Next, since $de = 1 \pmod{p-1}$, it follows from Problem 12.2 that $m^{de} \equiv m \pmod{p}$. (Problem 12.2 only applies if $m \not\equiv 0 \pmod{p}$. But if $m \equiv 0$ then $m^{de} \equiv 0 \equiv m$ too.)

Let's demonstrate. Suppose we pick $(p, e) = (3037, 31)$. Then to encrypt the first block $m = 1200$, we must compute 1200^{31} modulo 3037. Fortunately, this can be done quickly

using a computer (or by hand—plus a pocket calculator—using the square-and-multiply algorithm!). We find that $1200^{31} \equiv 1967 \pmod{3037}$. Doing the same thing to the remaining blocks, we arrive at the ciphertext

$$1967 \quad 2805 \quad 1678 \quad 1178 \quad 2377 \quad 1109 \quad 2047 \quad 2359.$$

(We do not attempt to decode this back into alphabet.)

To decrypt this ciphertext, we first determine the inverse d of 31 modulo 3036 by solving $31d \equiv 1 \pmod{3036}$. (Pay attention! The decryption key uses the inverse of e modulo $p-1$ not modulo p!) Using a computer (or the Euclidean algorithm), we find that $d \equiv 31^{-1} \equiv 1567 \pmod{3037}$. To decrypt the ciphertext block $c = 1967$, we compute

$$c^d \equiv 1967^{1567} \equiv 1200 \pmod{3031}.$$

We have recovered the first block of our original plaintext message.

Exercise 13.10

Use the Pohlig–Hellman cipher with key $(e, p) = (7, 250739)$ and blocksize $b = 3$ to:

(a) Encrypt the plaintext message `EUCLID ALGORITHM`.

(b) Decrypt the ciphertext 216369 50016 52858 112945.

Does the Pohlig–Hellman cipher satisfy our requirements 1–4? The encryption and decryption functions both use modular exponentiation, which can be performed efficiently on a computer. Furthermore, to determine the decryption exponent d from the encryption exponent e, we can simply run the Euclidean algorithm—which also is reasonably efficient. So requirements 1 and 2 are satisfied.

What about requirements 3 and 4? How secure is this cipher? A brute-force attack would be infeasible if p is sufficiently large, say $p \approx 2^{2048}$. But are there other attacks?

Let's think about what Eve would need to do to break the cipher. Assume she has access to a plaintext-ciphertext pair (m, c) and that she somehow knows the prime p in the key $k = (p, e)$. Her task is to determine e, since once she has e she can determine d and then decrypt any ciphertext she comes across. So Eve must solve the equation

$$c \equiv m^e \pmod{p}$$

for e. She is in effect looking for $e =$ "$\log_m(c)$". This is known as the **Discrete Logarithm Problem** (DLP). As far as is known, there is no efficient algorithm for solving the DLP if p is large. A naive trial-and-error approach would require about p steps, so if $p \approx 2^{2048}$ then the solar system would have collapsed into itself before our search for e concluded.

REMARK (Security of a Cryptosystem)

In designing a cryptosystem, the general idea is to prove that an attacker can circumvent the system if and only if they can solve a mathematical problem that is believed (or proved) to be *difficult to solve* in some precise sense (e.g. in terms of computational complexity).

The DLP is an example of such a problem. It is believed that there is no efficient (say, polynomial-time) algorithm that is capable of solving the DLP in general. However, nobody has been able to prove this so far. Perhaps you can try?

Lecture 13 Problems

13.1. Alice and Bob are communicating using an affine cipher with key $k = (k_1, k_2)$. You have intercepted the plaintext-ciphertext pairs $(m, c) = (\texttt{D}, \texttt{L}),\ (\texttt{K}, \texttt{X})$. Use this to discover k.

13.2. Suppose we were to replace the exponentiation encryption and decryption functions in the Pohlig–Hellman algorithm with the functions

$$E(m) \equiv em \pmod{p} \quad \text{and} \quad D(m) \equiv dm \pmod{p}$$

where $e \in (\mathbb{Z}/p\mathbb{Z})^\times$ and d is the inverse of e mod p. Is this cipher secure against known-plaintext attacks? You may assume that p is very large and that Eve knows p.

▶ 13.3. In this exercise you will examine the discrete logarithm $\log_3$ in $(\mathbb{Z}/7\mathbb{Z})^\times$. Let $c \in (\mathbb{Z}/7\mathbb{Z})^\times$ be arbitrary.

(a) Show that the equation $c \equiv 3^e \pmod{7}$ has a solution $e \in \mathbb{Z}$. [**Hint:** Calculate $3^2, 3^3, \ldots$ modulo 7.]

(b) Show that if e and e' are solutions to $c \equiv 3^e \pmod{7}$ then $e \equiv e' \pmod{6}$.

The above allows us to define a function $\log_3 \colon (\mathbb{Z}/7\mathbb{Z})^\times \to \mathbb{Z}/6\mathbb{Z}$ by specifying that $\log_3(c)$ is the unique $[e] \in \mathbb{Z}/6\mathbb{Z}$ such that $3^e \equiv c \pmod{7}$.

(c) Prove that the function $\log_3$ satisfies:

(i) $\log_3(1) \equiv 0 \pmod{6}$.

(ii) $\log_3(ab) \equiv \log_3(a) + \log_3(b) \pmod{6}$ for all $a, b \in (\mathbb{Z}/7\mathbb{Z})^\times$.

Lecture 14 The RSA Cryptosystem

"We stand today on the brink of a revolution in cryptography."

– W. Diffie and M. E. Hellman, *New Directions in Cryptography*

The three cryptosystems we saw in the previous lecture had one thing in common: the same key k was used in both their encryption and decryption functions. For example, in Pohlig–Hellman, the key $k = (p, e)$ allows us to compute the encryption function $e(m) \equiv m^e \pmod{p}$ *and* the decryption function $d(c) \equiv c^d \pmod{p}$ since d can be efficiently obtained from e as its inverse modulo $p - 1$. Conversely, e can be obtained from d. So (p, e) and (p, d) effectively contain the same information. Such systems, where the same key information is used on both the encryption side and decryption side, are called **symmetric key cryptosystems**.

If Alice and Bob are to use a symmetric key cryptosystem, they must somehow agree on a shared secret key. This can prove challenging if they have no secure way to communicate. (After all, wasn't this—the desire to communicate securely over insecure channels—the problem to begin with?)

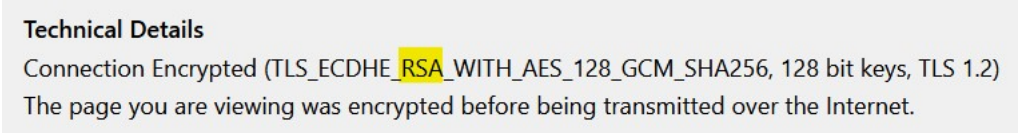
Enter **asymmetric** (or **public**) **key cryptosystems** (PKC). In a PKC, the key consists of two pieces: a private key k_{priv} and a public key k_{pub}. The public key is used to encrypt messages and the private key is used to decrypt ciphertexts. In practice, the public key (and the encryption function) are released to the public but the private key is kept secret. In this way, if Bob has the private key, then everyone (including Alice) can send him messages that only he can decrypt.

REMARK (Symmetric vs. Asymmetric)

PKCs suffer from the drawback that they are generally slower than symmetric cryptosystems. In practice, PKCs are used to share private keys that can then be used in symmetric cryptosystems.

The possibility of asymmetric encryption was put forward by Diffie and Hellman (in the paper quoted at the top of this page) and independently by Merkle. Nowadays there are various PKCs. The most popular by far is the RSA cryptosystem, named after its creators Rivest, Shamir and Adleman (although it was already known to the UK intelligence agency GCHQ!).

Next time you browse an HTTPS website secured by SSL/TLS, try to look up the certificate data to see what it entails. For example, this is what my browser tells me when I visit https://uwaterloo.ca:[12]



[12]The other fragments are also interesting: ECDHE is Elliptic Curve Diffie–Hellman Ephemeral key exchange; see Problem 14.1 for the basic idea. AES-128 is a popular symmetric key cryptosystem that uses 128-bit keys. GCM is Galois/Counter Mode, which is a mode of operation for blockciphers that involves the arithmetic of finite fields. SHA256 is the Secure Hash Algorithm, which is used to insert pseudorandomness into the protocol.

RSA has two phases: key generation (where k_{priv} and k_{pub} are created) and then encryption/decryption using the generated key $k = (k_{\text{priv}}, k_{\text{pub}})$.

RSA Key Generation

- Choose distinct large primes p and q, say $p, q \geq 2^{2048}$, and let $n = pq$.
- Choose an integer e that is coprime to $\varphi(n)$.
- The public key k_{pub} is the pair (n, e). *The primes p and q are to be kept secret.*
- Compute the inverse d of e modulo $\varphi(n)$.
- The private key k_{priv} is the integer d (and the primes p and q).

RSA Encryption and Decryption Functions

- Given the public key (n, e), the encryption function is defined by

$$e(m) \equiv m^e \pmod{n}.$$

  The plaintext space $\mathcal{M}$ is the set of integers m in the interval $0 \leq m < n$ that are coprime to n.

- Given the private key d, the decryption function is

$$d(c) \equiv c^d \pmod{n}.$$

Note the similarity to the Pohlig–Hellman cipher. The difference here is that the modulus n is composite. This creates a minor complication: we must now insist that m be coprime to n. In practice this is not an issue since we can always pad a plaintext message to achieve this coprimality requirement. If m is coprime to n, we can verify that d decrypts $e(m)$:

$$d(e(m)) \equiv d(m^e) \equiv m^{ed} \equiv m^1 \pmod{n},$$

where the last step follows from Problem 12.2 together with the fact that $ed \equiv 1 \pmod{\varphi(n)}$.[13]

Example 14.1 Suppose Alice wants to let Bob know that the password to her banking account is

`WONDERLAND`.

Bob generates an RSA key as follows. He picks $p = 1000183$ and $q = 2593697$ and sets

$$n = pq = 2594171646551.$$

He picks $e = 65537$, which he confirms is coprime to $\varphi(n) = (p-1)(q-1)$. Finally, he computes the inverse of e mod $\varphi(n)$ to be $d = 1675321858817$.

Bob publishes (e, n) and keeps d (and p and q) private.

[13] In fact, $d(e(m)) = m$ even if m and n are not coprime. (See Problem 14.2.) However, in practice, if we find ourselves with a message m that isn't coprime to n then we can use it to factor n. Indeed, we can efficiently compute $\gcd(m, n)$ and this gcd will be either p or q. Since factoring n breaks RSA (see the next page), it's more secure to only use m such that $\gcd(m, n) = 1$.

Alice will now encrypt her message using the public key (e, n). She first converts it into an integer:

W	O	N	D	E	R	L	A	N	D
22	14	13	03	04	17	11	00	13	03

She breaks this up into two blocks $m_1 = 2214130304$ and $m_2 = 1711001303$ (to get integers smaller than n). She encrypts these into the ciphertexts

$$c_1 \equiv e(m_1) \equiv m_1^e \equiv 2393950175804 \pmod{n}$$

and

$$c_2 \equiv e(m_2) \equiv m_2^e \equiv 738407767416 \pmod{n}$$

which she then sends to Bob.

Bob takes each ciphertext and then computes

$$d(c_1) \equiv c_1^d \equiv 2214130304 \pmod{n}$$

and

$$d(c_1) \equiv c_1^d \equiv 1711001303 \pmod{n}.$$

He has thus recovered Alice's m_1 and m_2.

Exercise 14.2 Bob has generated the public key $(e, n) = (3, 6319)$.

(a) Determine $\varphi(n)$ and compute the private key k.

(b) Encrypt the plaintext message `ZETA`. (Break it up into two pieces.)

(c) Decrypt the ciphertext 2173 5047.

Let's wrap up with a brief analysis of the security of the RSA cryptosystem.

How easily can Eve determine d given the public key (e, n)? Since d is the inverse of e mod $\varphi(n)$, if Eve can compute $\varphi(n)$, she can find d using the Euclidean algorithm.

How difficult is it to compute $\varphi(n)$? It's very easy if we can determine the prime factorization of n. (See Problem 11.5. We'll say more next lecture.) In fact, for integers of the form $n = pq$, computing $\varphi(n)$ is equivalent to factoring n. To see why, suppose you can compute $\varphi(n) = (p-1)(q-1) = pq - (p+q) + 1 = (n+1) - (p+q)$. Then you can determine the sum $p + q$. Since you also know the product $pq = n$, you can find p and q by solving the quadratic equation

$$(x - p)(x - q) = x^2 - (p + q)x + pq = 0.$$

Exercise 14.3 Given that $pq = 239777$ and $p + q = 1038$, determine p and q.

So, computing $\varphi(n)$ is just as difficult as factoring $n = pq$. If p and q are large primes, this is known to be a *difficult problem* in general (i.e. no efficient algorithm for factoring large integers n is known). Can we determine the private key d by some other means that don't involve computing $\varphi(n)$ or factoring n? The answer is *no*. It is known that if you can determine the private key d then you can use this information to factor n. So the problems of factoring n and determining d are equivalent.

We conclude that RSA is secure inasmuch as it is difficult for Eve to obtain d from (e, n) in general. However, what if Eve only wants to decrypt a specific ciphertext $c \equiv m^e \pmod{n}$? Strictly speaking, all that is necessary here is the ability to take an eth root modulo n. This *might be* an easier problem to solve. However, as far as we know, it is just as difficult as factoring n. (Though there is no proof that this is the case.)

So assuming that the problem of cracking RSA is equivalent to the problem of efficiently factoring large integers, which decades of evidence indicates is the case, we can surmise that RSA is secure. (In truth, the version of RSA presented here—which is called *Textbook RSA*—is susceptible to a variety of attacks in certain circumstances. For more information, see D. Boneh, *Twenty Years of Attacks on the RSA Cryptosystem.* Notices of the AMS, 46 (2), 203–213, 1999.)

14.1 Number Theoretic Problems Inspired by Cryptography

Our foray into cryptography has raised many interesting problems. Here is a sampling:

1. **How do we find large primes? / How do we efficiently test for primality?**

   To be secure against brute-force attacks, we found ourselves needing to work modulo a large prime or product of large primes. For example, in RSA, we should use primes that are roughly of size 2^{2048} or 2^{4096}. (We could use larger primes, of course, but that might be a waste of computer memory.) The Prime Number Theorem guarantees that there are plenty of such primes, but how do we find them?

2. **How do we solve the Discrete Logarithm Problem (DLP)?**

   That is, given $a, b \in (\mathbb{Z}/n\mathbb{Z})^\times$, how can we find $x \in \mathbb{Z}$ such that $a \equiv b^x \pmod{n}$? Strictly speaking, this is the DLP for the group $(\mathbb{Z}/n\mathbb{Z})^\times$. We can formulate the DLP in a general group (solve for $x \in \mathbb{Z}$ given $g = h^x$ in G). We are typically more interested in showing that the DLP cannot be solved efficiently, since the DLP forms the mathematical backbone of several widely used cryptosystems. If we can find a group G in which the DLP is provably difficult, then that would be good news.

3. **How do we compute the Euler phi function $\varphi(n)$?**

   In our discussion of RSA, we worked with $\varphi(n)$ where $n = pq$ is a product of two distinct primes. We discovered that computing $\varphi(n)$ in this case is equivalent to being able to factor n. What can be said about computing $\varphi(n)$ in general?

4. **How do we efficiently factor large integers?**

   This might seem to be a special case of the first question above, but this is not the case! Testing for primality turns out to be easier than factoring.

We will explore these problems over the next several lectures.

Lecture 14 Problems

14.1. The **Diffie–Hellman Key Exchange** is a protocol that allows Alice and Bob to share a secret key across a public channel. It works as follows:

- Alice and Bob publicly agree on prime p and an integer g.
- Alice selects a private key $a \in \mathbb{Z}$ and sends Bob the public key $A \equiv g^a \pmod{p}$.
- Bob selects a private key $b \in \mathbb{Z}$ and sends Alice the public key $B \equiv g^b \pmod{p}$.

(a) Prove that $A^b \equiv B^a \pmod{p}$. This is Alice and Bob's shared secret key K.

(b) Eve knows $A = g^a$, $B = g^b$, g and p. Her goal is to determine the secret key K from this information. This is known as the **Diffie–Hellman Problem**. Show that if Eve can efficiently solve the Discrete Logarithm Problem then she can efficiently solve the Diffie–Hellman Problem. [**Note:** The converse is an open problem.]

14.2. Let p and q be primes and let $e, d \in \mathbb{Z}$ satisfy $ed \equiv 1 \pmod{\varphi(pq)}$. The goal of this problem is to prove that $m^{ed} \equiv m \pmod{pq}$ for all $m \in \mathbb{Z}$. (Problem 12.2 gives us this for m coprime to pq.)

(a) Prove that $m^{ed} \equiv m \pmod{p}$ and $m^{ed} \equiv m \pmod{q}$.

(b) Use the Chinese Remainder Theorem to deduce that $m^{ed} \equiv m \pmod{pq}$.

14.3. Alice and Bob use RSA with the same modulus n but different encryption exponents e_A and e_B. Charles sends Alice and Bob the same message m using the keys (e_A, n) and (e_B, n). Eve intercepts the resulting ciphertexts c_A and c_B. Assuming that e_A and e_B are coprime, show that Eve can recover the original message m.

Lecture 15 Arithmetic Functions

The Euler phi function φ satisfies the following two properties:

(1) $\varphi(nm) = \varphi(n)\varphi(m)$ if $\gcd(n, m) = 1$.

(2) $\varphi(p^a) = p^a - p^{a-1}$ if p is prime and $a \in \mathbb{Z}_{>0}$.

This gives us a method for computing $\varphi(n)$ for any $n > 1$ as follows. First, determine the prime factorization $n = p_1^{a_1} \cdots p_k^{a_k}$ of n, and then

$$\varphi(n) = \varphi(p_1^{a_1}) \cdots \varphi(p_1^{a_1}) = (p_1^{a_1} - p_1^{a_1-1}) \cdots (p_k^{a_k} - p_k^{a_k-1}).$$

For example,

$$\varphi(200) = \varphi(2^3 5^2) = \varphi(2^3)\varphi(5^2) = (2^3 - 2^2)(5^2 - 5) = 80.$$

This beats trying to count all the integers in the interval $1 \leq a \leq 200$ that are coprime to 200.

What I want to do in this lecture and the next is prove properties (1) and (2). It's possible to prove both using counting arguments similar to those in the solutions to Problem 11.5. However, I will present a roundabout proof as an excuse to showcase some interesting mathematics that can be applied in more general situations.

The Euler φ function is an example of what is called an arithmetic function.

Definition 15.1

Arithmetic function

An **arithmetic function** is a function whose domain is $\mathbb{Z}_{>0}$ and whose range is a subset of $\mathbb{C}$.

An arithmetic function f is said to be

- **multiplicative** if $f(nm) = f(n)f(m)$ for all $n, m \in \mathbb{Z}_{>0}$ such that $\gcd(n, m) = 1$;
- **completely multiplicative** if $f(nm) = f(n)f(m)$ for all $n, m \in \mathbb{Z}_{>0}$.

The next proposition follows immediately from the above definitions.

Proposition 15.2

Suppose that $n = p_1^{a_1} \cdots p_k^{a_k}$ is the prime factorization of n. If f is a multiplicative function, then

$$f(n) = f(p_1^{a_1}) \cdots f(p_k^{a_k}).$$

If f is completely multiplicative, then

$$f(n) = f(p_1)^{a_1} \cdots f(p_k)^{a_k}.$$

We will be particularly interested in arithmetic functions $f(n)$ that have something to do with the arithmetic nature of n. Euler's $\varphi(n)$ and the p-adic valuation $v_p(n)$ are examples. The prime counting function $\pi(x)$ as we've defined it is not an arithmetic function according to the above definition since its domain is $\mathbb{R}_{>0}$. However, if we restrict the domain to $\mathbb{Z}_{>0}$ then $\pi(n)$ becomes an arithmetic function. Here are some more examples:

- The (little) **omega function** $\omega(n)$ = number of distinct prime divisors of n.

  For example, $\omega(1) = 0$, $\omega(8) = \omega(2^3) = 1$ and $\omega(12) = \omega(2^2 \cdot 3) = 2$.

- The (big) **omega function** $\Omega(n)$ = number of distinct prime divisors of n counted according to their valuation.

  For example, $\Omega(10) = \Omega(2^1 \cdot 5^1) = 1 + 1 = 2$, $\Omega(12) = \Omega(2^2 \cdot 3) = 2 + 1 = 3$ and $\Omega(100) = \Omega(2^2 \cdot 5^2) = 2 + 2 = 4$.

- The **divisor functions** $\sigma_k(n) = \sum_{d|n} d^k$, where the sum runs over the positive divisors of n and where $k \in \mathbb{R}$.

  For example
  $$\sigma_0(10) = \sum_{d|10} d^0 = 1^0 + 2^0 + 5^0 + 10^0 = 4$$
  and
  $$\sigma_1(10) = \sum_{d|10} d^1 = 1^1 + 2^1 + 5^1 + 10^1 = 18.$$
  Notice that $\sigma_0(n)$ is the number of positive divisors of n and $\sigma_1(n)$ is their sum. We set $d(n) = \sigma_0(n)$ and $\sigma(n) = \sigma_1(n)$.

Here are a couple of more "creative" examples:

- The **sum-of-squares function** $r_2(n) = \#\{(x, y) \in \mathbb{Z}^2 : x^2 + y^2 = n\}$. That is, $r_2(n)$ is the number of solutions to the equation $x^2 + y^2 = n$ with $x, y \in \mathbb{Z}$.

  For example, $r_2(1) = 4$ since there are four solutions to $x^2 + y^2 = 1$ (namely: $(x, y) = (\pm 1, 0)$ and $(x, y) = (0, \pm 1)$). On the other hand, $r_2(23) = 0$ since $23 \equiv 3 \pmod 4$ and we have seen that the sum of two squares cannot be congruent to 3 modulo 4.

- The **Ramanujan tau function** $\tau(n)$ is the coefficient of q^n in the expansion
  $$q \prod_{n=1}^{\infty} (1 - q^n)^{24} = q(1 - q)^{24}(1 - q^2)^{24} \cdots = \tau(1)q + \tau(2)q^2 + \tau(3)q^3 + \cdots .$$
  For example, $\tau(1) = 1$, $\tau(2) = -24$ and $\tau(3) = 252$. (Can you determine $\tau(4)$?)

  Ramanujan stumbled upon this function while investigating the divisor functions $\sigma_k(n)$. He stated three famous conjectures concerning $\tau(n)$. The first is that $\tau(n)$ is multiplicative; the second is a recursive formula for $\tau(p^k)$ for p a prime; and the third is the inequality $|\tau(n)| \leq n^{11/2} d(n)$, where $d(n) = \sigma_0(n)$ as above. The first two conjectures were proved quickly by Mordell. The third was much more difficult. Pierre Deligne won a Fields Medal in part for his proof of Ramanujan's third conjecture.

One of our two goals (1) and (2) is to prove that $\varphi(n)$ is multiplicative. Of the functions above, we've noted that $\tau(n)$ is multiplicative. What about the others?

Exercise 15.3 For each of $v_p(n)$, $\pi(n)$, $\omega(n)$, $\Omega(n)$ and $r_2(n)$, determine whether it is multiplicative, completely multiplicative or neither.

Are the divisor functions $\sigma_k(n)$ multiplicative? The answer is *yes*. The proof is actually not that difficult, so I encourage you to pause and try come up with it yourself before reading ahead.

We will need a simple, but very useful, lemma.

Lemma 15.4 Let $m, n \in \mathbb{Z}$ be coprime. If $d \mid mn$ then we express d in the form $d = d_1 d_2$ where $d_1 \mid m$ and $d_2 \mid n$. Conversely, if $d_1 \mid m$ and $d_2 \mid n$, then $d_1 d_2 \mid mn$.

Proof: If $m = 1$ or $n = 1$, this is trivial, so we may assume that $m, n \neq 1$. Consider then the prime factorizations of all integers involved. Since m and n are coprime, they do not share primes in common. So if $d \mid mn$ then we can break up its prime factorization into two pieces: one piece consisting of the primes appearing in m and the other consisting of the primes appearing in n. These are our desired d_1 and d_2. The converse follows from Proposition 4.3(b) since d_1 and d_2 must be coprime and they each divide mn by transitivity. ■

This lemma shows in particular that, if m and n are coprime, the divisors of mn are in one-to-one correspondence with divisors of m and n.

Proposition 15.5 The function $\sigma_k(n) = \sum_{d|n} d^k$ is multiplicative.

Proof: We must show that $\sigma_k(mn) = \sigma_k(m)\sigma_k(n)$ whenever $m, n \in \mathbb{Z}_{>0}$ are coprime. Lemma 15.4 implies that summing over the divisors d of mn is the same as summing over $d_1 d_2$ where $d_1 \mid m$ and $d_2 \mid n$. Thus,

$$\begin{aligned}
\sigma_k(mn) &= \sum_{d|mn} d^k \\
&= \sum_{\substack{d_1|m \\ d_2|n}} (d_1 d_2)^k \\
&= \sum_{d_1|m} \sum_{d_2|n} d_1^k d_2^k \\
&= \left(\sum_{d_1|m} d_1^k \right) \left(\sum_{d_2|n} d_2^k \right) \\
&= \sigma_k(m)\sigma_k(n),
\end{aligned}$$

as desired. ■

Exercise 15.6 Is $\sigma_k(n)$ completely multiplicative?

As a consequence of Propositions 15.2 and 15.5, we've reduced the problem of computing $\sigma_k(n)$ to the problem of computing $\sigma_k(p^i)$ where p is a prime. We can use this to obtain formulas for the number of positive divisors of n and their sum. See Problem 15.3.

If we analyze the proof of Proposition 15.5, we discover that there are two things that make it work: First, we are able to break up the sum $\sum_{d|mn}$ into $\sum_{d_1|m}\sum_{d_2|n}$ thanks to Lemma 15.4. Second, we are able to split up the summand d^k into $d_1^k d_2^k$ since the function $f(d) = d^k$ is multiplicative. So the same proof would show that every function of the form $F(n) = \sum_{d|n} f(d)$, with $f(d)$ multiplicative, is itself multiplicative. This prompts the following definition.

Definition 15.7
Summatory Function

If $f(n)$ is an arithmetic function, then its **summatory function** is the arithmetic function $F(n)$ defined by

$$F(n) = \sum_{d|n} f(d)$$

where the sum runs over the positive divisors d of n.

REMARK

You can think of the summatory function as a kind of "discrete" integral of the function $f(d)$. Perhaps a more appropriate integral would be a function of the form $F(x) = \sum_{d\leq x} f(d)$. Functions of this type are in fact studied in number theory; they are, confusingly, also called summatory functions.

For example, the summatory function of $f(n) = n^k$ is $F(n) = \sum_{d|n} d^k = \sigma_k(n)$.

Theorem 15.8

Let f be an arithmetic function and let F be the summatory function of f. Then f is multiplicative if and only if F is multiplicative.

If f is multiplicative, we can prove that its summatory function F is multiplicative by mimicking the proof of Proposition 15.5, as we've indicated above. The converse will be proved next lecture by means of *Möbius inversion*—a technique that allows us to express f in terms of F. (See Corollary 16.5.)

Our strategy for proving (1)—that is, that $\varphi(n)$ is multiplicative—will be to prove that its summatory function is multiplicative! We will end this lecture by determining the summatory function of $\varphi(n)$.

Theorem 15.9

For all $n \in \mathbb{Z}_{>0}$, we have

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

Proof: Let $S = \{1, \ldots, n\}$ and let $S_d = \{k \in S\colon \gcd(k, n) = d\}$. Observe that each $k \in S$ belongs to exactly one S_d, so the sets S_d partition S. Of course, S_d is empty if $d \nmid n$ since the condition $\gcd(k, n) = d$ is impossible in this case. Thus,

$$n = |S| = \sum_{d|n} |S_d|.$$

Now,

$$\begin{aligned} k \in S_d &\iff 1 \le k \le n \text{ and } \gcd(k,n) = d \\ &\iff 1 \le k \le n \text{ and } \gcd(k/d, n/d) = 1 \\ &\iff 1 \le k/d \le n/d \text{ and } \gcd(k/d, n/d) = 1. \end{aligned}$$

Setting $a = k/d$, we see that the size of S_d is equal to the number of integers in the interval $1 \le a \le n/d$ that are coprime to n/d. That is, $|S_d| = \varphi(n/d)$. Consequently,

$$n = \sum_{d|n} |S_d| = \sum_{d|n} \varphi(n/d).$$

Finally, note that as d runs over the divisors of n then so does n/d. Thus,

$$\sum_{d|n} \varphi(n/d) = \sum_{d|n} \varphi(d),$$

which completes the proof. ■

Thus, the summatory function of $\varphi(n)$ is the constant function $F(n) = n$. Since the latter is clearly multiplicative, it follows that $\varphi(n)$ is multiplicative. This completes our proof of (1)—except we still need to prove the reverse implication in Theorem 15.8. We'll do this next time.

Lecture 15 Problems

15.1. Let f and g be multiplicative functions. Prove that if $f(p^a) = g(p^a)$ for all primes p and all $a \in \mathbb{Z}_{\ge 0}$ then $f(n) = g(n)$ for all $n \in \mathbb{Z}_{>0}$.

15.2. Prove that if f is multiplicative then either $f(1) = 1$ or else $f(n) = 0$ for all $n \in \mathbb{Z}_{>0}$.

15.3. (a) Prove that

$$d(n) = \prod_{p|n} (v_p(n) + 1) \quad \text{and} \quad \sigma(n) = \prod_{p|n} \left(\frac{p^{v_p(n)+1} - 1}{p - 1} \right),$$

where each product is over the prime divisors p of n. [**Hint:** σ_k is multiplicative.]

(b) Determine $d(100)$ and $\sigma(100)$.

15.4. For $n \in \mathbb{Z}_{>0}$, let

$$\chi(n) = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \equiv 1 \pmod 4 \\ -1 & \text{if } n \equiv 3 \pmod 4. \end{cases}$$

(a) Prove that $\chi(n)$ is a completely multiplicative function.

(b) Determine whether $X(n) = \sum_{d|n} \chi(d)$ is multiplicative, completely multiplicative or neither.

(c) Show that $r_2(n) = 4X(n)$ for $n = 1, 2, \ldots, 10$. [**Note:** In fact, $r_2(n) = 4X(n)$ for all n.]

15.5. Determine all multiplicative functions $f(n)$ that satisfy

$$n \equiv m \pmod 3 \implies f(n) = f(m).$$

Lecture 16 Möbius Inversion

Let's recap. We wanted to prove that the Euler phi function $\varphi(n)$ is multiplicative, meaning that $\varphi(mn) = \varphi(m)\varphi(n)$ whenever $m, n \in \mathbb{Z}_{>0}$ are coprime. We claimed that an arithmetic function $f(n)$ is multiplicative if and only if its summatory function $F(n) = \sum_{d|n}$ is multiplicative, and proved the direction

$$f(n) \text{ is multiplicative} \quad \Longrightarrow \quad F(n) \text{ is multiplicative.}$$

In this lecture we will prove the $\Longleftarrow$ implication. Since the summatory function of $f(n) = \varphi(n)$ is $F(n) = n$ (Theorem 15.9), which is multiplicative, this will allows us to conclude that $\varphi(n)$ is multiplicative.

Our proof will involve expressing $f(n)$ in terms of its summatory function $F(n)$. Let's see why this ought to be possible. Consider the following values of F:

$$\begin{aligned}
F(1) &= f(1) \\
F(2) &= f(1) + f(2) \\
F(3) &= f(1) + f(3) \\
F(4) &= f(1) + f(2) + f(4) \\
F(5) &= f(1) + f(5) \\
F(6) &= f(1) + f(2) + f(3) + f(6).
\end{aligned}$$

We can "invert" the above system to obtain

$$\begin{aligned}
f(1) &= F(1) \\
f(2) &= F(2) - F(1) \\
f(3) &= F(3) - F(1) \\
f(4) &= F(4) - F(2) \\
f(5) &= F(5) - F(1) \\
f(6) &= F(6) - F(3) - F(2) + F(1).
\end{aligned}$$

Continuing in this manner, we begin to suspect that there is a relation of the form

$$f(n) = \sum_{d|n} \mu_d F(n/d)$$

where the coefficients μ_d are 0 or ± 1. This is indeed the case. With a little more thought, we can discover what μ_d should be. The upshot is as follows.

Definition 16.1

The Möbius Function

The **Möbius function** μ is defined by

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ has a repeated prime divisor} \\ 1 & \text{if } n = 1 \\ (-1)^k & \text{if } n \text{ is a product of } k \text{ distinct primes.} \end{cases}$$

for all $n \in \mathbb{Z}_{>0}$.

For example $\mu(4) = \mu(54) = 0$ since $2^2 \mid 4$ and $3^3 \mid 54$, while $\mu(30) = \mu(2{\cdot}3{\cdot}5) = (-1)^3 = -1$.

Exercise 16.2

(a) Determine $\mu(n)$ for $n = 1, 2, \ldots, 10$.

(b) Prove that $\mu(n)$ is multiplicative but not completely multiplicative.

The next result determines the summatory function of $\mu(n)$.

Theorem 16.3

For all $n \in \mathbb{Z}_{>0}$, we have

$$\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

(This can be expressed more compactly as: $\sum_{d|n} \mu(d) = \lfloor \frac{1}{n} \rfloor$.)

Proof: If $n = 1$ then the sum is simply $\mu(1) = 1$, so that's that. If $n > 1$ and $n = p_1^{a_1} \cdots p_k^{a_k}$ is the prime factorization of n, then since $\mu(d) = 0$ if d has a repeated prime factor, we have

$$\sum_{d|n} \mu(d) = \mu(1) + \sum_{p|n} \mu(p) + \sum_{p,q|n} \mu(pq) + \cdots + \mu(p_1 \cdots p_k),$$

where the sums are over the distinct prime divisors of n, pairs of distinct prime divisors of n, and so on. In the ith sum, we are summing over the $\binom{k}{i}$ choices of i distinct prime divisors of n. Since $\mu(p_{j_1} \cdots p_{j_i}) = (-1)^i$, it follows that

$$\sum_{n|d} \mu(d) = \sum_{i=0}^{k} \binom{k}{i} (-1)^i = (1-1)^k = 0,$$

where the binomial theorem was used to sum up the series. ■

With this in hand, we can now show how to recover f from its summatory function.

Theorem 16.4 **(Möbius Inversion Formula)**

Let f be an arithmetic function and let $F(n) = \sum_{d|n} f(d)$ be the summatory function of f. Then

$$f(n) = \sum_{d|n} \mu(d) F(n/d).$$

Proof: We have

$$\begin{aligned} \sum_{d|n} \mu(d) F(n/d) &= \sum_{d|n} \mu(d) \sum_{e|n/d} f(e) \\ &= \sum_{d|n} \sum_{e|n/d} \mu(d) f(e) \end{aligned}$$

In the above, we are summing up over all pairs (d, e) such that $d \mid e$ and $e \mid n/d$. Observe that if $e \mid n/d$ then $e \mid n$. Conversely, if $e \mid n$ then e will divide n/d if and only if d divides

n/e (in which case d will divide n too). So we may view the sum as being over all pairs (d, e) such that $e \mid n$ and $d \mid n/e$. Thus, we can interchange the roles of e and d in the expression above, leaving us with

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

By Theorem 16.3,

$$\sum_{d|n/e} \mu(d) = \begin{cases} 1 & \text{if } n/e = 1, \\ 0 & \text{otherwise.} \end{cases}$$

So in the above sum all terms except the $e = n$ term vanish, leaving us with $f(n)$. This completes the proof. ■

Let's record an important corollary.

Corollary 16.5 Let $F(n) = \sum_{d|n} f(d)$ be the summatory function of f. If F is multiplicative then f is multiplicative.

Proof: Möbius inversion gives $f(n) = \sum_{d|n} \mu(d) F(n/d)$. So if $m, n \in \mathbb{Z}_{>0}$ are coprime, then

$$\begin{aligned} f(mn) &= \sum_{d|mn} \mu(d) F(mn/d) \\ &= \sum_{d_1|m} \sum_{d_2|n} \mu(d_1 d_2) F(mn/(d_1 d_2)) && \text{(Lemma 15.4)} \\ &= \sum_{d_1|m} \mu(d_1) F(m/d_1) \sum_{d_2|n} \mu(d_2) F(n/d_2) && (\mu \text{ and } F \text{ are multiplicative}) \\ &= f(m) f(n), \end{aligned}$$

as desired. ■

Example 16.6 **(Möbius inversion applied to $\sigma_k(n)$)**

The divisor function $\sigma_k(n) = \sum_{d|n} n^k$ can be viewed as the summatory function of the power function $f(n) = n^k$. Applying Möbius inversion, we obtain

$$n^k = \sum_{d|n} \mu(d) \sigma_k(n/d).$$

For instance, if $k = 0$, this result gives the (non-obvious) identity

$$1 = \sum_{d|n} \mu(d) \sigma_0(n/d)$$

where $\sigma_0(n/d)$ is the number of positive divisors of n/d. To illustrate, let $n = 10$. Then

$$\begin{aligned}\sum_{d|10} \mu(d)\sigma_0(10/d) &= \mu(1)\sigma_0(10) + \mu(2)\sigma_0(5) + \mu(5)\sigma_0(2) + \mu(10)\sigma_0(1) \\ &= 1(4) + (-1)(2) + (-1)(2) + (1)(1) \\ &= 1.\end{aligned}$$

If we apply Möbius inversion to $\varphi(n)$ and its summatory function, we can finally prove both items (1) and (2) from the previous lecture.

Proposition 16.7

(1) $\varphi(nm) = \varphi(n)\varphi(m)$ if $\gcd(n, m) = 1$.

(2) $\varphi(p^a) = p^a - p^{a-1}$ if p is prime and $a \in \mathbb{Z}_{>0}$.

Proof: The summatory function of $\varphi(n)$ is the constant function $F(n) = n$ (Theorem 15.9). Since $F(n) = n$ is obviously multiplicative, item (1) follows from Corollary 16.5.

For item (2), we apply Möbius inversion to $f(n) = \varphi(n)$ and $F(n) = n$, obtaining

$$\varphi(n) = \sum_{d|n} \mu(d)\frac{n}{d}.$$

Letting $n = p^a$, we find

$$\varphi(p^a) = \sum_{d|p^a} \mu(p^a)\frac{p^a}{d}.$$

The positive divisors of p^a are the integers p^b with $0 \le b \le a$. If $b > 1$ then $\mu(p^b) = 0$, leaving us with

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

as desired. ■

By combining items (1) and (2), we obtain the following corollary.

Corollary 16.8

If $n = p_1^{a_1} \cdots p_k^{a_k}$ is the prime factorization of n, then

$$\varphi(n) = \prod_{i=1}^{k} \left(p_i^{a_i} - p_i^{a_i - 1}\right).$$

Exercise 16.9

Compute $\varphi(2024)$.

REMARK (Computing $\varphi(n)$)

Corollary 16.8 answers question 3 in Section 14.1 with the caveat that we have the prime factorization of n available to us. However, since there are no known efficient algorithms for factoring integers, this is not a very satisfying answer. Alas, it's expected that computing arbitrary values of $\varphi(n)$ is as hard as factoring arbitrary integers. We already saw this explicitly in the case where $n = pq$ was the product of two primes. We proved that knowing the value of $\varphi(n)$ allows one to determine p and q (see bottom of page 78).

Lecture 16 Problems

16.1. Let $n > 1$. Prove that

$$\varphi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right),$$

where the product is over the prime divisors p of n.

16.2. Let $n, m \in \mathbb{Z}_{>0}$. Prove:

(a) $\varphi(nm) = \varphi(n)\varphi(m)(d/\varphi(d))$, where $d = \gcd(m, n)$.

(b) If $n \mid m$ then $\varphi(n) \mid \varphi(m)$. [**Hint:** Try induction on m and use part (a).]

(c) $\varphi(n)$ is even for all $n \geq 3$.

16.3. The **van Mangoldt function** is defined by

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^a \text{ for some prime } p \text{ and } a \in \mathbb{Z}_{>0} \\ 0 & \text{otherwise.} \end{cases}$$

This function plays an important role in the proof of the Prime Number Theorem. Prove:

(a) $\sum_{d|n} \Lambda(d) = \log n$.

(b) $\Lambda(n) = -\sum_{d|n} \mu(d) \log d$.

16.4. If f and g are arithmetic functions, we define their **convolution** $f * g$ by

$$(f * g)(n) = \sum_{d|n} f(d)g(n/d).$$

Let f, g and h be arithmetic functions. Prove:

(a) $f * g = g * f$.

(b) $(f * g) * h = f * (g * h)$.

(c) $f * e = e * f$, where e is defined by $e(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1. \end{cases}$

(d) If $f(1) \neq 0$, then $f * f^\iota = e = f^\iota * f$, where f^ι is defined recursively by

$$f^\iota(1) = \frac{1}{f(1)} \quad \text{and} \quad f^\iota(n) = -\frac{1}{f(1)} \sum_{d|n, d<n} f^\iota(d) f(n/d) \text{ for } n > 1.$$

[**Note:** This shows that the set of arithmetic functions f with $f(1) \neq 0$ forms a commutative group with respect to the convolution operation. The identity element is the function e and the inverse of f is the function f^ι defined above.]

16.5. This problem is a continuation of the previous one.

(a) The **unit function** u is defined by $u(n) = 1$ for all $n \in \mathbb{Z}_{>0}$. Prove that $u * \mu = e$.

(b) Let f be an arithmetic function and let F be its summatory function. Prove that $F = f * u$ and hence deduce that $f = F * \mu$.

[**Note:** This gives a one-line proof of the Möbius inversion formula!]

Lecture 17 The Discrete Logarithm

Recall that the Discrete Logarithm Problem asks us to find $x \in \mathbb{Z}$ such that $a^x \equiv b \pmod{n}$, where a and b are integers coprime to n. The idea is that such a solution x, if one exists, ought to be called "$\log_a(b)$." There are a couple of subtleties that arise. Let's consider some examples.

Example 17.1 Find all $x \in \mathbb{Z}$, if any, that satisfy the given congruences.

(a) $3^x \equiv 6 \pmod{7}$.

(b) $4^x \equiv 6 \pmod{7}$.

Solution:

(a) Let's calculate powers of 3 modulo 7:

$$3^1 \equiv 3$$
$$3^2 \equiv 2$$
$$3^3 \equiv 6$$

and we're done! We found a solution, namely $x = 3$. But let's continue:

$$3^4 \equiv 4$$
$$3^5 \equiv 5$$
$$3^6 \equiv 1$$
$$3^7 \equiv 3$$
$$3^8 \equiv 2$$
$$3^9 \equiv 6.$$

We've discovered another solution: $x = 9$. What happened here is that the moment we arrived at $3^6 \equiv 1$, the pattern of values of 3^x began to repeat, so we were destined to find another solution. The powers of 3 repeat in cycles of length 6: 3, 2, 6, 4, 5, 1, and repeat. Since we found that $3^x \equiv 6$ when $x = 3$, then we will have $3^x \equiv 6$ for all $x = 3 + 6k$ with $k \in \mathbb{Z}$. Indeed:

$$3^{3+6k} = 3^3(3^6)^k \equiv 6(1)^k \equiv 6 \pmod{7}.$$

Based on our computations of powers of 3, these are the only exponents that produces a 6. Thus, there are infinitely many possible values for "$\log_3(6)$", namely $3 + 6k$ for $k \in \mathbb{Z}$.

(b) Let's calculate powers of 4 modulo 7:

$$4^1 \equiv 4$$
$$4^2 \equiv 2$$
$$4^3 \equiv 1$$

and now the pattern of values of 4^x will repeat: 4, 2, 1, 4, 2, 1, ... This time the cycle length is only 3, and 6 never occurs amongst these values. So there are no solutions to $4^x \equiv 6 \pmod{7}$. Thus, "$\log_3(4)$" is undefined.

We learn two things from Example 17.1:

1. There might be infinitely many solutions to $a^x \equiv b \pmod{n}$. That is, $\log_a(b)$ might be "multi-valued."
2. There might be no solutions to $a^x \equiv b \pmod{n}$. That is, $\log_a(b)$ might be undefined.

The latter is not too surprising, since for example in $\mathbb{R}$, $\log_e(-1)$ is undefined. The former might be surprising if you're only familiar with real logarithms, but if you know about complex logarithms, then you will recall that $\log_e(z)$ is multi-valued and is defined only up to an integer multiple $2\pi i$. This is analogous to the $\mathbb{Z}/7\mathbb{Z}$ logarithm $\log_3(6)$ which we've discovered is defined only up to an integer multiple of 6.

Let's look at another example.

Example 17.2 Find all $x \in \mathbb{Z}$, if any, that satisfy the given congruences.

(a) $2^x \equiv 5 \pmod{11}$.

(b) $3^x \equiv 5 \pmod{11}$.

Solution:

(a) The powers of 2 modulo 11 in order from 2^1 to 2^{10} are:

$$2,\ 4,\ 8,\ 5,\ 10,\ 9,\ 7,\ 3,\ 6,\ 1.$$

Here $e = 10$ is the smallest exponent $e \geq 1$ where $2^e \equiv 1$. So the powers of 2 modulo 11 will repeat in cycles of length 10.

Therefore, since we see that $2^4 \equiv 5$, it follows that the solutions to $2^x \equiv 5 \pmod{11}$ are all given by $x = 4 + 10k$ with $k \in \mathbb{Z}$. In more provocative terms:

$$\log_2(5) = 4 + 10k \quad (k \in \mathbb{Z}).$$

(b) The powers of 3 mod 11 in order from 3^1 to 3^5 are:

$$3,\ 9,\ 5,\ 4,\ 1.$$

This time $e = 5$ is the smallest exponent where $3^e \equiv 1$. So the powers of 3 mod 11 will repeat in cycles of length 5.

Since $3^3 \equiv 5$, it follows that the solutions to $3^x \equiv 5 \pmod{11}$ are all given by $x = 3+5k$ with $k \in \mathbb{Z}$. That is,

$$\log_3(5) = 3 + 5k \quad (k \in \mathbb{Z}).$$

Exercise 17.3 Find all $x \in \mathbb{Z}$ such that $4^x \equiv 5 \pmod{11}$.

It becomes apparent now that the smallest exponent $e \geq 1$ for which $a^e \equiv 1 \pmod{n}$ is of some importance, so let's give it a name.

Definition 17.4
Order, $\mathrm{ord}(a)$

Let $a \in \mathbb{Z}$ be coprime to n. The **order** of a modulo n, denoted by $\mathrm{ord}_n(a)$ or $\mathrm{ord}(a)$, is the smallest integer $e \geq 1$ such that $a^e \equiv 1 \pmod{n}$.

For example, our computations in Examples 17.1 and 17.2 show:

- In $(\mathbb{Z}/7\mathbb{Z})^\times$, $\mathrm{ord}_7(3) = 6$ and $\mathrm{ord}_7(4) = 3$.
- In $(\mathbb{Z}/11\mathbb{Z})^\times$, $\mathrm{ord}_{11}(2) = 10$ and $\mathrm{ord}_{11}(3) = 5$.

Some housekeeping: Since $a^{\varphi(n)} \equiv 1 \pmod{n}$ for all a coprime to n (by Theorem 12.4 (Euler's Theorem)), we know that there is *some* integer $e \geq 1$ for which $a^e \equiv 1 \pmod{n}$, so there must be a smallest such integer. So $\mathrm{ord}(a)$ is well-defined. Also note that if a is not coprime to n, then a is not a unit modulo n, so $a^e \not\equiv 1 \pmod{n}$ for all $e \geq 1$, so it makes no sense to speak of $\mathrm{ord}(a)$ in this case. Finally, if $a \equiv b \pmod{n}$ and if a and b are coprime to n, then $\mathrm{ord}(a) = \mathrm{ord}(b)$, so the order of a depends only on the congruence class of a in $(\mathbb{Z}/n\mathbb{Z})^\times$.

The following table lists sets of representatives of $(\mathbb{Z}/5\mathbb{Z})^\times$, $(\mathbb{Z}/7\mathbb{Z})^\times$ and $(\mathbb{Z}/10\mathbb{Z})^\times$ together with their orders.

a	values of a^e	$\mathrm{ord}_5(a)$
1	$\{1\}$	1
2	$\{2,4,3,1\}$	4
3	$\{3,4,2,1\}$	4
4	$\{4,1\}$	2

a	values of a^e	$\mathrm{ord}_7(a)$
1	$\{1\}$	1
2	$\{2,4,1\}$	3
3	$\{3,2,6,4,5,1\}$	6
4	$\{4,2,1\}$	3
5	$\{5,4,6,2,3,1\}$	6
6	$\{6,1\}$	2

a	values of a^e	$\mathrm{ord}_{10}(a)$
1	$\{1\}$	1
3	$\{3,9,7,1\}$	4
7	$\{7,9,3,1\}$	4
9	$\{9,1\}$	2

Exercise 17.5 Create a similar table for $(\mathbb{Z}/11\mathbb{Z})^\times$.

By studying these tables, we can make lots of conjectures. For instance, it appears that $\mathrm{ord}_n(a)$ is always a divisor of the order of $(\mathbb{Z}/n\mathbb{Z})^\times$—that is, it appears that $\mathrm{ord}_n(a) \mid \varphi(n)$. It also looks appears that $1, a, a^2, \ldots, a^{\mathrm{ord}_n(a)}$ are distinct mod n. Let's prove that both of these observations hold in general.

Proposition 17.6 Let $n \in \mathbb{Z}_{>0}$ and let $a \in \mathbb{Z}$ be coprime to n. If $a^k \equiv 1 \pmod{n}$ then $\mathrm{ord}_n(a) \mid k$. In particular, $\mathrm{ord}_n(a) \mid \varphi(n)$.

Proof: Apply the Remainder Theorem to write $k = \mathrm{ord}(a)q + r$ for some $q, r \in \mathbb{Z}$ with $0 \leq r < \mathrm{ord}(a)$. Then

$$1 \equiv a^k \equiv a^{\mathrm{ord}(a)q+r} \equiv (a^{\mathrm{ord}(a)})^q a^r \equiv a^r \pmod{n}.$$

It follows that $r = 0$ by minimality of $\mathrm{ord}(a)$, and hence $\mathrm{ord}(a) \mid k$. This proves the first part of the proposition. The second part follows from the first since $a^{\varphi(n)} \equiv 1 \pmod{n}$ by Euler's theorem. ■

Corollary 17.7 If $k = \text{ord}_n(a)$, then $1, a, a^2, \ldots, a^{k-1}$ are all distinct mod n.

Proof: If $a^i \equiv a^j \pmod{n}$ with $0 \leq i \leq j < k$ then $a^j(a^i)^{-1} \equiv 1 \pmod{n}$ hence $a^{j-i} \equiv 1 \pmod{n}$. So by the previous proposition, $k \mid j - i$. However, since $j - i < k$, this is only possible if $j - i = 0$. So if $i \neq j$ then $a^i \not\equiv a^j \pmod{n}$. ■

Example 17.8 Determine $\text{ord}_{17}(7)$.

Solution: Since $\varphi(17) = 16$, the only possible values for $\text{ord}_{17}(7)$ are 1, 2, 4, 8 and 16. We can compute

$$7^2 \equiv 15, \quad 7^4 \equiv 4 \quad \text{and} \quad 7^8 \equiv 16 \pmod{17}.$$

It follows that 7^{16} must be congruent to 1 mod 17, and thus $\text{ord}_{17}(7) = 16$.

Exercise 17.9 Determine $\text{ord}_{23}(5)$.

In the tables above (and in the previous example and exercise!), there always appears to be an integer a with $\text{ord}_n(a) = \varphi(n)$. For such an a, the set of values a^e mod n encompasses all of $(\mathbb{Z}/n\mathbb{Z})^\times$. This is of interest in the Discrete Logarithm Problem because it tells us that the equation $a^x \equiv b \pmod{n}$ has a solution for every b coprime to n. For instance, in Example 17.1, we saw that $\text{ord}_7(3) = 6$. Since $\varphi(7) = 6$, it follows that $3^x \equiv b \pmod{7}$ has a solution for all b. On the other hand, since $\text{ord}_7(4) = 3 < \varphi(7)$, there must be values of b for which $4^x \equiv b \pmod{7}$ has no solution, and indeed we saw that $b = 6$ is such a value.

Definition 17.10 **Primitive Root, Generator Mod n**

Let $a \in \mathbb{Z}$ be coprime to n. We say that a is a **primitive root** (or **generator**) modulo n if $\text{ord}_n(a) = \varphi(n)$.

For example:

- 2 and 3 are primitive roots modulo 5.
- 3 and 5 are primitive roots modulo 7.
- 3 and 7 are primitive roots modulo 10.

Exercise 17.11 Which integers are primitive roots modulo 11?

Alas, it is **not** true that there is always a primitive root mod n. Here is what we can say:

Theorem 17.12 There is a primitive root modulo n if and only if

$$n = 1,\ 2,\ 4,\ 2p^a,\ p^a$$

where p is an odd prime and $a \in \mathbb{Z}_{>0}$.

Exercise 17.13 Show that there is no primitive root mod 8.

Next lecture we will prove that primitive roots always exist mod p if p is a prime. We won't prove Theorem 17.12 in this course. However, let me mention that the following two facts are true:

- If g is a primitive root mod an odd prime p, then either g or $g+p$ is a primitive root mod p^a for all $a \geq 2$.
- If g is a primitive root mod p^a, then whichever of g and $g+p^a$ is odd will be a primitive root mod $2p^a$.

Let's close this lecture by returning to the Discrete Logarithm Problem. If g is a primitive root mod p, then for all b coprime to p, the equation $g^x \equiv b \pmod{p}$ has a solution with $x \in \mathbb{Z}$. You will prove in Problem 17.4 that any two solutions $x, x' \in \mathbb{Z}$ must be congruent mod $p-1$. Thus, there is a unique congruence class $[\ell] \in \mathbb{Z}/(p-1)\mathbb{Z}$ such that $g^\ell \equiv b \pmod{p}$. We denote this congruence class by $\log_g(b)$. This gives us a function

$$\log_g \colon (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{Z}/(p-1)\mathbb{Z}.$$

You will explore some properties of this function in the problem set below.

Lecture 17 Problems

▶ 17.1. Let p be an odd prime. Prove that $\operatorname{ord}(a) = 2$ if and only if $a \equiv -1 \pmod{p}$.

▶ 17.2. Let $a \in \mathbb{Z}$ be coprime to n, and let $k \in \mathbb{Z}_{>0}$. Prove:

$$\operatorname{ord}_n(a^k) = \frac{\operatorname{ord}_n(a)}{\gcd(\operatorname{ord}_n(a), k)}.$$

[**Hint:** Prove that the left-side divides the right, and vice versa. Proposition 17.6 will be helpful.]

17.3. Let $a, b \in \mathbb{Z}$ be coprime to n.

(a) Prove that if $\gcd(\operatorname{ord}_n(a), \operatorname{ord}_n(b)) = 1$ then $\operatorname{ord}_n(ab) = \operatorname{ord}_n(a) \cdot \operatorname{ord}_n(b)$.

(b) Give an example showing that $\operatorname{ord}_n(ab) \neq \operatorname{ord}_n(a) \cdot \operatorname{ord}_n(b)$ in general.

17.4. (a) Let $a \in \mathbb{Z}$ be coprime to n. Prove that $a^x \equiv a^y \pmod{n}$ if and only if $x \equiv y \pmod{\operatorname{ord}_n(a)}$.

(b) Let p be a prime and let g be a primitive root mod p. Prove that $g^x \equiv g^y \pmod{p}$ if and only if $x \equiv y \pmod{p-1}$.

▶ 17.5. Let p be a prime and let g be a primitive root mod p. Prove:

(a) $\log_g(1) \equiv 0 \pmod{p-1}$.

(b) $\log_g(ab) \equiv \log_g(a) + \log_g(b) \pmod{p-1}$ for all $a, b \in \mathbb{Z}$ coprime to p.

(c) $\log_g(a^k) \equiv k \log_g(a) \pmod{p-1}$ for all $a \in \mathbb{Z}$ coprime to p and all $k \in \mathbb{Z}$.
[**Note:** If $k < 0$ then a^k is to be interpreted as $(a^{-1})^{-k}$ mod p. So $a^{-2} = (a^{-1})^2$.]

17.6. Let p be a prime and let g be a primitive root mod p. Determine $\log_g(-1)$.

Lecture 18 Primitive Roots Mod p

Note: This material took up two in-class lectures. So what's here is really Lectures 18+19.

Our goal in this lecture is to prove that there is a primitive root modulo p if p is prime. (This is a special case of Theorem 17.12.) We will need some preliminary results concerning polynomial congruences.

Example 18.1 Find all solutions to the following congruences.

(a) $x^2 - 1 \equiv 0 \pmod{p}$, where p is a prime.

(b) $x^2 - 1 \equiv 0 \pmod{8}$.

Solution:

(a) If $x^2 - 1 \equiv 0 \pmod{p}$, then $p \mid (x^2 - 1) = (x-1)(x+1)$. So, by Euclid's Lemma, either $p \mid (x-1)$ or $p \mid (x+1)$. Thus, $x \equiv \pm 1 \pmod{p}$.

(b) We cannot use the same argument as in (a) since Euclid's Lemma no longer applies. If we try all possible values mod 8, we find that the solutions are given by $x \equiv \pm 1, \pm 3 \pmod{8}$.

Part (a) of the previous example shows that the polynomial $f(x) = x^2 - 1$ has at most 2 roots mod p. (In fact, it has one root mod 2 and two roots mod p if $p > 2$.) The next result is a generalization of this fact.

Theorem 18.2 Let $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ be a polynomial with integer coefficients and let p be a prime. Assume that $p \nmid a_n$. Then the congruence

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

has at most n distinct solutions mod p.

Proof: We proceed by induction on $n = \deg f(x)$.

If $n = 1$ then we are reduced to the linear congruence $a_1 x \equiv -a_0 \pmod{p}$ which, by Theorem 10.3, has exactly one solution since $\gcd(p, a_1) = 1$.

The proof of the inductive step will make use of a familiar fact from elementary algebra: if $x = a$ is a solution of $f(x) = 0$, then we can write $f(x) = (x-a)g(x)$ where $\deg g(x) = n-1$.

Assume now that the result is true for all polynomials of degree $\leq n-1$. If $f(x)$ has no roots, we're done. Otherwise, suppose that $f(a) \equiv 0 \pmod{p}$. Then

$$f(x) \equiv f(x) - f(a) \equiv a_1(x-a) + \cdots + a_n(x^n - a^n) \pmod{p}.$$

By applying the identity

$$x^k - a^k = (x-a)(x^{k-1} + x^{k-2}a + \cdots + xa^{k-2} + a^{k-1})$$

to each term above, we find that

$$f(x) \equiv (x-a)g(x) \pmod{p}$$

where $g(x)$ is a polynomial of degree $n-1$. So if $f(x) \equiv 0 \pmod{p}$ then $p \mid (x-a)g(x)$ and hence either

$$x - a \equiv 0 \pmod{p} \quad \text{or} \quad g(x) \equiv 0 \pmod{p}$$

by Euclid's Lemma. Also note that the leading coefficient of $g(x)$ is not divisible by p since otherwise the same would be true of $f(x) \equiv (x-a)g(x)$ contrary to assumption. So, by the inductive hypothesis, $g(x) \equiv 0 \pmod{p}$ has at most $n-1$ solutions. Thus, the original congruence $f(x) \equiv 0 \pmod{p}$ has at most n solutions (namely, $x \equiv a$ and the solutions of $g(x) \equiv 0 \pmod{p}$). ■

Warning: This theorem is false if the modulus is not prime. Indeed, in Example 18.1 we saw that the quadratic congruence $x^2 - 1 \equiv 0 \pmod{8}$ has four solutions.

Theorem 18.3 Let p be a prime and let $d \in \mathbb{Z}_{>0}$ be a divisor of $p-1$. Then the congruence

$$x^d - 1 \equiv 0 \pmod{p}$$

has exactly d solutions.

Proof: Write $p - 1 = dk$. Then

$$x^{p-1} - 1 = (x^d)^k - 1 = (x^d - 1)(\underbrace{x^{d(k-1)} + x^{d(k-2)} + \cdots + x^d + 1}_{g(x)}).$$

By Fermat's Little Theorem, $x^{p-1} - 1$ has exactly $p-1$ roots mod p. By Euclid's Lemma, each of these roots is either a root of $x^d - 1$ or of $g(x)$; conversely, any root of $x^d - 1$ and $g(x)$ must be a root of $x^{p-1} - 1$. By Theorem 18.2, these latter two polynomials have at most d and $dk - d$ roots, respectively. It follows that they must each have *exactly* d and $dk - d$ roots since otherwise $x^{p-1} - 1$ would have $< d + dk - d = p - 1$ roots. ■

We can now prove the existence of primitive roots mod p. It's instructive to see how the argument works in a couple of examples.

Example 18.4 Prove that there is a primitive root mod 11.

Solution: We want to show that there is an integer a such that $\text{ord}_{11}(a) = \varphi(11)$. The possible orders mod 11 are the divisors of $\varphi(11) = 10$, namely 1, 2, 5 and 10.

If $\text{ord}_{11}(a) = d$, then $x = a$ is a solution to $x^d - 1 \equiv 0 \pmod{11}$. Thus, by Theorem 18.3, there are at most 1, 2 and 5 elements of order 1, 2 and 5, respectively. This accounts for 1+2+5=8 of the classes in $(\mathbb{Z}/11\mathbb{Z})^\times$, leaving us with two classes whose orders must therefore be 10. These are our desired primitive roots.

Everything worked out nicely in the previous example! We won't always be so lucky, as the next example shows.

Example 18.5 Prove that there is a primitive root mod 13.

Solution: We want to show that there is an integer a such that $\mathrm{ord}_{13}(a) = \varphi(13)$. The possible orders mod 13 are the divisors of $\varphi(13) = 12$, namely 1, 2, 3, 4, 6 and 12.

If $\mathrm{ord}_{13}(a) = d$, then $x = a$ is a solution to $x^d - 1 \equiv 0 \pmod{13}$. Thus, by Theorem 18.3, there are at most 1, 2, 3, 4 and 6 elements of order 1, 2, 3, 4 and 6, respectively. This accounts for 1+2+3+4+6=16 of the classes in $(\mathbb{Z}/13\mathbb{Z})^\times$. Oops—we've over-counted!

Let's count a bit more carefully. The elements of order 6 satisfy the equation $x^6 - 1 \equiv 0 \pmod{13}$, but so do the elements of orders 2 and 3. In fact, if $\mathrm{ord}(g) = 6$ then $1, g, g^2, g^3, g^4, g^5$ are all distinct and they all satisfy $x^6 - 1 \equiv 0 \pmod{13}$ (since $(g^i)^6 = (g^6)^i \equiv 1^i \pmod{13}$). Thus, these six roots must be all of the roots of $x^6 - 1$ mod 13. On the other hand, since

$$\mathrm{ord}(g^i) = \frac{\mathrm{ord}(g)}{\gcd(i, \mathrm{ord}(g))}$$

by Problem 17.2 we see that of these roots, only g and g^5 have order 6.

Likewise, by considering the equation $x^4 - 1 \equiv 0 \pmod{13}$, we can no see that of its roots, at most 2 will have order 4 (in fact, exactly 2 will have order 4).

So now we've accounted for $1 + 2 + 3 + 2 + 2 = 10$ of the elements of $(\mathbb{Z}/13\mathbb{Z})^\times$, leaving us with 2 elements whose orders must therefore be 12. These are primitive roots.

We can do better still. There is exactly only 1 element of order 2 (of the roots $x = \pm 1$ of $x^2 - 1$, only $x = -1$ has order 2), and only 2 elements of order 3 (why?). So in fact there are $12 - (1 + 1 + 2 + 2 + 2) = 4$ primitive roots mod 13.

Exercise 18.6 Show, by an argument similar to the one in the previous two examples, that there is a primitive root mod 19.

The counting argument used above works in general.

Theorem 18.7 Let p be a prime and let $d \mid \varphi(p)$. Then the number of congruence classes in $(\mathbb{Z}/p\mathbb{Z})^\times$ of order d is equal to $\varphi(d)$.

Proof: Let N_d be the number of classes in $(\mathbb{Z}/p\mathbb{Z})^\times$ of order d. Since every class in $(\mathbb{Z}/p\mathbb{Z})^\times$ has order dividing $\varphi(p)$, and since there are $\varphi(p) = p - 1$ classes in total, we see that

$$p - 1 = \sum_{d \mid p-1} N_d.$$

Let's prove that $N_d \leq \varphi(d)$. If there are no classes of order d, then $N_d = 0$, and we're done. Otherwise, suppose a has order d. Then the d roots of $x^d - 1 \equiv 0 \pmod{p}$ are given by $1, a, \ldots, a^{d-1}$. Of these, since $\mathrm{ord}(a^i) = d/\gcd(d, i)$ (by Problem 17.2), only the a^i with i coprime to d have order d. Thus, $N_d = \varphi(d)$ since there are $\varphi(d)$ such i.

Consequently,

$$p - 1 = \sum_{d \mid p-1} N_d \leq \sum_{d \mid p-1} \varphi(d).$$

On the other hand, by Theorem 15.9,

$$\sum_{d|p-1} \varphi(d) = p - 1.$$

It follows that the inequality above is an equality. Looking at our argument, a strict inequality results only if $N_d < \varphi(d)$ for some d. So it must be the case that $N_d = \varphi(d)$ for all $d \mid p-1$, which is precisely what we wanted to prove. ■

Applying this theorem to the case where $d = \varphi(p)$, we obtain:

Corollary 18.8 **(Existence of Primitive Roots Mod p)**

Let p be a prime. There are $\varphi(p-1)$ distinct primitive roots mod p.

The following table illustrates the previous corollary.

p	$\varphi(p-1)$	primitive roots mod p
2	1	1
3	1	2
5	2	2, 3
7	2	3, 5
11	4	2, 6, 7, 8
13	4	2, 6, 7, 11
17	8	3, 5, 6, 7, 10, 11, 12, 14
19	6	2, 3, 10, 13, 14, 15

How do we find primitive roots in practice? The counting method in our existence proof above is horribly inefficient. Is there a method better than picking a random a and computing all the powers a^d where $d \mid p-1$? The following result shows that we don't have to compute a^d for *all* divisors of $p-1$.

Proposition 18.9 Let p be prime and let $a \in \mathbb{Z}$ be coprime to p. Then a is a primitive root mod p if and only if $a^{(p-1)/q} \not\equiv 1 \pmod{p}$ for all prime divisors q of $p-1$.

Proof: If $a^{(p-1)/q} \equiv 1 \pmod{p}$, then $\text{ord}(a) < p-1$ so a cannot be a primitive root. Conversely, assume that $a^{(p-1)/q} \not\equiv 1 \pmod{p}$ for all prime divisors q of $p-1$. Let $k = \text{ord}(a)$. We know that $k \mid p-1$. If $k \neq p-1$ then $(p-1)/k$ has a prime divisor q. But then $(p-1)/(kq)$ is an integer, and therefore

$$a^{(p-1)/q} = (a^k)^{(p-1)/qk} \equiv 1^{(p-1)/qk} \equiv 1 \pmod{p},$$

which is a contradiction. Thus, $k = p-1$ and so a is a primitive root mod p. ■

So here is how we can find a primitive root mod p: First, find the prime factors q of $p-1$. Next, pick a small integer a in the interval $2 \leq a < p$. Finally, compute $a^{(p-1)/q}$ mod p for all q. If none of these are 1, then you've found a primitive root. If one of them is 1, then repeat the same process with $a+1$. This actually works decently well in practice.[14]

[14] It's a polynomial time algorithm... if you assume the generalized Riemann Hypothesis.

Example 18.10 Let $p = 19$ and $a = 2$. Then $p - 1 = 18 = 2 \cdot 3^2$. Now compute:

$$2^{(p-1)/2} = 2^9 \equiv 18 \pmod{19}$$
$$2^{(p-1)/3} = 2^6 \equiv 7 \pmod{19}.$$

Since neither is congruent to 1, 2 must be a primitive root mod 19.

Exercise 18.11 Find a primitive root mod $p = 31$.

REMARK (Primitive Root Mysteries)

There are many open problems concerning primitive roots. For example, how many primes p have 2 as a primitive root? Of the primes ≤ 100, the following twelve primes do:

$$3,\ 5,\ 11,\ 13,\ 19,\ 29,\ 37,\ 53,\ 59,\ 61,\ 67,\ 83.$$

Since there are 25 primes ≤ 100, we see that 48% of them admit 2 as a primitive root. Approximately 40% of the primes $\leq 10^3$, 38% of the primes $\leq 10^4$, and 38% of the primes $\leq 10^5$ admit 2 as a primitive root.

Artin's Conjecture (1927): There are infinitely many primes that admit 2 as a primitive root. More precisely, the proportion of primes $\leq x$ that admit 2 as a primitive root tends to

$$\prod_{p \text{ prime}} \left(1 - \frac{1}{p(p-1)}\right) \approx 0.3739...$$

as $x \to \infty$.

Assuming a general form of the Riemann Hypothesis, Hooley was able to prove Artin's Cojecture in 1967. However, since the Riemann Hypothesis is still an open problem, so too is Artin's Conjecture.

If we ignore results that rely on the Riemann Hypothesis, we actually do not know of a single, specific integer a that is a primitive root modulo infinitely many primes. However, there is a remarkable result due to Gupta, M.R. Murty and Heath-Brown that says: Every prime number—with at most two exceptions—is a primitive root modulo infinitely many primes. So one of 2, 3 and 5 is definitely a primitive root for infinitely many primes, but we do not know which one!

Lecture 18 Problems

18.1. Let p be a prime.

(a) Prove that if g is a primitive root mod p then every primitive root mod p is congruent to exactly one integer in the set $\{g^i : 0 \leq i \leq p-1,\ \gcd(i, p-1) = 1\}$.

(b) Given that 2 is a primitive root mod 19, find all of the primitive roots mod 19.

▶ 18.2. Let p be a prime such that $p \equiv 1 \pmod{4}$.

(a) Prove that there exists an a such that $\text{ord}_p(a) = 4$.

(b) Prove that the congruence $x^2 + 1 \equiv 0 \pmod{p}$ has a solution. [**Note:** This shows that $\sqrt{-1} \in \mathbb{Z}/p\mathbb{Z}$ if $p \equiv 1 \pmod{4}$.]

▶ 18.3. Let p be a prime such that $p \equiv 3 \pmod{4}$. Prove that the congruence $x^2 + 1 \equiv 0 \pmod{p}$ has no solutions. [**Note:** This shows that $\sqrt{-1} \notin \mathbb{Z}/p\mathbb{Z}$ if $p \equiv 3 \pmod{4}$.]

18.4. Let p be a prime, and let

$$f(x) = (x-1)(x-2)\cdots(x-(p-1)) - (x^{p-1} - 1).$$

(a) What is the degree of $f(x)$?

(b) Show that $f(a) \equiv 0 \pmod{p}$ for all $a \in \mathbb{Z}$ coprime to p.

(c) Deduce from parts (a) and (b) that all of the coefficients of $f(x)$ are divisible by p.

(d) Use part (c) to deduce **Wilson's Theorem**:

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

18.5. Let p be a prime, and let g be a primitive root mod p.

(a) Prove that if p is odd, then $g^{(p-1)/2} \equiv -1 \pmod{p}$. [**Hint:** What is $\left(g^{(p-1)/2}\right)^2 \pmod{p}$?]

(b) Prove that $(p-1)! \equiv g^{p(p-1)/2} \pmod{p}$.

(c) Use parts (a) and (b) to give another proof of Wilson's Theorem (see previous problem).

Lecture 19 Applications of Primitive Roots

Solving Congruences

Let's begin by recalling the definition of the discrete logarithm mod p.

If g is a primitive root mod p then the $\{1, g, g^2, \ldots, g^{p-2}\}$ is a complete set of representatives for the unit group $(\mathbb{Z}/p\mathbb{Z})^\times$. Thus, for every $b \in \mathbb{Z}$ coprime to p, the congruence

$$g^x \equiv b \pmod{p}$$

has a unique solution $x = \ell \in \{0, 1, 2, \ldots, p-2\}$. We denote this unique solution by $\log_g(b)$ and we view it as a congruence class in $\mathbb{Z}/(p-1)\mathbb{Z}$. We've thus defined a function

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

which we call the **discrete logarithm** mod p to the base g. (Some textbooks call this the **index** and denote it by ind_g.) Here are some of its key properties, which you had already proved in Problem 17.5.

Proposition 19.1 Let p be a prime and let g be a primitive root mod p. Then:

(a) $\log_g(1) \equiv 0 \pmod{p-1}$.

(b) $\log_g(ab) \equiv \log_g(a) + \log_g(b) \pmod{p-1}$ for all $a, b \in \mathbb{Z}$ coprime to p.

(c) $\log_g(a^k) \equiv k \log_g(a) \pmod{p-1}$ for all $a \in \mathbb{Z}$ coprime to p and all $k \in \mathbb{Z}$.

REMARK

All of the above works equally well for any modulus n that admits a primitive root g. We can in the same way define a function

$$\log_g \colon (\mathbb{Z}/n\mathbb{Z})^\times \to \mathbb{Z}/\varphi(n)\mathbb{Z}$$

that satisfies the same congruences given in Proposition 19.1 except now they are all taken mod $\varphi(n)$. If you know some group theory, you will recognize that part (b) says that $\log_g$ is a group homomorphism; it is, in fact, an isomorphism.

Example 19.2 I claim that $g = 2$ is a primitive root mod $p = 13$. I will actually confirm this by computing all the powers of 2, but just as a refresher, let's also see how this quickly follows from Proposition 18.9: We have $p - 1 = 12 = 2^2 \cdot 3$, and since $2^{(p-1)/3} \equiv 3 \pmod{13}$ and $2^{(p-1)/2} \equiv 6 \pmod{13}$, we can safely conclude that 2 is a primitive mod 13.

Here are the powers of 2 mod 13:

$$2^0 \equiv 1, \quad 2^1 \equiv 2, \quad 2^2 \equiv 4, \quad 2^3 \equiv 8, \quad 2^4 \equiv 3, \quad 2^5 \equiv 6,$$

$$2^6 \equiv 12, \quad 2^7 \equiv 11, \quad 2^8 \equiv 9, \quad 2^9 \equiv 5, \quad 2^{10} \equiv 10, \quad 2^{11} \equiv 7.$$

So, for instance, we have that

$$\log_2(1) = 0, \quad \log_2(2) = 1, \quad \log_2(4) = 2, \quad \log_2(8) = 3, \quad \log_2(3) = 4, \quad \text{etc.}$$

We can record this information in a table:

x	1	2	3	4	5	6	7	8	9	10	11	12
$\log_2(x)$	0	1	4	2	9	5	11	3	8	10	7	6

Exercise 19.3 Create a table containing the values of $\log_3(x)$ for $x \in (\mathbb{Z}/7\mathbb{Z})^\times$.

In elementary algebra we can use logarithms to solve equations of the form $ax^b = c$. For example,

$$11x^5 = 4 \implies \log(11) + 5\log(x) = \log(4) \implies \log(x) = \log\left(\frac{4}{11}\right)^{1/5} \implies x = \left(\frac{4}{11}\right)^{1/5}.$$

We can do the same with the discrete logarithm!

Example 19.4 Find all solutions to the congruence $11x^5 \equiv 4 \pmod{13}$.

Solution: We will use $\log_g$ with the primitive root $g = 2$ mod 13. We find that

$$\log_2(11) + 5\log_2(x) \equiv \log_2(4) \pmod{12}$$

(Warning: The discrete logarithm goes from $(\mathbb{Z}/p\mathbb{Z})^\times$ to $\mathbb{Z}/(p-1)\mathbb{Z}$!) Referring to our table in the previous example, we see that $\log_2(11) = 7$ and $\log_2(4) = 2$, so the above congruence simplifies to

$$7 + 5\log_2(x) \equiv 2 \pmod{12} \implies 5\log_2(x) \equiv -5 \pmod{12}.$$

Since 5 is a unit mod 12, we can multiply through by 5^{-1} mod 12 to obtain

$$\log_2(x) \equiv -1 \equiv 11 \pmod{12}.$$

Thus, $x \equiv 2^{11} \equiv 7 \pmod{13}$, where 2^{11} was determined using our log table.

Exercise 19.5 Use $\log_3$ to find all solutions to the congruence $5x^{46} \equiv 6 \pmod{7}$.

Example 19.6 Prove that the Diophantine equation $5x^{16} - 13y^{16} = 3$ has no solutions in the integers.

Solution: Reducing mod 13, we arrive at the congruence

$$5x^{16} \equiv 3 \pmod{13}.$$

Applying $\log_2$ to both sides and setting $\ell = \log_2(x)$, we get

$$\log_2(5) + 16\log_2(x) \equiv \log_2(3) \pmod{12} \implies 16\ell \equiv 4 - 9 = 7 \pmod{12}.$$

This linear congruence in ℓ will have a solution if and only if $\gcd(16, 12) \mid 7$. Since $\gcd(16, 12) = 4$ does not divide 7, it follows there is no solution for $\ell = \log_2(x) \bmod 12$ and hence no solution for $x \bmod 13$. So there can be no solutions to the original Diophantine equation in the integers.

Of course, we didn't *have to* use $\log_2$ to deal with the congruence mod 13 in previous example. For instance, we could have used Fermat's Little Theorem to deduce that $x^{16} = x^{13}x^3 \equiv x \cdot x^3 \pmod{13}$, and we could have computed the inverse of 5 mod 13 to be 8, and thus we would have arrived at

$$5x^{16} \equiv 3 \pmod{13} \quad \Longrightarrow \quad x^4 \equiv 5^{-1} \cdot 3 \equiv 8 \cdot 3 \equiv 11 \pmod{13}.$$

Now it's just a matter of checking whether 11 is a 4th power mod 13, which we can do by inspection. The discrete logarithm just makes this process a bit more efficient. In fact, this brings us to our next topic...

Power Residues

Let's look at congruences of the form $x^k \equiv a \pmod{p}$, where p is a prime, $a \not\equiv 0 \pmod{p}$ and $k \geq 2$. The terminology below is due to Gauss.

Definition 19.7

kth Power Residue and Nonresidue, Quadratic Residue, Cubic Residue

Let $a \in \mathbb{Z}$ be coprime to p.

If the congruence $x^k \equiv a \pmod{p}$ has a solution, then we say that a is a **kth power residue** mod p. Otherwise, if the congruence has no solutions, then we say that a is a **kth power nonresidue** mod p.

In the special case where $k = 2$, we say respectively that a is a **quadratic** residue or nonresidue mod p. If $k = 3$, then we say **cubic** residue and nonresidue.

For example, 1 and 4 are quadratic residues mod 5, since $1 \equiv 1^2 \pmod{5}$ and $4 \equiv 2^2 \pmod{5}$, while 3 is a quadratic non-residue since there are no solution to $x^2 \equiv 3 \pmod{5}$ as we can check by inspection.

Exercise 19.8 Find all cubic residues and nonresidues mod 7.

The study of quadratic residues, and more generally kth power residues, was historically a major driving force in the development of number theory. In the next several lectures we will investigate the properties of quadratic residues. For now, we close this lecture with the following important characterization of kth power residues.

Theorem 19.9 (kth Power Residue Criterion)

Let p be a prime and let $a \in \mathbb{Z}$ be coprime to p. Then a is a kth power residue mod p if and only if

$$a^{(p-1)/d} \equiv 1 \pmod{p} \quad \text{where} \quad d = \gcd(k, p-1).$$

Furthermore, if a is a kth power residue, then the congruence $x^k \equiv a \pmod{p}$ has exactly d distinct solutions mod p.

Proof: Let g be a primitive root mod p. Then a is a kth power residue if and only if the congruence

$$x^k \equiv a \pmod{p} \qquad (*)$$

has a solution. Taking logs and setting $\ell = \log_g(x)$, we see that $(*)$ has a solution if and only if the linear congruence

$$k\ell \equiv \log_g(a) \pmod{p-1} \qquad (**)$$

has a solution which is the case if and only if $d = \gcd(k, p-1) \mid \log_g(a)$ by Theorem 10.3. Now,

$$\begin{aligned} d \mid \log_g(a) &\iff \log_g(a) = dk \text{ for some } k \in \mathbb{Z} \\ &\iff \frac{p-1}{d}\log_g(a) = (p-1)k \text{ for some } k \in \mathbb{Z} \\ &\iff \frac{p-1}{d}\log_g(a) \equiv 0 \pmod{p-1} \\ &\iff \log_g\left(a^{(p-1)/d}\right) \equiv 0 \pmod{p-1} \\ &\iff a^{(p-1)/d} \equiv 1 \pmod{p}. \end{aligned}$$

This proves the first part of the theorem. For the last part, simply note that every solution to $(*)$ gives a solution to $(**)$ and vice versa. By Theorem 10.3, if $(**)$ has solutions, then it has exactly d distinct solutions. ■

The special case $k = 2$ is a famous result due to Euler.

Corollary 19.10 **(Euler's Criterion—First Form)**

Let p be an odd prime and let $a \in \mathbb{Z}$ be coprime to p. Then

$$a^{(p-1)/2} \equiv \begin{cases} 1 \pmod{p} & \text{if } a \text{ is a quadratic residue mod } p \\ -1 \pmod{p} & \text{if } a \text{ is a quadratic nonresidue mod } p. \end{cases}$$

Proof: Since p is odd, $\gcd(p-1, 2) = 2$. So a will be a quadratic residue if and only if $a^{(p-1)/2} \equiv 1 \pmod{p}$. Now observe that

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

So $a^{(p-1)/2}$ is a root of the polynomial $x^2 - 1$ mod p, hence the only possible values of $a^{(p-1)/2}$ mod p are ± 1 (why?). Since the value 1 corresponds to the case where a is a quadratic residue, the value -1 must therefore correspond to the nonresidue case. ■

Example 19.11 Determine whether 11 is a 4th power residue mod 13.

Solution: Here $d = \gcd(4, 13-1) = 4$, and

$$11^{(13-1)/4} = 11^3 \equiv 5 \pmod{13}.$$

Since $11^{(13-1)/d} \not\equiv 1 \pmod{13}$, it follows that 11 is not a 4th power mod 13.

Lecture 19 Problems

▶ 19.1. Prove that there are exactly $(p-1)/d$ distinct kth power residues mod p, where $d = \gcd(k, p-1)$.

19.2. Prove that if $p \equiv 2 \pmod 3$, then all integers a coprime to p are cubic residues mod p.

19.3. Let p be an odd prime. Prove that if -1 is a 4th power residue mod p then $p \equiv 1 \pmod 8$.

19.4. Prove that there are infinitely many primes of the form $8k+1$. [**Hint:** Suppose $p_1, \ldots, p_n$ are all such primes and consider $N = (p_1 \cdots p_n)^4 + 1$. The previous problem can help.]

Lecture 20 Quadratic Residues

We've learned how to solve linear congruences $ax \equiv b \pmod{n}$. We can test for solvability using Theorem 10.3 and we can find solutions using the Euclidean algorithm.

Naturally, the next thing to do is to look at quadratic congruences. For simplicity, we'll work modulo a prime p. Consider the congruence

$$ax^2 + bx + c \equiv 0 \pmod{p}$$

where $a \not\equiv 0 \pmod{p}$. Let's assume that $p > 2$ (the case $p = 2$ is trivial), so that 2 is invertible mod p. Since a is also invertible mod p, we can complete the square:

$$a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a} \equiv 0 \pmod{p}.$$

Setting $y = x + b/2a$ and re-arranging, we arrive at

$$y^2 \equiv \frac{b^2 - 4ac}{4a^2} \pmod{p}.$$

The question now is how to solve this for y. Since any quadratic congruence can be brought into this form, we may as well just start with

$$x^2 \equiv a \pmod{p}.$$

You will now recognize (see Definition 19.7) that our question essentially is: When is a a *quadratic residue* mod p?

Definition 20.1
Legendre Symbol

Let p be an odd prime, and let $a \in \mathbb{Z}$. The **Legendre symbol** is defined by

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{p} \\ 1 & \text{if } a \text{ is a quadratic residue mod } p \\ -1 & \text{if } a \text{ is a quadratic nonresidue mod } p \end{cases}$$

$\left(\frac{a}{p}\right)$ is read as "a on p" and occasionally written as $(a|p)$.

For example,

$$\left(\frac{a}{5}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{5} \\ 1 & \text{if } a \equiv 1, 4 \pmod{5} \\ -1 & \text{if } a \equiv 2, 3 \pmod{5}. \end{cases}$$

Exercise 20.2

Determine $\left(\frac{a}{7}\right)$.

The Legendre symbol has many interesting properties. A convenient tool for demonstrating these properties is Corollary 19.10 which we re-state here in terms of the Legendre symbol.

Theorem 20.3 **(Euler's Criterion—Second Form)**

Let p be an odd prime and let $a \in \mathbb{Z}$ be coprime to p. Then

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

We can use Euler's criterion to determine $\left(\frac{a}{p}\right)$ without having to solve $x^2 \equiv a \pmod{p}$.

Example 20.4 Determine $\left(\frac{5}{17}\right)$.

Solution: Using Euler's criterion, we have

$$\left(\frac{5}{17}\right) \equiv 5^{(17-1)/2} \equiv 5^8 \equiv -1 \pmod{17}.$$

Since $\left(\frac{5}{17}\right)$ is either 1 or -1, it follows that $\left(\frac{5}{17}\right) = -1$ (an actual equality and not just a congruence mod 17).

Exercise 20.5 Determine $\left(\frac{8}{17}\right)$.

The next result presents some basic properties of the Legendre symbol.

Proposition 20.6 Let p be an odd prime and let $a, b \in \mathbb{Z}$ be coprime to p.

(a) If $a \equiv b \pmod{p}$ then $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

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

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

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

Proof:

(a) If $a \equiv b \pmod{p}$ then $x^2 \equiv a \pmod{p}$ has a solution if and only if $x^2 \equiv b \pmod{p}$ has a solution. So $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$.

(b) By Euler's criterion,

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

Since $\left(\frac{ab}{p}\right)$, $\left(\frac{a}{p}\right)$ and $\left(\frac{b}{p}\right)$ are each ± 1, it follows that

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

(c) a^2 is obviously a quadratic residue mod p.

(d) By Euler's criterion,

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

If $p \equiv 1 \pmod 4$ then $(p-1)/2$ is even, so $(-1)^{(p-1)/2} = -1$. If $p \equiv -1 \pmod 4$, then $(p-1)/2$ is odd, so $(-1)^{(p-1)/2} = -1$. ■

REMARK

If you know some group theory, you will recognize that parts (a) and (b) of Proposition 20.6 imply that the map

$$\left(\frac{\cdot}{p}\right) : (\mathbb{Z}/p\mathbb{Z})^\times \to \{\pm 1\}$$

defined by sending a to $\left(\frac{a}{p}\right)$ is a group homomorphism. The kernel of this map is the set (subgroup!) of quadratic residues mod p. The quadratic nonresidues are a coset of the quadratic residues in $(\mathbb{Z}/p\mathbb{Z})^\times$.

Our calculation of $\left(\frac{-1}{p}\right)$ can be used to show that a certain type of Diophantine equation has no solutions.

Example 20.7 Show that the equation $y^2 = x^3 - 5$ has no solutions in the integers.

Solution: Suppose for a contradiction that there are integer solutions. Reducing mod 4, the equation becomes $y^2 \equiv x^3 - 1 \pmod 4$. Since y^2 is either 0 or 1 mod 4, it follows that either $x^3 \equiv 1 \pmod 4$ or $x^3 \equiv 2 \pmod 4$. By inspection, the latter is impossible. Thus, $x^3 \equiv 1 \pmod 4$ and hence $x \equiv 1 \pmod 4$ (again, by inspection).

Now re-write the original equation $y^2 = x^3 - 5$ as

$$y^2 + 4 = x^3 - 1 = (x-1)(x^2 + x + 1).$$

Since $x \equiv 1 \pmod 4$, it follows that $x^2 + x + 1 \equiv 3 \pmod 4$. Also, $x^2 + x + 1$ is positive, odd and not equal to 1 (why?) hence it can be factored into a product of odd primes. If all prime factors were 1 mod 4, then $x^2 + x + 1$ would be 1 mod 4 too, but it's not. So $x^2 + x + 1$ must have a prime divisor p congruent to 3 mod 4. As $x^2 + x + 1 \equiv 0 \pmod p$, it follows that

$$y^2 + 4 \equiv 0 \pmod{p}.$$

Re-arranging and using the fact that 2 is invertible mod p (since p is odd), we can re-write this as

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

This shows that -1 is a quadratic residue mod p, that is, $\left(\frac{-1}{p}\right) = 1$. Since $p \equiv 3 \pmod 4$, this contradicts Proposition 20.6(d).

Exercise 20.8 Show that the equation $y^2 = x^3 + 11$ has no solutions in the integers.

REMARK (The Local-to-Global Principle Fails)

Our main tool thus far for showing that a Diophantine equation has no solutions is to reduce the equation modulo a cleverly chosen n and show that there are no solutions mod n. In a remark on page 53 we considered the converse to this process. Namely, if a Diophantine equation has solutions modulo n for every n, must it have a solution in the integers?

Alas, it's now time for me to break the bad news: The answer is *no*.

It can be shown that the equations in Example 20.7 and Exercise 20.8 have solutions mod n for every n, but as we just saw, they do not have integer solutions. Thus, these equations have solutions "locally everywhere" but not "globally."

Proposition 20.6 allows us to speed up Legendre symbol calculations. For instance, we have

$$\left(\frac{198}{23}\right) = \left(\frac{22\cdot 9}{23}\right) = \left(\frac{22}{23}\right)\left(\frac{9}{23}\right) = \left(\frac{-1}{23}\right)\left(\frac{3^2}{23}\right) = (-1)(1) = -1.$$

(From this we are able to conclude that the congruence $x^2 \equiv 198 \pmod{23}$ has no solutions without doing much work!)

More generally, if $a \in \mathbb{Z}$ has prime factorization $a = \pm q_1^{a_1}\cdots q_k^{a_k}$ then

$$\left(\frac{a}{p}\right) = \left(\frac{\pm 1}{p}\right)\left(\frac{q_1}{p}\right)^{a_1}\cdots\left(\frac{q_k}{p}\right)^{a_k}.$$

Since we've determined $\left(\frac{-1}{p}\right)$ in Proposition 20.6(c), and since $\left(\frac{p}{p}\right) = 0$, our next task is to determine $\left(\frac{q}{p}\right)$ where q is a prime different from p. We will take this up next lecture. However, I will leave you with this important exercise...

Exercise 20.9 Make some tables containing values of $\left(\frac{p}{q}\right)$ and $\left(\frac{q}{p}\right)$ where p and q are distinct primes.

Do you notice anything interesting? Make some conjectures and try to see if you can discover the results that we're going to discuss next time!

Lecture 20 Problems

20.1. Prove:

(a) The product of two quadratic residues is a quadratic residue.

(b) The product of two nonquadratic residues is a quadratic residue.

(c) The product of a quadratic residue and a quadratic nonresidue is a quadratic nonresidue.

20.2. Let p be an odd prime. Prove that there are $(p-1)/2$ quadratic residues and $(p-1)/2$ quadratic nonresidues mod p.

20.3. Let p be an odd prime.

(a) Show that the number of solutions to the congruence $x^2 \equiv a \pmod{p}$ is $1 + \left(\frac{a}{p}\right)$.

(b) Assume $a \not\equiv 0 \pmod{p}$. Show that the number of solutions to the congruence $ax^2 + bx + c \equiv 0 \pmod{p}$ is $1 + \left(\frac{b^2-4ac}{p}\right)$.

20.4. Let p and q be odd primes and suppose that $p = 4q + 1$. Prove that 2 is a primitive root mod p.

20.5. Let p be an odd prime and let g be a primitive root mod p. Prove that $\left(\frac{g}{p}\right) = -1$.

20.6. Let p be an odd prime and let $a \in \mathbb{Z}$ be coprime to p. Prove:

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

20.7. Let p be an odd prime and let $a \in \mathbb{Z}$ be coprime to p. Prove that the congruence $x^2 \equiv a \pmod{p^n}$ has a solution for all $n \in \mathbb{Z}_{>0}$ if and only if a is a quadratic residue mod p.

[**Hint:** Suppose $x = x_0$ is a solution to $x^2 \equiv a \pmod{p^n}$. Argue that there is some $m \in \mathbb{Z}$ such that $x = x_0 + mp^n$ is a solution to $x^2 \equiv a \pmod{p^{n+1}}$.]

Lecture 21 Quadratic Reciprocity

The fundamental theorem must certainly be regarded as one of the most elegant of its type.

– C.F. Gauss, *Disquisitiones Arithmeticae*

We now come to one of the deepest theorems in all of elementary number theory.

Theorem 21.1 **(Law of Quadratic Reciprocity)**

Let p and q be distinct odd primes. Then:

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

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

Exercise 21.2 Show that the statements of Theorem 21.1 are equivalent to:

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

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

The "reciprocity" in Theorem 21.1 refers to the (unexpected) relation between $\left(\frac{q}{p}\right)$ and $\left(\frac{p}{q}\right)$. Recall that $\left(\frac{a}{p}\right)$ has to do with the solvability of the equation $x^2 - a \equiv 0 \pmod p$. There is no reason to expect the behavior of the polynomial $x^2 - q \bmod p$ to have such a strong influence on the behavior of $x^2 - p \bmod q$. The fact that it *does* have an influence is one of the profound mysteries of number theory.

The first traces of the law of quadratic reciprocity are in the work of Fermat, who essentially discovered that the value $\left(\frac{-1}{p}\right)$ depended on the congruence class of $p \bmod 4$. In trying to extend Fermat's work, Euler noticed that the value of $\left(\frac{a}{p}\right)$ seems to depend on the congruence class of $p \bmod 4|a|$ (see Problem 21.4). He conjectured, but was not able to prove (a rarity for Euler!), a precise result that is equivalent to the Law of Quadratic Reciprocity. Legendre formulated the Law in the same way we have, and attempted to give a proof. However, his proof had gaps.

The first complete proof of quadratic reciprocity was given by Gauss in 1796 (when he was 19 years old). Gauss went on to give *eight* proofs during his lifetime. He cherished this result and referred to it in his mathematical diary as "aureum theorema" (golden theorem).

Today there are over 300 proofs of the law of quadratic reciprocity.[15] These proofs are not all completely different; one thing they certainly have in common is that they each rely on at least one *non-obvious* idea. Indeed, the law of quadratic reciprocity is significantly more difficult to prove than anything else we've seen in the course. For this reason, I'm going to postpone the proof until after we've looked at some applications.

The first application is to calculating Legendre symbols.

Example 21.3 Determine $\left(\frac{246}{347}\right)$.

Solution: Since $246 = 2 \cdot 3 \cdot 41$, we have

$$\left(\frac{246}{347}\right) = \left(\frac{2}{347}\right)\left(\frac{3}{347}\right)\left(\frac{41}{347}\right).$$

Now let's determine each of the Legendre symbols on the right with the help of the law of quadratic reciprocity:

$$\begin{aligned}
\left(\frac{2}{347}\right) &= -1 && (347 \equiv 3 \pmod 8)\\
\left(\frac{3}{347}\right) &= -\left(\frac{347}{3}\right) && (\text{since } 347 \equiv 3 \pmod 4)\\
&= -\left(\frac{2}{3}\right) && (\text{since } 347 \equiv 2 \pmod 3)\\
&= -(-1) && (\text{since } 3 \equiv 3 \pmod 8)\\
&= 1\\
\left(\frac{41}{347}\right) &= \left(\frac{347}{41}\right) && (\text{since } 41 \equiv 1 \pmod 4)\\
&= \left(\frac{19}{41}\right) && (\text{since } 347 \equiv 19 \pmod{41})\\
&= \left(\frac{41}{19}\right) && (\text{since 19 is prime})\\
&= \left(\frac{3}{19}\right) && (\text{since } 41 \equiv 3 \pmod{19})\\
&= -\left(\frac{19}{3}\right) && (\text{since } 19 \equiv 3 \pmod 4)\\
&= -\left(\frac{1}{3}\right) && (\text{since } 19 \equiv 1 \pmod 3)\\
&= -1.
\end{aligned}$$

Thus,

$$\left(\frac{246}{347}\right) = (-1)(1)(-1) = 1.$$

Note that this shows that 246 is a square mod 347. (In fact, $246 \equiv (\pm 151)^2 \pmod{347}$.)

[15] See https://www.mathi.uni-heidelberg.de/~flemmermeyer/qrg_proofs.html

Exercise 21.4 Determine $\left(\frac{30}{61}\right)$.

REMARK (Finding Square Roots mod p)

If we know that $\left(\frac{a}{p}\right) = 1$, how do we actually find a b such that $b^2 \equiv a \pmod{p}$? Of course, we can try squaring each $b = 1, 2, \ldots$ but this is not very efficient in general. For better approaches, look up the Tonelli–Shanks algorithm and Cipolla's algorithm.

Our next application of quadratic reciprocity is to the problem of determining, given an integer a, which primes p have that integer as a quadratic residue. Note that, so far, we have fixed p and asked for $\left(\frac{a}{p}\right)$ (a finite problem, since this depends only on a mod p). Now we are fixing a and asking for $\left(\frac{a}{p}\right)$ (an infinite problem). Remarkably, quadratic reciprocity converts the latter problem into the first!

Example 21.5 Determine $\left(\frac{3}{p}\right)$, where p is an odd prime.

Solution: If $p = 3$ then $\left(\frac{3}{3}\right) = 0$, so let's assume that $p \neq 3$. We consider two cases.

If $p \equiv 1 \pmod 4$, then

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

If $p \equiv 3 \pmod 4$, then

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

Using the Chinese Remainder Theorem (Problem 10.3), we can combine each pair of conditions mod 4 and mod 3 into a condition mod 12. For example, the conditions $p \equiv 1 \pmod 4$ and $p \equiv 1 \pmod 3$ are equivalent to $p \equiv 1 \pmod{12}$. The conditions $p \equiv 1 \pmod 4$ and $p \equiv 2 \pmod 3$ are equivalent to $p \equiv 5 \pmod{12}$. We end up with the following uniform result for all $p \neq 3$:

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 11 \pmod{12} \\ -1 & \text{if } p \equiv 5, 7 \pmod{12}. \end{cases}$$

Exercise 21.6 Determine $\left(\frac{5}{p}\right)$ and $\left(\frac{7}{p}\right)$ for all odd primes p.

Our final application is to Diophantine equations. Since the Legendre symbol is intimately related to quadratic congruences, it shouldn't be surprising that quadratic reciprocity might be able to say something about certain Diophantine equations.

Example 21.7 Show that the equation $x^2 - 43y^2 = 7$ has no solutions in the integers.

Solution: Reducing mod $p = 43$, we end up with

$$x^2 \equiv 7 \pmod{43}$$

Since

$$\left(\frac{7}{43}\right) = (-1)\left(\frac{43}{7}\right) = (-1)\left(\frac{1}{7}\right) = -1,$$

it follows that 7 is a quadratic nonresidue mod 43. So there are no solutions to the equation mod 43, hence no solutions in the integers.

The next example is more involved, but it showcases a variety of techniques that we've learned in the course, so it's worth the effort.

Example 21.8 Show that the only solution to $x^4 - 17y^4 = 2z^2$ in the integers is $(x, y, z) = (0, 0, 0)$.

Solution: We may assume without loss of generality that x, y, z are pairwise coprime (exercise!). Suppose that p is an odd prime divisor of z. Then $p \neq 17$ since otherwise p would divide x, contradicting the assumption that x and z are coprime. Now reduce the equation mod p to obtain

$$x^4 \equiv 17y^4 \pmod{p}.$$

Note that $p \nmid y$ since y and z are coprime, so y is invertible mod p. Thus, $17 \equiv (xy^{-1})^4 \pmod{p}$ is a quadratic residue mod p. So, by the law of quadratic reciprocity,

$$\left(\frac{p}{17}\right) = \left(\frac{17}{p}\right) = 1.$$

Furthermore, since $17 \equiv 1 \pmod{8}$, we also have that

$$\left(\frac{-1}{17}\right) = \left(\frac{2}{17}\right) = 1.$$

Assuming $z \neq 0$, we can write z a product of ± 1, 2 and odd primes. Thus, $\left(\frac{z}{17}\right) = 1$ by the multiplicativity of the Legendre symbol. So $z \equiv w^2 \pmod{17}$ for some w. But then

$$x^4 \equiv 2w^4 \pmod{17}.$$

Since $w \not\equiv 0 \pmod{17}$ (otherwise $17 \mid z$), the above shows that $2 \equiv (xw^{-1})^4 \pmod{17}$ is a 4th power residue mod 17. But this is not true (exercise!). The only way out of this contradiction is for z to not be a product of ± 1, 2 and odd primes, which can only be the case if $z = 0$. It easily follows then that $x = y = 0$ (exercise!).

Exercise 21.9 Fill in the gaps left in the above solution.

(a) Explain why we may assume that x, y and z are pairwise coprime.

(b) Prove that 2 is not a 4th power residue mod 17.

(c) Prove that if $z = 0$ then $x = y = 0$.

Lecture 21 Problems

21.1. In this problem you will use quadratic reciprocity to give a Euclid-style proof that there are infinitely many primes of the form $5k + 4$.

(a) Suppose that $p_1, \ldots, p_n$ are primes of the form $5k + 4$. Let $N = 5(p_1 \cdots p_n)^2 - 1$. Show that if a prime p divides N then $\left(\frac{p}{5}\right) = 1$.

(b) Deduce from part (a) that N must have a prime divisor of the form $5k + 4$.

(c) Conclude that there must be infinitely many primes of the form $5k + 4$.

21.2. Suppose that p is a prime of the form $p = 2^{2^n} + 1$ with $n > 1$. (Such a prime is called a **Fermat prime**; see Problem 6.2.)

(a) Determine $\left(\frac{3}{p}\right)$.

(b) Prove that 3 is a primitive root mod p.

21.3. Find all primes p for which the congruence $x^2 + x + 1 \equiv 0 \pmod{p}$ has a solution.

21.4. Let a be a non-zero integer, and let p and q be odd primes that do not divide a. Prove that if $p \equiv q \pmod{4|a|}$ then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.

21.5. Show that the equation $x^2 + 10xy - 6y^2 = 17$ has no solutions in the integers. [**Hint:** Try to complete the square on the left-side.]

Lecture 22 The Proof of Quadratic Reciprocity

The proof works, but it is remarkable in the fact that it gives us no insight at all into why the theorem is true. In particular, it does not yield any direct connection between "life mod p" and "life mod q." Every time I present the proof to students, I point out the feeling that yes, it comes out right, but it comes out right because the theorem is true. It's hard to claim (and I do not believe) that counting points in a rectangle explains why the theorem is true.

– F. Gouvêa

The task of presenting an illuminating, elementary proof of the law quadratic reciprocity is a challenging one. In my opinion, none of the known elementary proofs offer any insight as to why the result is true. They all rely on some counting trickery, where the same thing is tallied up in two ways, and when the two counts are compared the desired result magically appears.

The plan for this lecture is to present a version of Gauss's third proof of the law of quadratic reciprocity. The counting trickery will be based on a clever (and interesting) result known as Gauss's Lemma.

We need some notation. Let p be an odd prime and let

$$S = \{1, 2, \ldots, (p-1)/2\} \quad \text{and} \quad -S = \{-1, -2, \ldots, -(p-1)/2\}.$$

Notice that S and $-S$ are disjoint and that $S \cup -S$ forms a complete set of representatives for $(\mathbb{Z}/p\mathbb{Z})^\times$. For example, if $p = 7$ then $S = \{1, 2, 3\}$ and $-S = \{-1, -2, -3\}$, and we have $S \cup -S = \{1, 2, 3, -1, -2, -3\}$. Since $4, 5, 6 \equiv -3, -2, -1 \pmod 7$, we see that the integers in $S \cup -S$ represent every class in $(\mathbb{Z}/7\mathbb{Z})^\times$.

If $a \in \mathbb{Z}$ is coprime to p, let $n(a)$ denote the number of integers among

$$a, 2a, 3a, \ldots, ((p-1)/2)a$$

that have representatives in $-S$. For example, if $p = 7$ and $a = 3$, then

$$a, 2a, 3 = 3, 6, 9 \equiv 3, -1, 2 \pmod 7$$

so $n(a) = 1$. If $a = 4$, then

$$a, 2a, 3a = 4, 8, 12 \equiv -3, 1, -2 \pmod 7$$

so $n(a) = 2$.

Informally, $n(a)$ counts the number of sign changes that occur after we multiply the elements of S by a.

Lemma 22.1 **(Gauss's Lemma)**

Using the above notation, $\left(\dfrac{a}{p}\right) = (-1)^{n(a)}$.

For example,

$$\left(\frac{3}{7}\right) = (-1)^{n(3)} = (-1)^1 = -1$$

and

$$\left(\frac{4}{7}\right) = (-1)^{n(4)} = (-1)^2 = 1.$$

Note that we can confirm the latter by noting that $\left(\frac{4}{7}\right) = \left(\frac{2^2}{7}\right) = 1$.

Exercise 22.2 Use Gauss's Lemma to determine $\left(\frac{5}{7}\right)$.

Proof of Lemma 22.1 (Gauss's Lemma):

For each $s \in S$, we have $as \equiv \pm\tilde{s} \pmod{p}$ for some $\tilde{s} \in S$, since $S \cup -S$ is a complete set of representatives for all $(\mathbb{Z}/p\mathbb{Z})^\times$. Note that this $\tilde{s}$ is uniquely determined by s, because if there is a $t \in S$ such that $at \equiv \pm\tilde{s} \mod p$, then

$$as \equiv \pm at \pmod{p} \quad \iff \quad s \equiv \pm t \pmod{p}.$$

This is only possible if $s = t$ since both s and t are in S. So $s \mapsto \tilde{s}$ is an injection, hence a bijection, on the finite set S.

Now we will compute the product $\prod_{s\in S} as$ in two ways. First, we have

$$\prod_{s\in S} as = a^{|S|} \prod_{s\in S} s = a^{(p-1)/2} \prod_{s\in S} s \equiv \left(\frac{a}{p}\right) \prod_{s\in S} s \pmod{p},$$

where we used Euler's Criterion 20.3 in the last step. Second, we have

$$\prod_{s\in S} as \equiv \prod_{\tilde{s}\in S} \pm\tilde{s} = (-1)^{n(a)} \prod_{\tilde{s}\in S} s = (-1)^{n(a)} \prod_{s\in S} s \pmod{p},$$

where we've used the fact that $s \leftrightarrow \tilde{s}$ is a bijection.

Thus,

$$\left(\frac{a}{p}\right) \prod_{s\in S} s \equiv (-1)^{n(a)} \prod_{s\in S} s \pmod{p}.$$

Note that since all the elements $s \in S$ are units mod p, so is their product, and so we can cancel off the product from both sides, leaving us with

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

Since both sides are ± 1, they must be equal—not just congruent mod p (since p is odd). This completes the proof. ■

With this in hand, we can easily determine $\left(\frac{2}{p}\right)$.

Proof of Theorem 21.1 (Law of Quadratic Reciprocity) – Part (a):

In order to use Gauss's Lemma, we must count the number of sign changes in $2s$ ($s \in S$):

$$2,\ 4,\ 6, \ldots, p-1.$$

The values of s that flip signs are precisely those that satisfy

$$\frac{p-1}{2} < 2s \leq p-1$$

or, equivalently,

$$\frac{p-1}{4} < s \leq \frac{p-1}{2}.$$

There are precisely

$$n(2) = \frac{p-1}{2} - \left\lfloor \frac{p-1}{4} \right\rfloor$$

such values of s. We then have $\left(\frac{2}{p}\right) = (-1)^{n(2)}$ by Gauss's Lemma. Examining each of the possibilities of p mod 8, we find:

p	$n(2)$	$(-1)^{n(2)}$
$8k+1$	$4k - 2k = 2k$	$+1$
$8k+3$	$(4k+1) - 2k = 2k+1$	-1
$8k+5$	$(2k+2) - (2k+1) = 2k+1$	-1
$8k+7$	$(4k+3) - (2k+1) = 2k+2$	$+1$

This is precisely what we want to prove. ■

Proof of Theorem 21.1 (Law of Quadratic Reciprocity) – Part (b):

We will prove that $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$. Let

$$S_p = \{1, 2, \ldots, (p-1)/2\} \quad \text{and} \quad S_q = \{1, 2, \ldots, (q-1)/2\}.$$

We wish to apply Gauss's Lemma to q and S_p mod p, and then to p and S_q mod q. Thus, let $n_p(q)$ be the number of sign changes of qs ($s \in S_p$) mod p, and let $n_q(p)$ be the number of sign changes of ps ($s \in S_q$) mod q. Then, using Gauss's Lemma twice, we have

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

We need a better handle on how to count these sign flips. To this end, note that if $qs \equiv -s' \pmod{p}$ for some $s' \in S$ (so s contributes to $n_p(q)$) then we can write $qs = -s' + pt$ for some $t \in \mathbb{Z}$. We have $pt - qs = s' \in S_p$ therefore

$$0 < pt - qs \leq \frac{p-1}{2}.$$

This t must belong to S_q since $pt > qs > 0$ and

$$pt \leq \frac{p-1}{2} + qs \leq (q+1)\frac{p-1}{2} < \frac{q+1}{2}p$$

hence $t < (q+1)/2$ and therefore, since q is odd, $t \leq (q-1)/2$.

This shows that

$$n_p(q) = |\{(s,t) \in S_p \times S_q \colon 0 < pt - qs \leq (p-1)/2\}|\,.$$

By the same argument,

$$n_q(p) = |\{(s,t) \in S_p \times S_q \colon 0 \leq qs - pt \leq (q-1)/2\}|\,.$$

We can re-write this as

$$n_q(p) = |\{(s,t) \in S_p \times S_q \colon -(q-1)/2 \leq pt - qs < 0\}|\,.$$

Now since $pt - qs \neq 0$ when $(s,t) \in S_p \times S_q$, if we let

$$X = \{(s,t) \colon -(q-1)/2 \leq pt - qs \leq (p-1)/s\}$$

then we find that

$$n_p(q) + n_q(p) = |X|.$$

Now, given $(s,t) \in X$, let $(s',t') = ((p+1)/2 - s, (q+1)/2 - t)$. It's easy to check that $(s',t') \in X$. Thus, $f(s,t) = (s',t')$ defines a function $f \colon X \to X$. This function is a bijection (with inverse $f^{-1} = f$ itself). Thus, the number of elements in X is equal to 2 times the number of x such that $f(x) \neq x$ (since each such x pairs with a unique $f(x) \neq x$) plus the number of fixed points x such that $f(x) = x$. The only potential fixed point of f is $((p+1)/4, (q+1)/4)$ which only exists if both p and q are 3 mod 4. Thus, $|X|$ is odd if and only if p and q are 3 mod 4 and $|X|$ is even in all other cases. But then the same is true of

$$n_p(q) + n_q(p) = |X|$$

and hence

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{n_p(q)+n_q(p)}$$

is exactly as stated by the law of quadratic reciprocity. This completes the proof. ■

If you made it this far, congratulations!

In some textbooks, the set $S_p \times S_q$ is displayed as a rectangle of lattice points in the plane, and the set X is identified as a strip along the diagonal of the rectangle. The function f is then a half-rotation about the midpoint of X. It might be instructive to try to draw these pictures and try to make sense of the above manipulations. But, as Gouvêa says in the opening quote to the lecture, these manipulations and pictures don't really tell the story of *why* quadratic reciprocity is true. They merely verify that it *is* true.

REMARK (A Better Proof?)

It's a bit of a letdown that one of the great theorems of elementary number theory has such a clunky proof. The good news is: *illuminating proofs exist!* The best one, in my opinion, requires ideas from algebraic number theory and Galois theory, so it's unfortunately a bit too advanced for this course.

Lecture 22 Problems

22.1. Use Gauss's Lemma to prove that $\left(\dfrac{-1}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4 \\ -1 & \text{if } p \equiv 3 \pmod 4 \end{cases}$.

Lecture 23 Primality Testing

Having tackled the second and third problems on page 79, we now turn our attention to the first problem: How do we test if a positive integer n is a prime number? We of course have our naive test (Corollary 7.2) that asks us to run through the integers $1, 2, \ldots, \lfloor\sqrt{n}\rfloor$ and check if any of them divide n. Obviously, this is extremely inefficient if n is large.

In this lecture we will learn about about more efficient *primality tests*. The underlying idea is simple: Identify a property that all primes satisfy, and then check if n satisfies it as well. If n doesn't, then we know for sure that n isn't prime. If n does, then n may or may not be a prime—but if n passes several of these checks, then we grow more confident in the primality of n.

Here are two things that every prime p must satisfy:

- Fermat's Little Theorem: If a is coprime to p then $a^{p-1} \equiv 1 \pmod{p}$.
- Euler's Criterion: If a is coprime to p, and if p is odd, then $a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) \pmod{p}$.

We will leverage them to create two primality tests.

The Fermat Test

To test whether $n \in \mathbb{Z}_{>0}$ is prime, we check if Fermat's Little Theorem holds for n.

- If $\gcd(a, n) = 1$ and $a^{n-1} \equiv 1 \pmod{n}$, then we say that n **passes the Fermat test** for the base a.
- If $\gcd(a, n) = 1$ and $a^{n-1} \not\equiv 1 \pmod{n}$, then we say that n **fails the Fermat test** for the base a.

If n is prime, it will pass the Fermat test for all bases a coprime to n. So if an integer n passes the test, then it "behaves like" a prime (although it may not actually be prime—see below). If n fails one instance of the Fermat test, then n is definitely not prime.

Example 23.1 Determine if $n = 119$ is prime.

Solution: Clearly $\gcd(2, n) = 1$, and we have

$$2^{n-1} = 2^{118} \equiv 30 \pmod{119}.$$

Since $2^{n-1} \not\equiv 1 \pmod{119}$, 119 fails the Fermat test for $a = 2$. Thus, 119 is **not** prime.

It's interesting that we're able to assert this without having found any nontrivial divisors of 119. (In fact, $119 = 7 \times 17$.)

Example 23.2

Determine if $n = 341$ is prime.

Solution: We have $\gcd(2, n) = 1$ and

$$2^{n-1} = 2^{340} \equiv 1 \pmod{341}.$$

So 341 passes the Fermat test for the base $a = 2$. However, if we try $a = 3$, we find that $\gcd(3, n) = 1$ and

$$3^{n-1} = 3^{340} \equiv 56 \pmod{341}.$$

So 341 fails the Fermat test for the base $a = 3$. Thus, 341 is **not** prime. (In fact, $341 = 11 \times 31$.)

Numbers like $n = 341$ that are composite but that pass the Fermat test are "pretending" to be prime. We call them pseudoprimes.

Definition 23.3

Base-a Pseudoprime

Let n be an odd positive integer and let $a \in \mathbb{Z}$ be coprime to n. If n is composite but passes the Fermat test for the base a then we say that n is a **base-a (Fermat) pseudoprime**.

So 341 is a base-2 pseudoprime. Pseudoprimes are relatively rare, although it is known that there are infinitely many of them. Of the integers up to 1000, only three are base-2 pseudoprimes, namely

$$341, \quad 561 \quad \text{and} \quad 645.$$

Of the integers up to one billion, only 5595 are base-2 pseudoprimes.

We were able to deal with $n = 341$ by applying the Fermat test with $a = 3$. After all, for the Fermat test to detect that n is composite, all we need is a single $a < n$ such that $a^{n-1} \not\equiv 1 \pmod{n}$.

Example 23.4

Determine if $n = 8911$ is prime.

Solution: We compute that $\gcd(2, n) = \gcd(3, n) = \gcd(5, n) = \gcd(7, n) = 1$ and

$$\begin{aligned} 2^{n-1} &= 2^{8910} \equiv 1 && \pmod{8911} \\ 3^{n-1} &= 3^{8910} \equiv 1 && \pmod{8911} \\ 5^{n-1} &= 5^{8910} \equiv 1 && \pmod{8911} \\ 7^{n-1} &= 7^{8910} \equiv 1274 && \pmod{8911}. \end{aligned}$$

The first three congruences don't give any conclusive information, but the last one tells us that 8910 is composite.

As long as we can find a suitable base a quickly, the Fermat test works rather well. Unfortunately, there are composite numbers n that are pseudoprimes for all bases a coprime to n. Such numbers are called **Carmichael numbers**. The smallest one is $561 = 3 \times 11 \times 17$ and the next smallest one is $1105 = 5 \times 13 \times 17$. Alford, Granville and Pomerance proved in 1994 that there are infinitely many Carmichael numbers. However, they are extremely rare. There are only 646 Carmichael numbers below one billion.

For a Carmichael number n, the first a such that $a^{n-1} \not\equiv 1 \pmod{n}$ will be a divisor of n (why?). The existence of Carmichael numbers is bad news for the Fermat test since finding a base for which the test fails is just as difficult as finding a nontrivial divisor.

Exercise 23.5 Show that $a^{561-1} \equiv 1 \pmod{561}$ for all a coprime to 561.

REMARK (History of Fermat's Little Theorem)

Fermat discovered his "Little Theorem" while trying to determine which integers of the form $2^n - 1$ are prime (these are called **Mersenne primes**); that is, he was trying to devise a primality test! He actually proved (or, well, claims to have proved!) the following statement, which is stronger than what is usually called Fermat's Little Theorem.

If $p \nmid a$ then there exists an integer d such that $a^d \equiv 1 \pmod{p}$ and the smallest such integer d divides $p-1$ and divides any n such that $a^n \equiv 1 \pmod{p}$.

In modern language, Fermat discovered the notion of $\operatorname{ord}_p(a)$ and the fact that $\operatorname{ord}_p(a) \mid \varphi(p)$.

The Miller–Rabin Test

There is a way around the Carmichael number problem. It involves a slightly more penetrating look at the Fermat congruence $a^{n-1} \equiv 1 \pmod{n}$. Let me illustrate with an example.

Example 23.6 Let $n = 341$, which we know passes the Fermat test for the base $a = 2$. That is, we know that $2^{n-1} \equiv 1 \pmod{n}$, or equivalently, that n divides

$$2^{340} - 1 = (2^{170} - 1)(2^{170} + 1) = (2^{85} - 1)(2^{85} + 1)(2^{170} + 1).$$

If n were prime, then n would divide one of the factors on the right (by Euclid's Lemma). Thus, one of the following congruences must be true:

$$\begin{aligned} 2^{85} &\equiv 1 \pmod{341} \\ 2^{85} &\equiv -1 \pmod{341} \\ 2^{170} &\equiv -1 \pmod{341}. \end{aligned}$$

However, we have

$$\begin{aligned} 2^{85} &\equiv 32 \pmod{341} \\ 2^{170} &\equiv 1 \pmod{341}. \end{aligned}$$

So none of the congruences hold. Consequently, n cannot be prime.

The method in the example generalizes to give the **Miller–Rabin test**, which I will now describe.

Suppose we want to test $n \in \mathbb{Z}_{>0}$ for primality. We may certainly assume that n is odd. Let's also assume that n has passed the Fermat test for the base a, so that $\gcd(a, n) = 1$ and

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

Writing $n = 2^v m + 1$, where m is odd, we see that n divides

$$\begin{aligned} a^{n-1} - 1 &= a^{2^v m} - 1 \\ &= (a^{2^{v-1} m} - 1)(a^{2^{v-1} m} + 1) \\ &= (a^{2^{v-2} m} - 1)(a^{2^{v-2} m} - 1)(a^{2^{v-1} m} + 1) \\ &\vdots \\ &= (a^m - 1)(a^m + 1)(a^{2m} + 1) \cdots (a^{2^{v-1} m} + 1). \end{aligned}$$

If n were prime, then by Euclid's Lemma, n would divide one of the factors on the right, and so one of the following congruences must hold:

- $a^m \equiv 1 \pmod{n}$, or
- $a^{2^i m} \equiv -1 \pmod{n}$ for some $0 \leq i < v$.

Notice that the powers of a above can be obtained by repeated squaring:

$$a^m \to a^{2m} \to a^{4m} \to \cdots \to a^{2^{v-1} m}.$$

So we can check whether the congruences hold in sequential order.

- If $a^m \equiv \pm 1 \pmod{n}$, we can stop. One of the congruences has been satisfied. We say the n has **passed the Miller–Rabin test** for the base a.
- If $a^m \not\equiv \pm 1 \pmod{n}$, then we compute a^{2m}.
  - If $a^{2m} \equiv 1 \pmod{n}$, we can stop. All the other $a^{2^i m}$ will also be congruent to 1, and so none of the congruences hold. Thus, a is composite. We say the n has **failed the Miller–Rabin test** for the base a.
  - If $a^{2m} \equiv -1 \pmod{n}$, we can stop. One of the congruences has been satisfied. We say the n has **passed the Miller–Rabin** test for the base a.
- If $a^{2m} \not\equiv \pm 1 \pmod{n}$, then we compute a^{4m}, and repeat the analysis above. Etc.

So if one of the congruences holds, then n passes the Miller–Rabin test and is "behaving like" a prime (though, just like with the Fermat test, n is not necessarily prime—see below). Otherwise, if none of the congruences hold, then n fails the Miller–Rabin test and is definitely composite.

Example 23.7

Let's apply the Miller–Rabin test to the Carmichael number $n = 561$. We know that n satisfies the Fermat test for the base $a = 2$.

In this case, $n = 2^4 \cdot 35 + 1$, so $m = 35$ and $v = 4$. We compute

$$\begin{aligned} 2^{35} &\equiv 263 \not\equiv \pm 1 \pmod{561} \\ 2^{2\cdot 35} &\equiv 166 \not\equiv \pm 1 \pmod{561} \\ 2^{2^2\cdot 35} &\equiv 67 \not\equiv \pm 1 \pmod{561} \\ 2^{2^3\cdot 35} &\equiv 1 \pmod{561}. \end{aligned}$$

At this point we stop and note that n has failed the Miller–Rabin test. Thus, 561 is composite.

Example 23.8

Let's apply the Miller–Rabin test to $n = 2047$. In this case, $n = 2 \cdot 1023 + 1$, so $m = 1023$ and $v = 1$, and we only have to check whether $a^m \equiv \pm 1 \pmod{n}$.

With $a = 2$, we compute

$$2^{1023} \equiv 1 \pmod{2047}.$$

So n passes the Miller–Rabin test for the base $a = 2$.

With $a = 3$, we compute

$$3^{1023} \equiv 1565 \not\equiv \pm 1 \pmod{2047}.$$

So n has failed the Miller–Rabin test for the base $a = 3$. Thus, n is composite. Indeed, $2047 = 23 \times 89$.

Exercise 23.9

Determine if $n = 15841$ is prime using the Miller–Rabin test.

Composite numbers like 2047 that pass the Miller–Rabin test also deserve to be called pseudoprimes. However, they are more sneaky than Fermat pseudoprimes.

Definition 23.10
Strong Pseudoprime

Let n be an odd positive integer and let $a \in \mathbb{Z}$ be coprime to n. If n is composite but passes the Miller–Rabin Test for the base a then we say that n is a **base-a strong pseudoprime**.

So 2047 is a base-2 strong pseudoprime. By design, any integer n that passes the Miller–Rabin test also passes the Fermat test. Thus, a strong pseudoprime is automatically a pseudoprime. The converse is false ($n = 561$ is a counterexample).

Strong pseudoprimes are rarer than pseudoprimes. There are only 1282 base-2 strong pseudoprimes below one billion (compared to 5597 pseudoprimes). Of these, only three are also strong pseudoprimes for the bases $a = 3$ and $a = 5$, namely

$$25326001, \quad 161304001 \quad \text{and} \quad 960946321.$$

So assuming we don't apply it to one of these numbers, the Miller–Rabin test can correctly detect primality up to 10^9—we just have to use the three bases $a = 2, 3, 5$. Unfortunately,

10^9 is not large enough for cryptographic applications, where we often need primes of size 2^{2048} or 2^{4096}.

However, here is some good news. First, there are no "strong Carmichael numbers" (i.e. integers n that pass the Miller–Rabin test for all bases coprime to n) like there were for the Fermat test. Second, Rabin proved that if we perform the Miller–Rabin test on a composite n using k randomly chosen bases, then the probability that n passes these k tests is $\leq 4^{-k}$. So the probability that a pseudoprime survives $k = 100$ iterations of the Miller–Rabin test is $\leq 4^{-100}$, i.e., it is practically zero. This gives us a fairly reliable *probabilistic* primality test. This testing scheme is actually used in practice. Integers that pass it, and are therefore very probably prime, are called **industrial-grade primes**!

REMARK (Deterministic Primality Testing)

If one assumes a version of the generalized Riemann hypothesis, then it can be proved that every composite integer n fails the Miller–Rabin test for some base $a \leq 2(\log n)^2$. This gives us an algorithm for *proving* primality (and not just establishing *probable primality*)—we just have to apply the Miller–Rabin test to all bases up to this bound. However, in practice, the probabilistic test is more efficient since it requires significantly fewer computations.

The Solovay–Strassen Test and the Jacobi Symbol

The Fermat and Miller–Rabin tests were based on Fermat's Little Theorem and Euclid's Lemma. I will now describe a primality test that makes use of the fact that primes satisfy Euler's Criterion: If p is an odd prime and if a is coprime to p, then

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

Thus, to test n for primality, we can take an integer a coprime to n, and determine whether

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

But wait! The Legendre symbol is only defined for primes p, so we need to explain what we mean by $\left(\frac{a}{n}\right)$.

Definition 23.11
Jacobi symbol

Let $n \in \mathbb{Z}_{>0}$ be odd and let $a \in \mathbb{Z}$. If the prime factorization of n is

$$n = p_1^{a_1} \cdots p_k^{a_k}$$

then the **Jacobi symbol** is defined by

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{a_1} \cdots \left(\frac{a}{p_k}\right)^{a_k},$$

where $\left(\frac{a}{p_i}\right)$ is the Legendre symbol (well-defined since p_i is necessarily an odd prime).

For example,

$$\left(\frac{11}{45}\right) = \left(\frac{11}{3^2 \cdot 5}\right) = \left(\frac{11}{3}\right)^2 \left(\frac{11}{5}\right) = (1)\left(\frac{1}{5}\right) = 1.$$

It should be pointed out immediately that, unlike with the Legendre symbol, the fact that $\left(\frac{a}{n}\right) = 1$ does **not** imply that a is a square mod n. Indeed,

$$\left(\frac{2}{15}\right) = \left(\frac{2}{3}\right)\left(\frac{2}{5}\right) = (-1)(-1) = 1$$

but 2 is not a square mod 15 (because, for instance, it's not a square mod 3).

This aside, the Jacobi symbol satisfies many properties that the Legendre symbol satisfies, including the law of quadratic reciprocity.

Proposition 23.12 Let $n \in \mathbb{Z}_{>0}$ be an odd and let $a, b \in \mathbb{Z}$ be coprime to n.

(a) If $a \equiv b \pmod{n}$ then $\left(\frac{a}{n}\right) = \left(\frac{b}{n}\right)$.

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

(c) $\left(\frac{a^2}{n}\right) = 1$.

Theorem 23.13 **(Law of Quadratic Reciprocity—Jacobi Symbol Version)**

Let n and m be coprime positive integers. Then:

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

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

(c) $\left(\frac{m}{n}\right)\left(\frac{n}{m}\right) = (-1)^{\frac{n-1}{2}\frac{m-1}{2}}$.

Exercise 23.14 Prove Proposition 23.12 and Theorem 23.13.

Before returning to primality testing, let me point out that the law of quadratic reciprocity for the Jacobi symbol allows us to calculate Legendre symbols without having to factor integers.

Example 23.15 Determine the Legendre symbol $\left(\frac{585}{673}\right)$. [Note: 673 is prime.]

Solution: Using our old method for calculating Legendre symbols, we would begin by obtaining the prime factorization $585 = 3^2 \times 5 \times 13$ so that

$$\left(\frac{585}{673}\right) = \left(\frac{3}{673}\right)^2\left(\frac{5}{673}\right)\left(\frac{13}{673}\right) = \left(\frac{5}{673}\right)\left(\frac{13}{673}\right).$$

We would then proceed by calculating the two Legendre symbols on the right.

However, with the Jacobi symbol, we can note that $673 \equiv 1 \pmod 4$ and then immediately apply quadratic reciprocity to get

$$\left(\frac{585}{673}\right) = \left(\frac{673}{585}\right).$$

Now, as $673 \equiv 88 \pmod{585}$, we have

$$\left(\frac{673}{585}\right) = \left(\frac{88}{585}\right) = \left(\frac{2}{585}\right)^3 \left(\frac{11}{585}\right)$$

where in the last step we simply pulled out the factors of 2. Since $585 \equiv 1 \pmod 8$, we have that

$$\left(\frac{2}{585}\right) = 1$$

leaving us with

$$\left(\frac{673}{585}\right) = \left(\frac{11}{585}\right) = \left(\frac{585}{11}\right) = \left(\frac{2}{11}\right) = -1.$$

Thus,

$$\left(\frac{585}{673}\right) = -1.$$

Exercise 23.16 Determine the Legendre symbol $\left(\frac{655}{719}\right)$.

Now, back to primality testing. To test an odd positive integer n for primality, we simply check if n satisfies Euler's Criterion.

- If $\gcd(a, n) = 1$ and $a^{(n-1)/2} \equiv \left(\frac{a}{n}\right) \pmod n$, then we say that n **passes the Solovay–Strassen test** for the base a.
- If $\gcd(a, n) = 1$ and $a^{(n-1)/2} \not\equiv \left(\frac{a}{n}\right) \pmod n$, then we say that n **fails the Solovay–Strassen test** for the base a.

Just as with the Fermat and Miller–Rabin tests, if n passes the Solovay–Strassen test, we cannot assert that n is prime. However, if n fails the Solovay–Strassen test, then n definitely composite.

Example 23.17 Determine if $n = 779$ is prime.

Solution: With $a = 2$, we have

$$a^{(n-1)/2} = 2^{389} \equiv 471 \pmod{779}.$$

We immediately conclude that n has failed the Solovay–Strassen test since 471 is not congruent to $\left(\frac{2}{n}\right) = \pm 1 \bmod 779$. So 779 is composite.

Example 23.18

Determine if $n = 3277$ is prime.

Solution: With $a = 2$, we have

$$a^{(n-1)/2} = 2^{1638} \equiv -1 \pmod{3277}.$$

Since $3277 \equiv -3 \pmod 8$, we have $\left(\frac{2}{3277}\right) = -1$. Thus, n passes the Solovay–Strassen test for the base $a = 2$.

With $a = 3$, we have

$$a^{(n-1)/2} = 3^{1638} \equiv 434 \pmod{3277}.$$

Since this isn't congruent to ± 1, n has failed the Solovay–Strassen test. So n is composite.

A composite number that passes the Solovay–Strassen test for the base a is called a base-a **Euler pseudoprime**. Every Euler pseudoprime is a Fermat pseudoprime but not conversely. Every strong pseudoprime is an Euler pseudoprime but not conversely. Although I won't be proving these assertions, I will just note that they follow easily from the definitions. A result that lies a little deeper is the following, which was proved by Solovay and Strassen.

Proposition 23.19

Assume n is composite. If we perform the Solovay–Strassen test on n using k randomly chosen bases $a < n$, then the probability that n passes these k tests is $\leq 2^{-k}$.

This result says that performing, say, $k = 100$ iterations of the Solovay–Strassen test is a good probabilistic primality test (though not as good as $k = 100$ iterations of Miller–Rabin).

Lecture 23 Problems

23.1. Let $n \in \mathbb{Z}_{>0}$.

(a) Show that if $a^{n-1} \equiv 1 \pmod n$ for all $a \not\equiv 0 \pmod n$, then n is prime.

(b) Why does part (a) not contradict the fact that Carmichael numbers exist?

23.2. Show that every composite Fermat number $F_n = 2^{2^n} + 1$ is a base-2 pseudoprime.

23.3. Let $n \in \mathbb{Z}_{>0}$ and let $a, b \in \mathbb{Z}$ be coprime to n. Prove/disprove:

(a) If n is a base-a and base-b pseudoprime, then n is a base-ab pseudoprime.

(b) If n is a base-a pseudoprime then n is a base-a' pseudoprime, where a' is the inverse of a mod n.

23.4. Prove that if n is a base-a Euler pseudoprime, then n is a base-a Fermat pseudoprime. Give an example to show that the converse is false.

Lecture 24 The Gaussian Integers

> *[...] we soon recognized that the principles of arithmetic which were usable until then were in no way sufficient to build the general theory. Rather such a theory necessarily required an infinite enlargement to some extent of the field of higher arithmetic.*
>
> – C.F. Gauss

One of the curious facts about number theory is that sometimes to prove theorems about $\mathbb{Z}$ we have to work in larger number systems. In this lecture we will begin our investigation of the set of **Gaussian integers**

$$\mathbb{Z}[i] = \{a + ib \colon a, b \in \mathbb{Z}\}.$$

Here, i satisfies $i^2 = -1$, so we may view $\mathbb{Z}[i]$ as sitting inside the set of complex numbers $\mathbb{C}$. Gauss introduced $\mathbb{Z}[i]$ primarily as a tool to help him investigate *higher laws of reciprocity* (specifically, quartic reciprocity).

We will be a little more modest. Our primary short-term goal is to determine which integers are sums of two squares. That is, for which $n \in \mathbb{Z}$ are there solutions to the Diophantine equation

$$n = x^2 + y^2?$$

The Gaussian integers come into play because of the following identity

$$x^2 + y^2 = x^2 - i^2 y^2 = (x - iy)(x + iy).$$

So if we can somehow develop a theory of "unique factorization into primes" in $\mathbb{Z}[i]$ (whatever that means), then perhaps it will shed some light on our problem.

Towards a Fundamental Theorem of Arithmetic in $\mathbb{Z}[i]$

To prove unique factorization into primes in $\mathbb{Z}$, we made use of the following:

- The fact that $\mathbb{Z}$ is closed under addition, subtraction and multiplication but not division (which leads to the notion of divisibility).
- Theorem 2.3 (The Remainder Theorem).
- GCDs and Proposition 3.7 (Bézout's Lemma).
- Primes, coprimality and Lemma 5.3 (Euclid's Lemma).

We will try to generalize each to $\mathbb{Z}[i]$. The first is rather easy—it's clear that the sum, difference and product of Gaussian integers is itself a Gaussian integer.[16] For example,

$$(2 + 3i)(4 - 7i) = 8 - 14i + 12i + 21 = 29 - 2i.$$

For divisibility, we make the same definition we had in $\mathbb{Z}$.

[16] What I am getting at here is that $\mathbb{Z}[i]$, like $\mathbb{Z}$, is a *ring*.

Definition 24.1
Divides

Let $\alpha, \beta \in \mathbb{Z}$, with $\beta \neq 0$. We say that β **divides** α if $\alpha = \beta\gamma$ for some $\gamma \in \mathbb{Z}[i]$. We write $\beta \mid \alpha$ if β divides α and $\beta \nmid \alpha$ otherwise.

For example,

$$5 = 1^2 + 2^2 = (1+2i)(1-2i)$$

so $(1+2i) \mid 5$ and $(1-2i) \mid 5$. On the other hand, $(1+2i) \nmid (3+4i)$. To prove this, we must show that there is no Gaussian integer $\gamma = a + bi$ such that

$$3 + 4i = (a+bi)(1+2i).$$

This is easy to do by expanding and comparing real and imaginary parts. Alternatively, we can leverage the fact that we are working in $\mathbb{C}$, where we can make the observation that for $\alpha, \beta \in \mathbb{Z}[i]$,

$$\beta \mid \alpha \quad \iff \quad \frac{\alpha}{\beta} \in \mathbb{Z}[i].$$

Here, $\dfrac{\alpha}{\beta}$ is a complex number and we are to check that it is in $\mathbb{Z}[i]$. For example,

$$\frac{3+4i}{1+2i} = \frac{3+4i}{1+2i}\frac{1-2i}{1-2i} = \frac{11}{5} - \frac{2}{5}i.$$

Since this is not in $\mathbb{Z}[i]$, it follows that $(1+2i) \nmid (3+4i)$.

Being able to multiply by the complex conjugate of $1+2i$ was very helpful above, and in general we introduce the following important function.

Definition 24.2
Norm

The **norm** of $\alpha = a + ib \in \mathbb{Z}[i]$ is defined to be

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

Equivalently, $N(\alpha) = \alpha\overline{\alpha} = |\alpha|^2$ where $\overline{\alpha} = a - ib$ is the complex conjugate of α and $|\alpha|$ is the complex magnitude.

For example,

$$N(5) = 5^2 = 25 \quad \text{and} \quad N(3+2i) = 3^2 + 4^2 = 13.$$

Proposition 24.3

Let $\alpha, \beta \in \mathbb{Z}[i]$. Then:

(a) $N(\alpha) \in \mathbb{Z}_{\geq 0}$ and $N(\alpha) = 0$ if and only if $\alpha = 0$.

(b) $N(\alpha\beta) = N(\alpha)N(\beta)$. That is, the norm is (completely) multiplicative.

(c) If $\alpha \mid \beta$ then $N(\alpha) \mid N(\beta)$.

Proof: Part (a) is immediate from the definition of $N(\alpha)$. For part (b), note that

$$N(\alpha\beta) = |\alpha\beta|^2 = |\alpha|^2|\beta|^2 = N(\alpha)N(\beta)$$

and for part (c) we have

$$\alpha \mid \beta \implies \beta = \gamma\alpha \implies N(\beta) = N(\gamma)N(\alpha) \implies N(\alpha) \mid N(\beta).$$

This completes the proof. ■

Exercise 24.4 Show that the converse to part (c) is false: $N(\alpha) \mid N(\beta)$ does not imply that $\alpha \mid \beta$.

The norm is useful because it allows us to move from $\mathbb{Z}[i]$ to $\mathbb{Z}$ in a way that respects multiplication. It also allows us to speak of the "size" of a Gaussian integer, and with this we can state and prove the Remainder Theorem.

Theorem 24.5 **(The Remainder Theorem in $\mathbb{Z}[i]$)**

Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$. Then there exists $q, r \in \mathbb{Z}[i]$ such that

$$\alpha = q\beta + r \quad \text{and} \quad 0 \leq N(r) < N(\beta).$$

Proof: The idea of the proof is similar to the proof in $\mathbb{Z}$. We have to locate the multiple $q\beta$ that is closest to α, and then $r = \alpha - q\beta$ will be our desired remainder. The multiples of β tessellate the plane with squares of side length $|\beta| = \sqrt{N(\beta)}$ and α will lie in one of these squares. *Needs picture— will add later.* The distance from α to the closest vertex $q\beta$ is $\leq$ the distance from the center of the square to that vertex. Thus,

$$|\alpha - q\beta| \leq \frac{1}{2}\sqrt{2}|\beta| < |\beta|.$$

So with q as above, if we put $r = \alpha - q\beta$ then $N(r) < N(\beta)$, as desired. ■

Exercise 24.6 Show that, unlike in $\mathbb{Z}$, we cannot require q and r to be unique. [**Hint:** Examine the proof and come up with an example of an $\alpha \in \mathbb{Z}[i]$ where there are multiple suitable values for q and r.]

In $\mathbb{Z}$, we can find q and r by dividing and then splitting into the the floor plus what's left over (the *fractional part*). For example,

$$\frac{10}{4} = 2.5 = 2 + \frac{1}{2}$$

hence

$$10 = 2 \cdot 4 + 2.$$

We can do something similar in $\mathbb{Z}[i]$. For example,

$$\frac{5+8i}{3+i} = \frac{23}{10} + \frac{19}{10}i = (2+i) + \frac{3}{10} + \frac{9}{10}i$$

hence

$$5 + 8i = (2+i)(3+i) + \left(\frac{3}{10} + \frac{9}{10}i\right)(3+i) = (2+i)(3+i) + 30i.$$

Lecture 24 Problems

▶ 24.1. Let $\alpha, \beta, \gamma \in \mathbb{Z}[i]$. Prove that if $\alpha \mid \beta$ and $\alpha \mid \gamma$ then $\alpha \mid \beta x + \gamma y$ for all $x, y \in \mathbb{Z}[i]$.

24.2. Let $a, b \in \mathbb{Z}$. Prove that $a \mid b$ in $\mathbb{Z}[i]$ if and only if $a \mid b$ in $\mathbb{Z}$.

▶ 24.3. For each $n \in \{1, 2, 3, 4, 5\}$, determine all $\alpha \in \mathbb{Z}[i]$ such that $N(\alpha) = n$.

24.4. Let $\alpha, \beta \in \mathbb{Z}[i]$. Prove or disprove: If $\alpha \mid \beta$ and $\beta \mid \alpha$ then $\alpha = \pm\beta$.

24.5. Determine all Gaussian integers that divide 2.

Lecture 25 GCDs in $\mathbb{Z}[i]$

We continue our journey towards stating and proving a version of the Fundamental Theorem of Arithmetic in $\mathbb{Z}[i]$. Last time we formulated the Remainder Theorem for division in $\mathbb{Z}[i]$, and now our task is to introduce greatest common divisors and prove versions of the Bézout and Euclid lemmas. Everything will work out splendidly (suspiciously so).

One thing we had to contend with so far is the lack of uniqueness in the Remainder Theorem. This issue will also reappear when we try to define $\gcd(\alpha, \beta)$. What's at play is the existence of unexpected divisors of 1 in $\mathbb{Z}[i]$.

Proposition 25.1 **(Units in $\mathbb{Z}[i]$)**

Let $u \in \mathbb{Z}[i]$. The following statements are equivalent.

(a) $u \mid 1$ in $\mathbb{Z}[i]$.

(b) $\dfrac{1}{u} \in \mathbb{Z}[i]$.

(c) $N(u) = 1$.

(d) $u \in \{\pm 1, \pm i\}$.

Proof: We will prove that (a) $\Longrightarrow$ (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) $\Longrightarrow$ (a).

(a) $\Longrightarrow$ (b) is obvious. For (b) $\Longrightarrow$ (c), start with

$$u\frac{1}{u} = 1$$

and take norms of both sides to get

$$N(u)N\left(\frac{1}{u}\right) = N(1) = 1.$$

Since the norm of a Gaussian integer is a positive integer, the above equation implies $N(u) = N(1/u) = 1$.

Next, for (c) $\Longrightarrow$ (d), suppose that $u = a + ib$ (with $a, b \in \mathbb{Z}$) and $N(u) = 1$. Then

$$a^2 + b^2 = 1.$$

Thus, either $(a, b) = (\pm 1, 0)$ or $(a, b) = (0, \pm 1)$. This gives $u = \pm 1$ or $u = \pm i$, respectively.

Finally, for (d) $\Longrightarrow$ (a), we just note that $1 = (1)(1)$, $1 = (-1)(-1)$ and $1 = i(-i)$. ■

Definition 25.2 **Unit**

If $u \in \mathbb{Z}[i]$ satisfies any of the equivalent properties given in Proposition 25.1 then we say that u is a **unit** in $\mathbb{Z}[i]$.

REMARK

You should be reminded of the definition of units modulo n. (See especially the Remark following Definition 11.2.) What we have proved here is that

$$\mathbb{Z}[i]^\times = \{\pm 1, \pm i\}.$$

If we know that $\alpha \mid \beta$ then we also have that $u\alpha \mid \beta$ for any unit u (why?). For example, the two different-looking factorizations

$$5 = (1+2i)(1-2i) = (2-i)(2+i)$$

are actually not all that different since

$$1+2i = i(2-i) \quad \text{and} \quad 1-2i = (-i)(2+i)$$

that is, the divisors differ by units. This is similar in principle to the factorizations

$$10 = 2 \cdot 5 = (-2) \cdot (-5)$$

in $\mathbb{Z}$. We consider these factorizations to be essentially the same. Generally, in questions of divisibility, we should not differentiate too much between a divisor and a unit multiple of a divisor. This prompts the following definition.

Definition 25.3
Associates

Gaussian integers $\alpha, \beta \in \mathbb{Z}[i]$ are said to be **associates** if $\alpha = u\beta$ for some unit $u \in \mathbb{Z}[i]$.

The associates of $1+2i$ are

$$1+2i = (1)(1+2i), \quad -1-2i = (-1)(1+2i), \quad -2+i = i(1+2i) \quad \text{and} \quad 2-i = (-i)(1+2i).$$

We would like to now define the greatest common divisor of two Gaussian integers α and β to be their "largest" common divisor, by which we mean the common divisor δ with maximum norm $N(\delta)$. This is well-defined since if $\delta \mid \alpha$ then $N(\delta) \leq N(\alpha)$ so the norm of divisors of α is bounded. However, there is no unique such element. Indeed, if δ is a common divisor with maximum norm then so is $u\delta$ for any unit u.

The converse is also true.

Indeed, if $\delta, \delta' \in \mathbb{Z}[i]$ are common divisors with the same maximum norm then $\delta \mid \delta'$. (We will prove this below. This is the statement that a common divisor divides the gcd.) Thus, $\delta'/\delta \in \mathbb{Z}[i]$ and so

$$N(\delta'/\delta) = N(\delta')/N(\delta) = 1.$$

This shows that δ'/δ is a unit. Therefore, δ and δ' are associates.

Definition 25.4
Greatest Common Divisor, gcd

Let $\alpha, \beta \in \mathbb{Z}[i]$ not both be zero. A **greatest common divisor** of α and β is a common divisor $\delta \in \mathbb{Z}[i]$ of maximum norm.

By what we have just noted, any two greatest common divisors are associates, so we will abuse notation and simply write $\gcd(\alpha, \beta)$ to denote one of the four possible choices. We also define $\gcd(0, 0) = 0$.

At this point we can devise a version of the Euclidean algorithm for calculating $\gcd(\alpha, \beta)$, but we will forego this since we want to get to the Fundamental Theorem of Arithmetic as quickly as possible.

Theorem 25.5 **(Bézout's Lemma in $\mathbb{Z}[i]$)**

Let $\alpha, \beta \in \mathbb{Z}[i]$ not both be zero. Then there exist $x, y \in \mathbb{Z}[i]$ such that

$$\gcd(\alpha, \beta) = \alpha x + \beta y.$$

Before diving into the proof, let's note that the statement actually makes sense. Since there are 4 possible values of $\gamma = \gcd(\alpha, \beta)$, if we can write any one of them in the form $\gamma = \alpha x + \beta y$ then we can write the other three in this form as well. Indeed, we have $-\gamma = \alpha(-x) + \beta(-y)$ and $\pm i\gamma = \alpha(\pm ix) + \beta(\pm iy)$.

Proof: We will mimic the proof of Proposition 3.7 (Bézout's Lemma). Thus, let $S = \{\alpha x + \beta y \colon x, y \in \mathbb{Z}[i]\}$. Let $\gamma \in S$ be a non-zero element with smallest possible norm. Such an element exists by the well-ordering principle (applied to the set of norms $N(s)$ of non-zero elements $s \in S$).

Write $\gamma = \alpha x_0 + \beta y_0$. We claim that $\gamma \mid \alpha$ and $\gamma \mid \beta$. To prove this, apply the Remainder Theorem to write

$$\alpha = \gamma q + r$$

with $0 \leq N(r) < N(\gamma)$. Since $r = \alpha - \gamma q = (1 - qx_0)\alpha + (-qy_0)\beta$ is in S, we must have that $N(r) = 0$ by minimality of γ. Thus, $r = 0$. This shows that $\gamma \mid \alpha$. A similar argument shows that $\gamma \mid \beta$.

This shows that γ is a common divisor of α and β. Next, we must prove that γ has the largest norm amongst all common divisors δ. To prove this, write $\alpha = \delta z$ and $\beta = \delta w$ with $z, w \in \mathbb{Z}[i]$. Then

$$\gamma = \alpha x_0 + \beta y_0 = \delta(z x_0 + w y_0).$$

So $\delta \mid \gamma$ and therefore $N(\delta) \mid N(\gamma)$ by Proposition 24.3(c). Since norms are non-negative integers, it follows that $N(\delta) \leq N(\gamma)$, as required. ■

REMARK

The final paragraph in the preceding proof shows that every common divisor of α and β divides $\gcd(\alpha, \beta)$.

Our next target is Euclid's Lemma.

Definition 25.6 **Coprime**

We say that $\alpha, \beta \in \mathbb{Z}[i]$ are **coprime** if $\gcd(\alpha, \beta) = 1$.

In $\mathbb{Z}$, two integers are coprime if and only if ± 1 are their only common divisors. This generalizes in the appropriate way to $\mathbb{Z}[i]$.

Exercise 25.7 Prove that $\gcd(\alpha, \beta) = 1$ if and only if the only common divisors of α and β are ± 1 and $\pm i$.

Example 25.8 Show that $1 - 2i$ and $1 + 2i$ are coprime.

Solution: Let δ be a common divisor of $1 - 2i$ and $1 + 2i$. Then δ divides their sum

$$(1 - 2i) + (1 + 2i) = 2.$$

Hence $N(\delta) \mid N(2) = 4$. On the other hand, $N(\delta) \mid N(1 - 2i) = 1^2 + 2^2 = 5$. So $N(\delta)$ is a positive integer that divides both 4 and 5. Consequently, $N(\delta) = 1$ and so δ is a unit. It follows that $1 - 2i$ and $1 + 2i$ are coprime.

Example 25.9 Let $a, b \in \mathbb{Z}$. Show that if a and b are coprime in $\mathbb{Z}$ then a and b are coprime in $\mathbb{Z}[i]$.

Solution: By Bézout's Lemma in $\mathbb{Z}$, we can write

$$1 = ax + by$$

for some $x, y \in \mathbb{Z}$. If $\delta \in \mathbb{Z}[i]$ is a common divisor of a and b, then $\delta \mid ax + by = 1$, so δ is a unit. That is, the only Gaussian integer common divisors of a and b are units. So a and b must be coprime in $\mathbb{Z}[i]$.

Corollary 25.10 **(Euclid's Lemma in $\mathbb{Z}[i]$)**

Let $\alpha, \beta, \gamma \in \mathbb{Z}[i]$. If $\alpha \mid \beta\gamma$ and if α and β are coprime then $\alpha \mid \gamma$.

Exercise 25.11 Prove Corollary 25.10. [**Hint:** Mimic the proof of Proposition 4.3(a).]

Lecture 25 Problems

25.1. Let $\alpha, \beta \in \mathbb{Z}[i]$. Prove that α and β are coprime if and only if there exist $x, y \in \mathbb{Z}[i]$ such that $\alpha x + \beta y = 1$.

25.2. Let $a, b \in \mathbb{Z}$. Show that the gcd of a and b in $\mathbb{Z}$ is equal to their gcd if viewed as Gaussian integers (up to units).

25.3. Let $\alpha, \beta \in \mathbb{Z}[i]$ with $\beta \neq 0$. Let $\alpha = \beta q + r$ be as in the Remainder Theorem (Theorem 24.5). Prove that $\gcd(\alpha, \beta) = \gcd(\beta, r)$. [**Note:** This is a $\mathbb{Z}[i]$ version of Lemma 3.3.]

25.4. Let $\alpha, \beta, \gamma \in \mathbb{Z}[i]$ and assume that α and β are coprime. Show that if $\alpha \mid \gamma$ and $\beta \mid \gamma$ then $\alpha\beta \mid \gamma$. [**Note:** This is a $\mathbb{Z}[i]$ version of Proposition 4.3.]

Lecture 26 Gaussian Primes and Unique Factorization

We are almost ready to state and prove the Fundamental Theorem of Arithmetic in $\mathbb{Z}[i]$.

Definition 26.1
Gaussian Prime, Rational Prime

A **Gaussian prime** is a Gaussian integer $\pi \in \mathbb{Z}[i]$ that is a non-zero, non-unit whose only divisors are units and associates of π.

For clarity, a prime $p \in \mathbb{Z}$ will be referred to as a **rational prime**.

If π is a Gaussian prime and if $\pi = \alpha\beta$ then one of α and β must be a unit and the other one must be an associate of π. So the only factorizations of π in $\mathbb{Z}[i]$ are trivial—they are all of the form

$$\pi = u(u^{-1}\pi) \quad \text{where } u \text{ is a unit.}$$

This is analogous to how the only factorizations of a rational prime p in $\mathbb{Z}$ are

$$p = (1)p = (-1)(-p).$$

Let's look at some examples.

Example 26.2

The rational prime $p = 2$ is **not** a Gaussian prime. Indeed,

$$2 = 1^2 + 1^2 = (1+i)(1-i)$$

is a non-trivial factorization of 2 in $\mathbb{Z}[i]$. Note that

$$2 = (1+i)(1-i) = i(-i+1)(1-i) = i(1-i)^2$$

so 2 is in fact a unit times a square in $\mathbb{Z}[i]$. Weird.

Similarly, the rational prime $p = 5$ is not a Gaussian prime since

$$5 = 1^2 + 2^2 = (1+2i)(1-2i).$$

However, this time the divisors are not associates.

Example 26.3

Prove that $\pi = 1 - i$ is a Gaussian prime. (Note that this is one of the two divisors of 2 from the preceding example.)

Solution: Suppose $\alpha \mid \pi$. Then $N(\alpha) \mid N(\pi)$. Since $N(\pi) = 1^2 + 1^2 = 2$, it follows that $N(\alpha)$ is either 1 or 2. If $N(\alpha) = 1$ then α is a unit. If $N(\alpha) = 2$ then I claim that α is an associate of 2. To see why, write $\pi = \alpha\beta$ and then note that $N(\pi) = N(\alpha)N(\beta)$ forces $N(\beta)$ to be 1 and hence β must be a unit.

This shows that the only divisors of π are units and associates of π. Thus π must be a Gaussian prime.

Example 26.4

Prove that 3 is a Gaussian prime.

Solution: Suppose $\alpha \mid 3$. Then $N(\alpha) \mid N(3)$. Since $N(3) = 9$, it follows that $N(\alpha)$ is either 1, 3 or 9. If $N(\alpha) = 1$ then α is a unit and if $N(\alpha) = 9$ then, by an argument similar to the one in the preceding example, α is an associate of 3. So it suffices to prove that the case $N(\alpha) = 3$ cannot occur.

To see why this is so, note that if $\alpha = a + ib$ then $N(\alpha) = a^2 + b^2$ cannot be equal to 3 since a sum of squares cannot be equal to 3 mod 4.

The arguments in the preceding examples generalize to give the following result.

Proposition 26.5

(a) If $N(\alpha) = p$ is a rational prime, then α is a Gaussian prime.

(c) If p is a rational prime such that $p \equiv 3 \pmod 4$, then p is a Gaussian prime.

Exercise 26.6

Prove Proposition 26.5.

We now have a big supply of Gaussian primes. For example, each of the following is a Gaussian prime:

$$7, \quad 11, \quad 2+3i, \quad 2-3i, \quad 5+2i, \quad 5-2i.$$

In fact, we know what all Gaussian primes look like.

Theorem 26.7

(Classification of Gaussian Primes)

Every Gaussian prime is a unit multiple of one of the following Gaussian primes.

- $1 - i$.
- A rational prime such that $p \equiv 3 \pmod 4$.
- $\pi = a + ib$ where $N(\pi) = a^2 + b^2 = p$ is a rational prime such that $p \equiv 1 \pmod 4$.

Proposition 26.5 confirms that everything in the above list is a Gaussian prime. To prove that these are all possible Gaussian primes (up to associates) we will want to use the Fundamental Theorem of Arithmetic. So let's move towards that end.

Lemma 26.8

(Euclid's Lemma for Gaussian Primes)

Let $\pi \in \mathbb{Z}[i]$ be a Gaussian prime. Then if $\pi \mid \alpha\beta$ then $\pi \mid \alpha$ or $\pi \mid \beta$.

More generally, if $\pi \mid \alpha_1 \cdots \alpha_k$ then $\pi \mid \alpha_j$ for some j.

Proof: If π and α are not coprime, then $\gcd(\pi, \alpha)$ would be a non-unit divisor of π hence must be an associate of π. This shows that $\pi \mid \alpha$. Otherwise, if π and α are coprime, then $\pi \mid \beta$ by Corollary 25.10.

The general statement follows by induction. ■

Theorem 26.9 **(Fundamental Theorem of Arithmetic in $\mathbb{Z}[i]$)**

Every non-zero Gaussian integer that is not a unit can be written as a product of Gaussian primes in a unique way (up to re-ordering and associates).

For example, the factorizations

$$5 = (1+2i)(1-2i) = (2+i)(2-i)$$

are viewed as being the same same since $(1+2i) = i(2-i)$ and $(1-2i) = (-i)(2+i)$, so the factors are the same up to associates.

Proof of Theorem 26.9:

Existence: Suppose the theorem were false. Then there would exist a non-zero Gaussian integer α that is not a product of Gaussian primes. By the well-ordering principle, there would exist such an α of least norm. Note that α itself cannot be prime so $\alpha = \beta\gamma$ with $N(\beta) < N(\alpha)$ and $N(\gamma) < N(\alpha)$. By minimality, it follows that β and γ are products of Gaussian primes, and hence so is $\alpha = \beta\gamma$. Contradiction!

Uniqueness: If $\alpha = \pi_1 \cdots \pi_k = \pi_1' \cdots \pi_l'$ then π_1 must divide $\pi_1' \cdots \pi_l'$ hence (re-labeling the π_j' if necessary) $\pi_1 \mid \pi_1'$. So we can cancel off π and π_1' prime from both sides, leaving us with

$$\pi_2 \cdots \pi_k = u\pi_2' \cdots \pi_l'$$

where u is a unit. Now repeat the argument. ■

Example 26.10 Write $\alpha = 3 + 24i$ as a product of Gaussian primes.

Solution: We have $\alpha = 3(1+8i)$ and 3 is a Gaussian prime. So we must factor $1+8i$ into Gaussian primes.

If π is a prime divisor of $1+8i$ then $\pi \mid (1+8i)(1-8i) = 1+64 = 65$. (In general, if $\pi \mid \alpha$ then $\pi \mid N(\alpha) = \alpha\overline{\alpha}$.) Now since $65 = 5 \cdot 13$, we know that either $\pi \mid 5$ or $\pi \mid 13$.

We have

$$5 = 1^2 + 2^2 = (1+2i)(1-2i) \quad \text{and} \quad 13 = 2^2 + 3^2 = (2+3i)(2-3i).$$

So π must be one of the above four prime divisors. Note that only one of the primes divisors of 5 can be a divisor of $1+8i$. (If both were, then their product 5 would divide $1+8i$, but it does not.) Likewise, only one of the prime divisors of 13 can be a divisor of $1+8i$.

So now we just check the various possible products. We have

$$(1+2i)(2+3i) = -4+7i$$

which is no good. On the other hand,

$$(1-2i)(2+3i) = 8-i.$$

This is not exactly $1+8i$ but it is an associate: $i(8-i) = (1+8i)$. Thus,

$$1+8i = i(1-2i)(2+3i) = (2+i)(2+3i).$$

Consequently,

$$3+24i = 3(2+i)(2+3i)$$

and each of the three factors on the right is a Gaussian prime.

Exercise 26.11 Write $\alpha = 4 - 18i$ as a product of Gaussian primes.

We now have the tools to determine which integers can be written as the sum of two squares. Let's deal with the case of primes now, and we'll take up the general case next lecture.

Theorem 26.12 (Primes of the Form $p = x^2 + y^2$)

Let p be a rational prime. Then $p = x^2 + y^2$ with $x, y \in \mathbb{Z}$ if and only if $p = 2$ or $p \equiv 1 \pmod 4$.

Proof: If $p = 2$ then $p = 1^2 + 1^2$. So assume p is odd. If $p \equiv 3 \pmod 4$, then p cannot be the sum of two squares (see, e.g., Exercise 2.12). So it remains to prove that if $p \equiv 1 \pmod 4$ then p is a sum of two squares.

If $p \equiv 1 \pmod 4$ then $\left(\frac{-1}{p}\right) = 1$. Thus, there exists an $n \in \mathbb{Z}$ such that $n^2 \equiv -1 \pmod p$ and hence

$$p \mid n^2 + 1 = (n + i)(n - i).$$

If p were a Gaussian prime, then it would divide one of the factors on the right. However, $p \nmid n + i$ and $p \nmid n - i$. (Since neither of $\frac{n}{p} \pm \frac{1}{p}i$ is a Gaussian integer.) Thus, p is not a Gaussian prime, and so we can factor it into a product of two non-units, i.e., we can write

$$p = \alpha\beta \quad \text{with} \quad 1 < N(\alpha) < N(p) \quad \text{and} \quad 1 < N(\beta) < N(p).$$

Thus, $N(\alpha) = p$ since $N(p) = p^2$. Writing $\alpha = x + iy$, we find that

$$p = N(x + iy) = x^2 + y^2,$$

as required. ■

You should take a moment to appreciate all the ingredients that went into this proof. Would you have expected the Legendre symbol $\left(\frac{-1}{p}\right)$ to pop up?

Let's close the lecture by completing the proof of the classification of Gaussian primes. The idea behind the proof is simple (and ends up being important in algebraic number theory): Where can we find Gaussian primes? *Answer:* They occur as divisors of rational primes! So by factoring all rational primes p in $\mathbb{Z}[i]$, we obtain all possible Gaussian primes.

Proof of Theorem 26.7 (Classification of Gaussian Primes): Let π be a Gaussian prime. Let's start by proving that π must divide a rational prime. The key observation is that $\pi \mid N(\pi)$ since $N(\pi) = \pi\overline{\pi}$. However, $N(\pi)$ is an integer greater than 1 (why?), so it is a product of rational primes. By Euclid's Lemma, since $\pi \mid N(\pi)$, π must divide one of these rational primes—call it p.

If $p = 2$, then we have determined in Examples 26.2 and 26.3 that the Gaussian prime divisors of 2 are unit multiples of $1 + i$.

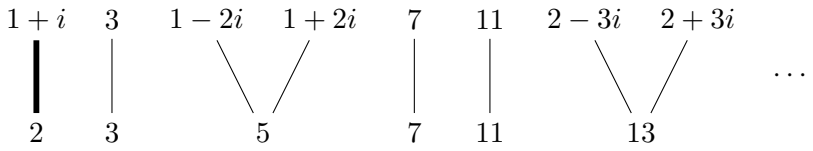
If p is odd then either $p \equiv 1 \pmod 4$ or $p \equiv 3 \pmod 4$. If $p \equiv 3 \pmod 4$ then p is a Gaussian prime (by Proposition 26.5 (b)) so since $\pi \mid p$, π must be a unit multiple of p.

Finally, if $p \equiv 1 \pmod 4$ then $p = a^2 + b^2 = (a - ib)(a + ib)$ for some $a, b \in \mathbb{Z}$. Since $\pi \mid p$, it follows from Euclid's Lemma that π divides one of $a \pm ib$. Writing $a \pm ib = \pi\alpha$ and taking norms, we find that

$$p = N(a \pm ib) = N(\pi)N(\alpha).$$

Since p is prime and $N(\pi) > 1$, it must be the case that $N(\alpha) = 1$. Thus, α is a unit. This shows that π is a unit multiple of $a \pm ib$. So, in all cases, π is a unit multiple of one of the Gaussian primes in the given list. ■

The diagram below lists the Gaussian primes obtained by factoring the first several rational primes.



The line above $p = 2$ was drawn a bit thicker to emphasize the fact that 2 factors into $(1 + i)^2$ (times a unit). It's almost as though 2 wants to behave like 5 and 13 which split into two factors, but its two factors happen to coincide... (This phenomenon is known as *ramification.*)

REMARK (Primes of the Form $p = x^2 + ny^2$)

Fermat was the first to prove that an odd prime p is the sum of two squares if and only if $p \equiv 1 \pmod 4$. The proof we've presented above is due to Richard Dedekind, one of the founders of algebraic number theory—the subject whose goal is to generalize number theory from $\mathbb{Z}$ to number systems such as $\mathbb{Z}[i]$.

Fermat also claimed:

$$p = x^2 + 2y^2 \iff p \equiv 1, 3 \pmod 8 \text{ or } p = 2$$
$$p = x^2 + 3y^2 \iff p \equiv 1 \pmod 8 \text{ or } p = 3.$$

Euler was able to prove both of these results (after considerable effort), as well as our result concerning $p = x^2 + y^2$. In the course of this, he discovered—but could not prove—the Law of Quadratic Reciprocity. Euler's proofs hinged on the fact that prime divisors of integers of the form $x^2 + ny^2$, where $\gcd(x, y) = 1$, are themselves also of the form $x^2 + ny^2$. This is true if $n = 1, 2, 3$ but not if $n = 5$. Euler conjectured that

$$p = x^2 + 5y^2 \iff p \equiv 1, 9 \pmod{20} \text{ or } p = 5.$$

This was proved by Lagrange, who developed the theory of binary quadratic forms for this purpose.

The story of primes of the form $p = x^2 + ny^2$ is a fascinating one that quickly gets intertwined with some of the deepest results in number theory (e.g. quadratic reciprocity, as Euler had noticed). For example, it is a theorem that, for $p \neq 2, 7$,

$$p = x^2 + 14y^2 \iff \begin{cases} \text{the congruences} \\ u^2 = -14 \pmod p \text{ and } (v^2 + 1)^2 = 8 \pmod p \\ \text{have solutions.} \end{cases}$$

The first congruence is asserting that $\left(\frac{-14}{p}\right) = 1$, which we can understand via quadratic reciprocity. The second congruence is hinting at something more mysterious.

For much more on this, I recommend David Cox's beautiful book, *Primes of the Form* $x^2 + ny^2$, which would make good reading after you have finished PMATH 340.

Lecture 26 Problems

26.1. Let $\pi \in \mathbb{Z}[i]$ be a non-zero, non-unit. Prove that π is a Gaussian prime if and only if whenever $\pi \mid \alpha\beta$ then $\pi \mid \alpha$ or $\pi \mid \beta$.

26.2. Let $p \equiv 1 \pmod 4$ be a rational prime. Prove that p is a sum of two squares in a unique way (up to signs and re-ordering). That is, if $p = a^2 + b^2 = c^2 + d^2$, then $(c, d) = (\pm a, \pm b)$ or $(c, d) = (\pm b, \pm a)$. [**Hint:** Connect the expression of p as a sum of two squares to the factorization of p into a product of Gaussian squares and use unique factorization.]

▶ 26.3. Prove that if $n, m \in \mathbb{Z}$ can be written as the sum of two squares then nm can also be written as the sum of two squares. [**Hint:** If $n = x^2 + y^2$ then $n = N(x + iy)$.]

▶ 26.4. Let $\alpha, \beta \in \mathbb{Z}[i]$ be coprime. Prove that if $\alpha\beta = \gamma^n$ for some $n \geq 2$ then each of α and β is a unit times an nth power of a Gaussian integer.

26.5. Prove that there are infinitely many Gaussian primes.

Lecture 27 Sums of Squares and Pythagorean Triples

Our goal in this lecture is to tackle two Diophantine equations from Lecture 1, namely

$$x^2 + y^2 = n \quad \text{and} \quad x^2 + y^2 = z^2.$$

The first asks us to determine the integers n that can be expressed as the sum of two squares. The second asks for integer solutions to the Pythagorean equation—and, therefore, for the possible right-angled triangles with integer side lengths.

27.1 Sums of Two Squares

We have already determined which primes p can be written as the sum of two squares. They are $p = 2$ and all odd primes congruent to 1 mod 4. What about composite integers?

One neat observation is that if n and m are each the sum of two squares, then so is their product nm. (That is, the set of integers that are sums of two squares is "closed under multiplication".) This is Problem 26.3 but let's restate it here again as a lemma.

Lemma 27.1 $(a^2 + b^2)(A^2 + B^2) = (aA - bB)^2 + (aB + bA)^2$.

Proof: Of course, we can just multiply out both sides and show that they coincide. However, here is a more conceptual approach. Note that

$$a^2 + b^2 = N(a + ib) \quad \text{and} \quad A^2 + B^2 = N(A + iB).$$

So, since the norm is multiplicative, we have

$$\begin{aligned}(a^2 + b^2)(A^2 + B^2) &= N((a + ib)(A + iB)) \\ &= N((aA - bB) + i(aB + bA)) \\ &= (aA - bB)^2 + (aB + bA)^2,\end{aligned}$$

as desired. ■

Exercise 27.2 State and prove an analogous result for integers of the form $a^2 + Db^2$.

For example, since

$$5 = 1^2 + 2^2 \quad \text{and} \quad 13 = 2^2 + 3^2$$

we discover that

$$65 = 5 \cdot 13 = (1 \cdot 2 - 2 \cdot 3)^2 + (1 \cdot 2 + 2 \cdot 3)^2 = 4^2 + 7^2.$$

Using $5 = 2^2 + 1^2$, we also find that

$$65 = (2 \cdot 2 - 1 \cdot 3)^2 + (2 \cdot 3 + 1 \cdot 2)^2 = 1^2 + 8^2.$$

As another example, we have

$$72 = 8 \cdot 9 = (2^2 + 2^2)(3^2 + 0^2) = 6^2 + 6^2.$$

Theorem 27.3 **(Integers of the form $n = x^2 + y^2$)**

Let $n \in \mathbb{Z}_{>0}$. Then n can be written as a sum of two squares if and only if $v_p(n)$ is even for all primes $p \equiv 3 \pmod 4$.

The condition in the theorem says that, in the prime factorization of n, all primes congruent to 3 mod 4 must appear to an even power (possibly 0). For example, $65 = 5 \cdot 13$ and $72 = 2^3 3^2$ both satisfy this condition (and as seen above they are sums of two squares). On the other hand, $275 = 5^2 \cdot 11$ doesn't since $v_{11} = 1$; thus $275 = x^2 + y^2$ has no solutions in the integers.

Proof: If $n = 1$ then the assertion is trivially true, so assume that $n > 1$ and consider the prime factorization

$$n = 2^a \prod_i p_i^{e_i} \prod_j q_j^{f_j}$$

where the p_i are congruent to 1 mod 4 and the q_j are congruent to 3 mod 4. We must prove that n is a sum of two squares if and only if all of the f_j are even.

Assume all of the $f_j = 2h_j$ are even. Then since

- $2 = 1^2 + 1^2$ is a sum of two squares,
- each p_i is a sum of two squares, and
- each $q_j^{f_j} = (q_j^{h_j})^2 + 0^2$ is a sum of two squares,

n is a sum of two squares by a repeated application of Lemma 27.1.

Conversely, assume that n is a sum of two squares, say $n = x^2 + y^2$. We will must show that all of the f_j are even. Suppose to the contrary that some f_j is odd. We may assume, without loss of generality, that $q = q_j$ doesn't divide x (exercise!). Then since $q \mid n$, we find that

$$x^2 + y^2 \equiv 0 \pmod q$$

and therefore, by multiplying through by x^{-1} mod q and re-arranging, we have

$$-1 = (yx^{-1})^2 \pmod q.$$

This shows that -1 is a square mod q, which is a contradiction since $q \equiv 3 \pmod 4$. ■

Exercise 27.4 Explain why we are allowed to assume that $q \nmid x$ in the final paragraph of the preceding proof.

Now that we know *which* integers n are sums of two squares, the natural follow-up question is: How do you actually find $x, y \in \mathbb{Z}$ so that $n = x^2 + y^2$? There is an obvious brute-force approach: Try every positive integer $x \leq \sqrt{n}$ until you find one such that $n - x^2$ is a perfect square. This is not very efficient—and there are more efficient algorithms—but it will suffice for our purposes.

The other question we might ask is: How many ways is n a sum of two squares? For example, above we discovered that 65 has two such representations:

$$65 = 4^2 + 7^2 = 1^2 + 8^2.$$

Are there any others?

More generally, let $r_2(n)$ denote the number of solutions (x, y) to the equation $n = x^2 + y^2$ with $x, y \in \mathbb{Z}$ (which we allow to be negative or zero). For example, $r_2(5) = 8$, since

$$5 = (\pm 1)^2 + (\pm 2)^2 = (\pm 2)^2 + (\pm 1)^2$$

and there are no other representations of 5 as a sum of two squares.

The following theorem, which we state without proof, tells us how to determine $r_2(n)$ from the prime factorization of n.

Theorem 27.5 Suppose that

$$n = 2^a \prod_i p_i^{e_i} \prod_j q_j^{f_j}$$

where the p_i are primes congruent to 1 mod 4 and the q_j are primes congruent to 3 mod 4. Then

$$r_2(n) = \begin{cases} 4 \prod (1 + e_i) & \text{if all of the } f_j \text{ are even} \\ 0 & \text{otherwise.} \end{cases}$$

For example,

- $r_2(5) = 4(1 + 1) = 8$ just as we've determined above. More generally, if p is any prime congruent to 1 mod 4, then $r_2(p) = 8$. Thus there are 8 ways of writing p as $p = x^2 + y^2$. Taking into account sign changes $(\pm x, \pm y)$ (which multiply the count by 4; since $x \neq 0$ and $y \neq 0$) and switching the order from (x, y) to (y, x) (which double the count; since we can be sure there are no solutions with $x = y$), this shows that p can be *uniquely* as a sum of two squares. (Refer back to Problem 26.2.)
- Since $65 = 5 \cdot 13$, we find that $r_2(65) = 4(1 + 1)(1 + 1) = 16$. Again, signs and re-ordering multiply the count by 8 (since 65 is not a perfect square and since $65 \neq x^2 + x^2$). So there are really only two essentially distinct representations of 65 as a sum of two squares, namely the ones we found above:

$$65 = 4^2 + 7^2 = 1^2 + 8^2.$$

- Consider now $50 = 2 \cdot 5^2$. Then $r_2(50) = 4(1+2) = 12$. This time we have $50 = x^2 + x^2$ (with $x = \pm 5$). Ignoring these four solutions, we are left with 8. These will be of the form (x, y) with $x \neq y$. Also, neither x nor y can be zero since 50 is not a perfect square. Thus, sign changes and order-swaps multiply the solution count by 8—meaning, there is only one other essentially different way of writing 50 as a sum of two squares. It is given by

$$50 = 1^2 + 7^2 = (\pm 1)^2 + (\pm 7)^2 = (\pm 7)^2 + (\pm 1)^2.$$

Exercise 27.6 Determine $r_2(100)$ and then find all $x, y \in \mathbb{Z}$ such that $100 = x^2 + y^2$.

REMARK (Sums of k Squares)

- Gauss and Legendre independently proved that a positive integer n is a sum of three squares if and only if n is not of the form $n = 4^a(8b+7)$ where $a, b \in \mathbb{Z}_{\geq 0}$. Gauss was also able to find a formula for $r_3(n)$, the number of solutions to $x^2 + y^2 + z^2 = n$ with $x, y, z \in \mathbb{Z}$. His formula is in terms of things we haven't covered in this course, so I won't give it here.
- Lagrange proved that *every* positive integer is the sum of four squares, and Jacobi proved that
$$r_4(n) = 8 \sum_{\substack{d|n \\ 4\nmid d}} d,$$
where the sum is over the positive divisors of n that are not divisible by 4. For example, $r_4(3) = 8(1+3) = 32$, and indeed
$$3 = (\pm 1)^2 + (\pm 1)^2 + (\pm 1)^2 + 0^2.$$
There are 8 choices of sign and 4 choices for where to place 0, giving a total of 32 different representations.
- Because of Lagrange's theorem, we know that every positive integer can be written as a sum of k squares for any $k \geq 4$ (just use enough 0s). Of course, a more interesting question is whether we can use non-zero squares. One result in this direction is that every integer $n > 33$ can be written as a sum of five positive squares. A proof for $n > 169$ is given in Corollary 6.27 of Niven, Montgomery and Zucker; and then you can check $34 \leq n \leq 169$ by hand (for example, $169 = 5^2 + 6^2 + 6^2 + 6^2 + 6^2$).

27.2 Pythagorean Triples

Recall that a **Pythagorean triple** is a triple (a, b, c) of positive integers satisfying the Pythagorean equation
$$a^2 + b^2 = c^2.$$
A Pythagorean triple (a, b, c) is called **primitive** if $\gcd(a, b, c) = 1$. We will determine all primitive Pythagorean triples. (We can then find all Pythagorean triples by scaling the primitive ones.)

Lemma 27.7 If (a, b, c) is a primitive Pythagorean triple, then:

(a) a, b, c are pairwise coprime.

(b) One of a and b is odd and the other is even.

(c) c is odd.

Proof:

(a) If d divides any two of a, b or c then it must divide the third, since $a^2 + b^2 = c^2$. So d divides $\gcd(a, b, c) = 1$ and therefore $d = \pm 1$. So any two of a, b, c are coprime.

(b) Since $\gcd(a, b, c) = 1$, a, b, c cannot all be even. So at least one of a and b must be odd (because if they were both even then c would be even). If they're both odd then $a^2 + b^2 \equiv 1 + 1 \equiv 2 \pmod 4$. Hence $a^2 + b^2$ cannot be equal to c^2 since 2 is not a square mod 4.

(c) This follows from part (b). ■

Theorem 27.8 **(Classification of Primitive Pythagorean Triples)**

Let (a, b, c) be a primitive Pythagorean triple. Then c is odd and only one of a and b is even. Assume that a is odd and b is even. Then there exist coprime integers $m, n \in \mathbb{Z}_{>0}$ one of which is even and the other one odd such that $m > n$ and

$$a = m^2 - n^2, \quad b = 2mn \quad \text{and} \quad c = m^2 + n^2.$$

Note that if $(a, b, c) = (m^2 - n^2, 2mn, m^2 + n^2)$ are as in the theorem, then

$$\begin{aligned} a^2 + b^2 &= (m^2 - n^2)^2 + (2mn)^2 \\ &= m^4 - 2n^2m^2 + n^4 + 4m^2n^2 \\ &= m^4 + 2m^2n^2 + n^4 \\ &= (m^2 + n^2)^2 \\ &= c^2. \end{aligned}$$

So every such triple is a Pythagorean triple. The theorem asserts that *every* primitive Pythagorean triple can be obtained in this fashion.

Proof of Theorem 27.8: We have

$$c^2 = (a + ib)(a - ib).$$

Assume for the moment that $a + ib$ and $a - ib$ are coprime. Then, by Problem 26.4, $a + ib$ must be a unit times a square, say

$$a + ib = u(m + ni)^2 = u((m^2 - n^2) + 2mni).$$

where $u \in \{\pm 1, \pm i\}$ is a unit. If $u = \pm i$ then by equating real parts we fin that $a = \pm 2mn$, contrary to our assumption that a is odd. Thus $u = \pm 1$, so

$$a = \pm(m^2 - n^2) \quad \text{and} \quad b = \pm 2mn.$$

By swapping m and n if necessary to arrange that $m > n$, we can assume that the signs above are $+$ (since $a > 0$). It then follows that

$$c^2 = a^2 + b^2 = (m^2 - n^2)^2 + (2mn)^2 = (m^2 + n^2)^2$$

and thus

$$c = m^2 + n^2$$

since $c > 0$. I'll leave it as an exercise for you to check that m and n must be coprime and that one must be odd and the other must be even.

It remains to prove that $a + ib$ and $a - ib$ are coprime in $\mathbb{Z}[i]$. Let π be a Gaussian prime that divides both $a \pm ib$. Then

$$\pi \mid (a+ib) - (a-ib) = 2ib \quad \text{and} \quad \pi \mid (a+ib) + (a-ib) = 2a.$$

Thus, π divides $2b$ (since i is a unit) and $2a$. Note that π cannot divide 2, since if otherwise we would have $\pi = u(1+i)$ for some unit i and hence $N(\pi) = 2$. However, since $\pi \mid a + ib$ we have that $N(\pi) \mid N(a^2 + b^2) = c^2$ and c is odd, so $N(\pi)$ must be odd. Thus $\pi \nmid 2$ and so by Euclid's Lemma π must divide a and b and hence will divide $\gcd(a, b)$—but $\gcd(a, b) = 1$ by Lemma 27.7. Contradiction. So no Gaussian prime divides both $a \pm ib$, and therefore they must be coprime. ■

Exercise 27.9 Show that if $m > n$ are positive integers and $(a, b, c) = (m^2 - n^2, 2mn, m^2 + n^2)$ is a primitive Pythagorean triple, then $\gcd(m, n) = 1$ and one of m and n must be even and the other one must be odd.

We can now easily generate Pythagorean triples (a, b, c). Here's a list of the first few primitive ones.

m	n	$a = m^2 - n^2$	$b = 2mn$	$c = m^2 + n^2$
2	1	3	4	5
3	2	5	12	13
4	1	15	8	17
4	3	7	24	25
5	2	21	20	29
5	4	9	40	41

Lecture 27 Problems

27.1. For $n \in \mathbb{Z}_{>0}$, let $r_2^+(n)$ be the number of pairs of *positive* integers (x, y) such that $x^2 + y^2 = n$. Prove:

$$r_2^+(n) = \begin{cases} \frac{1}{4} r_2(n) & \text{if } n \text{ is not a perfect square} \\ \frac{1}{4} r_2(n) - 1 & \text{if } n \text{ is a perfect square.} \end{cases}$$

27.2. Determine all positive integers n that can be written as the difference of two squares.
[**Hint:** Go through $n = 1, 2, 3, \ldots$ and see if you can spot any patterns.]

27.3. (a) Find all primitive Pythagorean triples (a, b, c) with $b = 10$ or prove that none exist.
(b) Find all Pythagorean triples (a, b, c) (not necessarily primitive) with $b = 18$ or prove that none exist.

27.4. Let (a, b, c) be a Pythagorean triple. Prove that $3 \mid ab$ and $60 \mid abc$.

Lecture 28 The Mordell Equation

In this lecture we will investigate the equation

$$y^2 = x^3 + k,$$

where $k \in \mathbb{Z}$ is non-zero. This equation is named after L.J. Mordell who published substantial results about it. In particular, Mordell proved that, for any given $k \neq 0$, this equation has finitely many solutions in $\mathbb{Z}$. We've seen two special cases in Example 20.7 and Exercise 20.8. Historically, Fermat had claimed and Euler had proved[17] that the only integer solutions to $y^2 = x^3 - 2$ are $(x, y) = (3, \pm 5)$.

In the previous couple of lectures, we tackled the Diophantine equations

$$x^2 + y^2 = n \quad \text{and} \quad x^2 + y^2 = z^2$$

by utilizing the factorization

$$x^2 + y^2 = (x + iy)(x - iy)$$

in $\mathbb{Z}[i]$, and then leveraging our development of number theory in the Gaussian integers.

Our approach to the Mordell equation will be similar. We will begin with the factorization

$$x^3 = y^2 - k = (y - \sqrt{k})(y + \sqrt{k})$$

and then we will "do number theory" in the extended number system

$$\mathbb{Z}[\sqrt{k}\,] = \{a + b\sqrt{k} \colon a, b \in \mathbb{Z}\}.$$

This approach was pioneered by Euler, who more generally studied "arithmetic surds" such as $a\sqrt{k} + b\sqrt{l}$. There will be a couple of surprises in store for us (just like there were for Euler!).

The Case $k = -1$

Let's look at the Mordell equation

$$y^2 = x^3 - 1 \quad \Longleftrightarrow \quad x^3 = y^2 + 1.$$

The factorization trick

$$x^3 = (y + i)(y - i)$$

places us in the now-familiar world of Gaussian integers.

Lemma 28.1 If $x, y \in \mathbb{Z}$ are such that $x^3 = y^2 + 1$ then y must be even.

Proof: If to the contrary y were odd, then x^3 hence x would be even. Consequently, $x^3 \equiv 0 \pmod 4$ and therefore $y^2 \equiv -1 \pmod 4$. But -1 is not a square mod 4. Contradiction. ■

[17] Well, almost. He was a bit too fast and loose with the rules of number theory.

Theorem 28.2 The only solution to $y^2 = x^3 - 1$ in the integers is $(x, y) = (1, 0)$.

Proof: I claim that $y + i$ and $y - i$ are coprime in $\mathbb{Z}[i]$. I will prove this momentarily, but let's grant it for now. Then, since their product is a cube, it must be the case that $y \pm i$ is a unit times a cube in $\mathbb{Z}[i]$ (by Problem 26.4). Coincidentally, the units in $\mathbb{Z}[i]^\times$, namely ± 1 and $\pm i$, are themselves cubes: $\pm 1 = (\pm 1)^3$ and $\pm i = (\mp i)^3$.

So if we have $y + i = u\alpha^3$ with u a unit, we might as well absorb the unit into the cube. Thus,

$$y + i = (a + bi)^3 = (a^3 - 3ab^2) + (3a^2b - b^3)i.$$

Equating imaginary parts, we find

$$1 = b(3a^2 - b^2)$$

so

$$b = \pm 1 \quad \text{and} \quad 3a^2 - b^2 = \pm 1.$$

The second equation gives $3a^2 - 1 = \pm 1$. If $b = 1$ then $3a^2 = 2$ which has no solutions in $\mathbb{Z}$. If $b = -1$ then $3a^2 = 0$ so $a = 0$. Consequently,

$$y = a^3 - 3ab^2 = 0$$

and

$$x^3 = y^2 + 1 = 1.$$

Thus, $(x, y) = (1, 0)$, as claimed.

It remains to prove that $y + i$ and $y - i$ are coprime. To see this, let π be a Gaussian prime that divides both. Then

$$\pi \mid (y + i) - (y - i) = 2i.$$

So $\pi \mid 2$ since i is a unit. Thus, $\pi = u(1 + i)$, where $u \in \mathbb{Z}[i]$ is a unit, since $1 + i$ is the only Gaussian prime (up to units) that divides 2. On the other hand, $\pi \mid y + i$ hence $N(\pi) \mid N(y + i) = y^2 + 1$ which is odd by Lemma 28.1. This is a contradiction since $N(\pi) = N(1 + i) = 2$. So $y \pm i$ do not have a common Gaussian prime divisor, hence they must be coprime. ■

The Case $k = 2$

Let's now apply the same approach to the equation

$$y^2 = x^3 + 2 \quad \Longleftrightarrow \quad x^3 = y^2 - 2 = (y - \sqrt{2})(y + \sqrt{2}).$$

This factorization is taking place in the number system

$$\mathbb{Z}[\sqrt{2}\,] = \{a + b\sqrt{2} \colon a, b \in \mathbb{Z}\}.$$

Assume (!) for the moment that this number system obeys the same arithmetic rules as the regular integers $\mathbb{Z}$ and the Gaussian integers $\mathbb{Z}[i]$. If we can prove that $y - \sqrt{2}$ and $y + \sqrt{2}$ are coprime, then we would be able to deduce that each is a unit times a cube, and

then upon absorbing the unit into the cube like we did in the previous section, we end up considering

$$y + \sqrt{2} = (a + b\sqrt{2})^3 = (a^3 + 6ab^2) + (3a^2b + 2b^3)\sqrt{2}.$$

By equating coefficients of $\sqrt{2}$, we end up with

$$1 = 3a^2b + 2b^3 = b(3a^2 + 2b^2).$$

There are no solutions to this equation because if b is non-zero then the right-side will be greater than 1 in absolute value, while if $b = 0$ the right-side is 0. Thus, it appears we have proved:

Theorem (?): The equation $y^2 = x^3 + 2$ has no solutions in the integers.

Wait—we still have to prove that $y - \sqrt{2}$ and $y + \sqrt{2}$ are coprime! To this end suppose that π is a prime divisor of both. Then

$$\pi \mid (y + \sqrt{2}) - (y - \sqrt{2}) = 2\sqrt{2} = (\sqrt{2})^3.$$

I claim that $\sqrt{2}$ itself is also prime, that is, it cannot be factored into two elements of $\mathbb{Z}[\sqrt{2}]$. To prove this, let's define a norm function by

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

You can easily check that this norm satisfies the properties of the Gaussian norm as given in Proposition 24.3. So here is the proof that $\sqrt{2}$ is prime. Suppose

$$\sqrt{2} = \alpha\beta.$$

Then, upon taking norms, we find

$$2 = N(\sqrt{2}) = N(\alpha)N(\beta).$$

Since $N(\alpha)$ and $N(\beta)$ are integers and 2 is prime, one of them must be ± 1. Thus, one of α or β must be a unit. So $\sqrt{2}$ admits no non-trivial factorizations. Consequently, since $\pi \mid (\sqrt{2})^3$, we must have that $\pi = \sqrt{2}$ (or a unit multiple of $\sqrt{2}$). On the other hand, we are assuming that

$$\pi \mid y + \sqrt{2}.$$

Thus, $N(\pi) \mid N(y + \sqrt{2})$, or equivalently, $2 \mid y^2 + 2$. However, by an argument similar to the one in Lemma 28.1, we can show that y is odd. We've arrived at our desired contradiction! Thus, $y + \sqrt{2}$ and $y - \sqrt{2}$ are coprime.

Before reading ahead, stop and attempt the next exercise.

Exercise 28.3 Inspect the previous argument and try to identity any errors.

There must, in fact, be errors because the equation $y^2 = x^3 + 2$ has solutions in the integers—for example, $(x, y) = (-1, \pm 1)$.[18]

Your first objection to the proof above should be that we assumed (!) at the very beginning that $\mathbb{Z}[\sqrt{2}]$ is as nice $\mathbb{Z}[i]$. This actually turns out to be not far from the truth. Indeed, we can prove a Remainder Theorem, define primality and gcds, all the way up to unique

[18] These are actually the *only* integer solutions, but we won't prove that here.

factorizations into primes, completely analogously to what we did with $\mathbb{Z}[i]$. Once we do this, the proof that $y+\sqrt{2}$ and $y-\sqrt{2}$ (for odd $y \in \mathbb{Z}$) that we gave above is fully correct!

The fundamental error occurred when we claimed that we could "absorb units into the cube." We were able to this in $\mathbb{Z}[i]$ because the units ± 1 and $\pm i$ themselves were cubes. Is this true in $\mathbb{Z}[\sqrt{2}\,]$?

To find the units in $\mathbb{Z}[\sqrt{2}\,]$, we need to look for $\alpha = a + b\sqrt{2}$ such that $N(\alpha) \mid 1$, i.e.

$$a^2 - 2b^2 \mid 1 \quad \iff \quad a^2 - 2b^2 = \pm 1.$$

We will study this equation next lecture. For now let me just mention that it has infinitely many solutions. For example, $(a,b) = (3,2)$ is a solution since

$$N(3+2\sqrt{2}) = 3^2 - 2 \cdot 2^2 = 1.$$

And we can now generate infinitely many solutions by taking powers of $3+2\sqrt{2}$, since

$$N((3+2\sqrt{2})^n) = (N(3+2\sqrt{2}))^n = 1^n = 1.$$

For example,

$$(3+2\sqrt{2})^2 = 17 + 12\sqrt{2}$$

gives the solution

$$17^2 - 2 \cdot 12^2 = 1.$$

So, where does that leave us? Since there are infinitely many units in $\mathbb{Z}[\sqrt{2}\,]$, to deal with the equation $x^2 = y^2 - 2$, we need a more refined approach that involves considering something like

$$y + \sqrt{2} = u(a + b\sqrt{2})^3$$

where u is a unit. This leads to a complicated case-by-case analysis. Too complicated for us to take up here! (That said, there is a modification of this approach that works nicely. It uses some ideas from algebraic number theory that are beyond the scope of PMATH 340.)

The Case $k = -11$

Okay, $y^2 = x^3 + 2$ didn't quite work out so well, but we won't be deterred! Let's now consider

$$y^2 = x^3 - 11 \quad \iff \quad x^3 = y^2 + 11 = (y - \sqrt{-11})(y + \sqrt{-11}).$$

This factorization puts us in

$$\mathbb{Z}[\sqrt{-11}\,] = \{a + b\sqrt{-11} \colon a, b \in \mathbb{Z}\}.$$

Since we don't want to fall into the same trap again, let's determine the units before going further. The norm function

$$N(a + b\sqrt{-11}) = a^2 + 11b^2$$

satisfies the same properties as the Gaussian norm (in fact, it's the same function restricted to the subset $\mathbb{Z}[\sqrt{-11}\,] = \mathbb{Z}[11i] \subseteq \mathbb{Z}[i]$!). So to find units, we need to find elements of norm ± 1, which leads us to the equation

$$a^2 + 11b^2 = \pm 1.$$

It's easy to see that the only solution has $b = 0$ and therefore $a = \pm 1$. So the only units in $\mathbb{Z}[\sqrt{-11}\,]$ are ± 1, and we're safe! The units are cubes.

So now assuming that $y \pm \sqrt{-11}$ are coprime, we deduce as before that

$$y + \sqrt{-11} = (a + b\sqrt{-11})^3 = (a^3 - 33ab^2) + (3a^2b - 11b^3)\sqrt{-11}.$$

(We've "absorbed" the unit into the cube, which is OK here.) Equating coefficients of $\sqrt{-11}$, we find that

$$(3a^2 - 11b^2)b = 1$$

so

$$b = \pm 1 \quad \text{and} \quad 3a^2 - 11b^2 = \pm 1.$$

The latter equation leads to $3a^2 - 11 = \pm 1$. The only solutions to this are $a = \pm 2$ (and therefore $b = 1$). Thus,

$$y = a^3 - 33ab^2 = \pm 58$$

and hence

$$x^3 = y^2 + 11 = 3375 = 15^3.$$

So it appears we have proved:

Theorem (?): The only integer solutions to $y^2 = x^3 - 11$ are $(x, y) = (15, \pm 58)$.

Okay—I still have to show you why $y \pm \sqrt{-11}$ are coprime. But believe me—they are. The proof is similar to the one I gave for $y \pm \sqrt{2}$. I will omit the details.

Exercise 28.4 Inspect the previous argument and try to identify any errors.

Alas, there must be an error, since we've missed the "obvious" solution $(x, y) = (3, \pm 4)$ to $y^2 = x^3 - 11$.

The error this time is more subtle. It has to do with the fact that $\mathbb{Z}[\sqrt{-11}\,]$, unlike its friends $\mathbb{Z}[\sqrt{2}\,]$ and $\mathbb{Z}[i]$, is not so well-behaved. There is no "Fundamental Theorem of Arithmetic" that holds in this number system: we do not have have unique factorization into primes. Indeed, the solution we missed gives us

$$(4 - \sqrt{-11})(4 + \sqrt{-11}) = 3^3.$$

It can be shown that $4 \pm \sqrt{-11}$ and 3 are primes[19] in $\mathbb{Z}[\sqrt{-11}]$. So we have discovered two fundamentally different factorizations of 27 in $\mathbb{Z}[\sqrt{-11}]$! Weird.

Lessons Learned?

The moral here is that we should not be too fast and loose when it comes to carrying out $\mathbb{Z}$-style manipulations in larger number systems, since there can be some severe differences. However, not all hope is lost. If you're interested in learning more, you should take a course in algebraic number theory, like PMATH 441!

[19] Actually, the correct word here is *irreducible*. In algebra, an element in a ring is said to be **irreducible** if it can't be factored non-trivially. A **prime** is an element that satisfies Euclid's Lemma. In nice rings, like $\mathbb{Z}$ and $\mathbb{Z}[i]$ (and $\mathbb{Z}[\sqrt{2}\,]$), these notions coincide. But in others, like $\mathbb{Z}[\sqrt{-11}\,]$, there are irreducible elements that aren't prime.

Lecture 28 Problems

28.1. At several points in this lecture, we arrived at an equation of the form

$$a + b\sqrt{k} = c + d\sqrt{k}$$

and then we seemingly deduced that we must have $a = c$ and $b = d$.

Assuming that $a, b, c, d, k \in \mathbb{Z}$ and that k is not a perfect square, give a proof that this is a valid deduction.

28.2. Prove Fermat's claim about the equation $y^2 = x^3 - 2$ by using the factorization

$$(y - \sqrt{-2})(y + \sqrt{-2}) = x^3.$$

(That is, prove that the only integer solutions are $(x, y) = (3, \pm 5)$.) Clearly state any assumptions concerning $\mathbb{Z}[\sqrt{-2}\,]$ that you make.

28.3. Consider the equation $y^2 = x^3 - 9$ with $x, y \in \mathbb{Z}$.

(a) Prove that y is even.

(b) Prove that $3 \nmid y$.

(c) Prove that $y - 3i$ and $y + 3i$ are coprime in $\mathbb{Z}[i]$.

(d) Use parts (a)–(c) and the factorization $(y+3i)(y-3i) = x^3$ to find all integer solutions to the equation $y^2 = x^3 - 9$.

Lecture 29 The Pell Equation

A person solving this problem within a year is a mathematician.

– Brahmagupta

In trying to solve the Diophantine equation $y^2 = x^3 + 2$, we ran into the equation

$$a^2 - 2b^2 = 1.$$

This was in connection to the existence of units in $\mathbb{Z}[\sqrt{2}\,]$. In general, the equation

$$a^2 - Db^2 = 1,$$

where D is not a perfect square, is called the **Pell equation**. The connection to units in $\mathbb{Z}[\sqrt{D}\,]$ is a relatively modern development. This equation has a rich history, having been studied by the Indian mathematicians Brahmagupta and Bhaskara in the 7th and 12th centuries, respectively. Brahmagupta's quote above is in reference to the equation

$$x^2 - 92y^2 = 1$$

where the smallest solution in positive integers is given by $(x, y) = (1151, 120)$.

In Europe, Fermat was the first mathematician to take interest in the equation. He challenged his contemporaries to solve the equations

$$x^2 - 61y^2 = 1 \quad \text{and} \quad x^2 - 109y^2 = 1,$$

adding that he chose small numbers so they wouldn't have to work too hard. Fermat was being a troll. The smallest positive solutions to these two equations are

$$(x, y) = (1766319049, 226153980)$$

and

$$(x, y) = (158070671986249, 15140424455100),$$

respectively. Of all Pell equations $x^2 - Dy^2 = 1$ with small D, these are the first two whose smallest solutions are enormous. Fermat must have known this. Incidentally, Bhaskara had already considered and solved $x^2 - 61y^2 = 1$ a few centuries prior, but Fermat didn't know about this.

So why do we call it the Pell equation? Euler is to blame. He accidentally attributed the equation to the English mathematician John Pell (who was only reporting on the work of William Brouncker), and the name has stuck. This is probably for the best, because "Pell equation" is easier to say than "Brahmagupta–Bhaskara–Brouncker–Fermat equation".

The Solution Set of the Pell Equation

For $D \in \mathbb{Z}$, let

$$S_D = \{(x, y) \in \mathbb{Z}^2 \colon x^2 - Dy^2 = 1.\}$$

and

$$S_D^+ = \{(x, y) \in S_D \colon x > 0 \text{ and } y > 0\}.$$

These are the sets of integer (resp. positive integer) solutions to the equation $x^2 - Dy^2 = 1$. The nature of these solution sets depends drastically on whether D is positive or negative.

Theorem 29.1 Let S_D be as above. Then:

(a) If $D < -1$, then $S_D = \{(\pm 1, 0)\}$ and $S_D^+ = \varnothing$.

(b) If $D = -1$, then $S_D = \{(\pm 1, 0), (0, \pm 1)\}$ and $S_D^+ = \varnothing$.

(c) If $D > 1$ is not a perfect square, then S_D and S_D^+ are both infinite sets.

Proof: Parts (a) is easy: If $D < -1$ then $x^2 - Dy^2$ will be > 1 if $y \neq 0$, and if $y = 0$ then we must have $x = \pm 1$. Part (b) can be proved similarly. Part (c) is considerably more difficult and won't be proved here (but see below). ■

Exercise 29.2 Determine S_D and S_D^+ if $D = d^2$ is a perfect square.

The most interesting thing about S_D is that it has a group structure: we can use known solutions to generate new solutions. To see how this works, we first introduce the **norm** function on $\mathbb{Z}[\sqrt{D}\,]$:

$$N(a + b\sqrt{D}) = (a + b\sqrt{D})(a - b\sqrt{D}) = a^2 - Db^2.$$

The next lemma follows immediately from this definition.

Lemma 29.3 The solution set S_D corresponds to the elements in $\mathbb{Z}[\sqrt{D}\,]$ of norm 1. More precisely,

$$S_D = \{(a, b) \in \mathbb{Z}^2 \colon N(a + b\sqrt{D}) = 1\}.$$

Now, it can be easily checked that N is multiplicative, in the sense that $N(\alpha\beta) = N(\alpha)N(\beta)$ for $\alpha, \beta \in \mathbb{Z}[\sqrt{D}\,]$. Thus, if α and β have norm 1, then so does their product $\alpha\beta$. Using this, we can define a group operation $*$ on S_D by

$$(a, b) * (u, v) = (au + Dbv, av + bu).$$

This definition comes from

$$(a + b\sqrt{D})(u + v\sqrt{D}) = (au + Dbv) + (av + bu)\sqrt{D}.$$

Exercise 29.4 **(The Group S_D)**

Verify that $*$ as defined above satisfies the group axioms:

(a) If $(a, b), (u, v) \in S_D$, prove that $(a, b) * (u, v) \in S_D$.

(b) Prove that $*$ is commutative and associative.

(c) Prove that $(a, b) * (1, 0) = (a, b)$ for all $(a, b) \in S_D$. Thus, the identity element is given by the trivial solution $(1, 0) \in S_D$.

(d) If $(a, b) \in S_D$, determine the inverse $(a, b)^{-1}$. [**Hint:** What is $(a + b\sqrt{D})^{-1}$?]

So, given a non-trivial solution $(a,b) \in S_D$, we can attempt to generate new solutions by taking powers of (a,b):

$$(a,b)^2 = (a,b) * (a,b), \qquad (a,b)^3 = (a,b)^2 * (a,b), \qquad \dots$$

Let's see how this works in an example.

Example 29.5

The Pell equation $x^2 - 2y^2 = 1$ has the solution $(3,2)$. We will want to repeatedly multiply by $(3,2)$ using the formula

$$(a,b) * (3,2) = (3a + 2(2b), 3a + 2b) = (3a + 4b, 2a + 3b).$$

Using this, we obtain the following solutions:

$$\begin{aligned}
(a,b)^2 &= (3,2) * (3,2) = (3(3) + 4(2), 2(3) + 3(2)) = (17,12) \\
(a,b)^3 &= (17,12) * (3,2) = (3(17) + 4(12), 2(17) + 3(12)) = (99,70) \\
(a,b)^4 &= (99,70) * (3,2) = (3(99) + 4(70), 2(99) + 3(70)) = (577,408) \\
(a,b)^5 &= (577,408) * (3,2) = (3(577) + 4(408), 2(577) + 3(408)) = (3363,2378).
\end{aligned}$$

Amazingly, the process in the preceding example generates *all* solutions to $x^2 - 2y^2 = 1$ in the positive integers! This is a special case of the following general theorem.

Theorem 29.6

Assume that $D \in \mathbb{Z}_{>0}$ is not a perfect square. Let $(a,b) \in S_D^+$ be a positive solution with smallest possible a. Then

$$S_D^+ = \{(a,b)^n : n \in \mathbb{Z}_{>0}\}.$$

The positive solution $(a,b) \in S_D^+$ with minimal a is called the **fundamental solution** to the Pell equation $x^2 - Dy^2 = 1$. The preceding theorem asserts that the fundamental solution generates all other positive solutions. In the next example, we will show how this works in the case $D = 2$. The general proof will follow the same steps.

Example 29.7 **(Solutions to $x^2 - 2y^2 = 1$)**

Show that $(3,2)$ is the fundamental solution to $x^2 - 2y^2 = 1$ and verify that

$$S_2^+ = \{(3,2)^n : n \in \mathbb{Z}_{>0}\}.$$

Proof: I'll let you check that the positive solution (a,b) with minimal a occurs when $a = 3$. Thus, $(3,2)$ is the fundamental solution.

Now, we know that $(3,2)^n$ will be a solution for all $n \geq 1$, and it's also obvious that it will be a positive solution since computing $(3,2)^n$ involves adding and multiplying positive integers. Thus, all that remains is to prove that every positive solution is equal to $(3,2)^n$ for some $n \geq 1$.

Here is the key idea. Given a positive solution (x,y) with $x > 3$, multiply it by $(3,2)^{-1}$ to get the solution (x',y'). I claim (and will prove below) that (x',y') is positive and that

$x' < x$. If $x' \leq 3$ then $(x', y') = (3, 2)$ by minimality; if $x' > 3$, then we can multiply (x', y') by $(3, 2)^{-1}$ to get a smaller positive solution (x'', y'') with $x'' < x'$. Continuing this way (i.e., by repeatedly multiplying by $(3, 2)^{-1}$), we must eventually arrive at $(3, 2)$ after finitely many steps, since otherwise we would have an infinite decreasing sequence

$$x > x' > x'' > \cdots$$

of positive integers, which is absurd. Thus, we must have that

$$(x, y) * \underbrace{(3, 2)^{-1} * \cdots * (3, 2)^{-1}}_{k \text{ times}} = (3, 2)$$

for some $k \geq 1$, and therefore

$$(x, y) = (3, 2)^{k+1},$$

as desired.

So here is what must be proved. Given $(x, y) \in S_D^+$ with $x > 3$, let

$$\begin{aligned}(u, v) &= (x, y) * (3, 2)^{-1} \\ &= (x, y) * (3, -2) \\ &= (3x - 4y, -2x + 3y).\end{aligned}$$

Then I claim:

(i) $u > 0$ and $v > 0$.

(ii) $u < x$.

To prove (i), note that $x^2 = 1 + 2y^2 > 2y^2$ so $x > \sqrt{2}y$ (since x and y are positive). Thus,

$$u = 3x - 4y > (3\sqrt{2} - 4)y > 0.$$

Next, note that

$$\begin{aligned}v > 0 &\iff -2x + 3y > 0 \\ &\iff 9y^2 > 4x^2 \\ &\iff 9\frac{x^2 - 1}{2} > 4x^2 \\ &\iff x^2 > 3 \\ &\iff x > 3.\end{aligned}$$

So $v > 0$ since we are assuming that $x > 3$.

Finally, to prove (ii), note that

$$(x, y) = (u, v) * (3, 2) = (3u + 4v, 2u + 3v)$$

so that

$$x = 3u + 4v > 3u > u,$$

as desired.

Exercise 29.8 Show that the only positive integer solution (a, b) to $x^2 - 2y^2 = 1$ with $a \leq 3$ is $(a, b) = (3, 2)$.

The proof technique we went through in the previous example is known as an **infinite descent**. The general idea is to create a process for going from a certain type of object of given "size" to another object of the same type of smaller "size". If size is always a positive integer, this process (the "descent") must eventually terminate since we cannot have an infinite decreasing sequence of positive integers.

Proof of Theorem 29.6 (Sketch): We will follow the descent procedure explained in Example 29.7. Let (a, b) be the fundamental solution. Given a positive solution (x, y) with $x > a$, consider

$$(u, v) = (x, y)(a, b)^{-1} = (x, y)(a, -b) = (xa - Dby, -bx + ay).$$

As in Example 29.7, we can show that $(u, v) \in S_D^+$ and that $u < x$. Thus, after finitely many iterations of this process, we must arrive at a solution with (u', v') with $u' \leq a$ and hence $(u', v') = (a, b)$ by minimality. That is, there exists a $k \geq 1$ such that

$$(x, y)(a, b)^{-k} = (a, b) \quad \iff \quad (x, y) = (a, b)^{k+1},$$

as desired. ■

There are now two questions that must be addressed:

1. How do we find a fundamental solution?
2. After having determined the positive solution set S_D^+, how do we determine the full solution set S_D?

The first question is a bit tricky. The obvious brute-force method (try $x = 1, 2, \ldots$ until $(x^2 - 1)/D$ is a perfect square) is not very efficient since fundamental solutions tend to be rather large, like we saw for $x^2 - 61y^2 = 1$ and $x^2 - 109y^2 = 1$. We will discuss a better approach next lecture.

On the other hand, the second question has an easy answer.

Proposition 29.9 Assume $D \in \mathbb{Z}_{>0}$ is not a perfect square and let $(a, b) \in S_D^+$ be the fundamental solution. Then

$$S_D = \{\pm(a, b)^n : n \in \mathbb{Z}\}.$$

Note: By convention, $(a, b)^0 = (1, 0)$ is the identity element in the group S_D and negative powers are given by $(a, b)^{-m} = ((a, b)^{-1})^m = (a, -b)^m$.

Exercise 29.10 Prove Proposition 29.9. [**Hint:** If $(x, y) \in S_D$ is not $(\pm 1, 0)$, then one of (x, y), $(-x, y)$, $(x, -y)$ or $(-x, -y)$ will be in S_D^+.]

Lecture 29 Problems

29.1. Find the fundamental solution to $x^2 - Dy^2 = 1$ for all non-square $D \leq 10$ by trial and error. [**Hint:** You don't have to try every $x = 1, 2, \ldots$. The equation $x^2 = 1 + Dy^2$ gives a condition on $x \bmod D$.]

29.2. Find five solutions to $x^2 - 5y^2 = 1$ with $x, y \in \mathbb{Z}_{>0}$.

29.3. Assume $D > 0$ is non-square and let (a, b) be the fundamental solution to $x^2 - Dy^2 = 1$.

(a) Suppose (u, v) is a solution to $x^2 - Dy^2 = m$. Devise a method for generating new solutions to $x^2 - Dy^2 = m$.

(b) Prove that there are infinitely many integer solutions to $x^2 - 2y^2 = 14$.

29.4. Prove that there are no integer solutions to $x^2 - 3y^2 = 14$.

29.5. Prove that if the **negative Pell equation** $x^2 - Dy^2 = -1$ has integer solutions then $v_p(D) = 0$ for all primes $p \equiv 3 \pmod 4$. [**Note:** The converse is false. See Problem 30.3.]

Lecture 30 Continued Fractions

The previous lecture ended with a lingering question: How do we find the fundamental solution to the Pell equation $x^2 - Dy^2 = 1$? This lecture will explain how.

The main idea is that a solution (x, y) with $y \neq 0$ provides a rational number x/y that is a good approximation to $\sqrt{D}$ since

$$x^2 - Dy^2 = 1 \quad \iff \quad \left(\frac{x}{y}\right)^2 - D = \frac{1}{y^2}.$$

If y is large, then $1/y^2 \approx 0$, and therefore

$$\frac{x}{y} \approx \sqrt{D}.$$

A little more precisely, we have

$$\left|\frac{x}{y} - \sqrt{D}\right| = \frac{|x - y\sqrt{D}|}{|y|} = \frac{|x^2 - Dy^2|}{|y||x + y\sqrt{D}|} \leq \frac{1}{2y^2\sqrt{D}}. \tag{$*$}$$

So, as a first step to solving $x^2 - Dy^2 = 1$, we are going to explain how to find good rational approximations to $\sqrt{D}$ via *continued fractions*.

Note: There's a lot that can be said about continued fractions but we will confine ourselves to the bits relevant to the Pell equation; there will also be very few proofs (since they tend to be a bit too technical).

Example 30.1 Suppose we want to approximate $\sqrt{2}$ with a rational number. We begin by observing that

$$\sqrt{2} = 1 + 0.4142\cdots.$$

We can re-write this as

$$\sqrt{2} = 1 + \frac{1}{\frac{1}{0.4142\cdots}}.$$

Since

$$\frac{1}{0.4142\cdots} = 2.4142\cdots = 2 + 0.4142\cdots \tag{$**$}$$

we have

$$\begin{aligned}\sqrt{2} &= 1 + \frac{1}{2 + 0.4142\cdots} \\ &= 1 + \frac{1}{2 + \frac{1}{\frac{1}{0.4142\cdots}}}.\end{aligned}$$

The bottom-most fraction is the one that appeared in ($**$). So we can repeat this process over and over, leaving us with

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}.$$

This is the so-called continued fraction expansion of $\sqrt{2}$. Let's pause to introduce some terminology.

Definition 30.2

Continued Fraction, Linear Form

An expression of the form

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}$$

where $a_0 \in \mathbb{Z}_{\geq 0}$ and $a_i \in \mathbb{Z}_{>0}$ for $i \geq 1$ is called a **continued fraction**. It is typically denoted in **linear form** by $[a_0; a_1, a_2, \ldots]$.

In the preceding example, we found that the continued fraction expansion of $\sqrt{2}$ is $[1; 2, 2, \ldots]$. We will write this as $[1; \overline{2}]$, with the overline indicating that 2 is repeating. Before explaining the connection to rational approximations, let's look at one more example.

Example 30.3

Let's find the continued fraction expansion of $\sqrt{3}$. We have

$$\sqrt{3} = 1 + 0.7320\cdots = 1 + \frac{1}{\frac{1}{0.7320\cdots}}.$$

Now,

$$\frac{1}{0.7320\cdots} = 1.3660\cdots = 1 + 0.3660\cdots = 1 + \frac{1}{\frac{1}{0.3660\cdots}}.$$

So

$$\sqrt{3} = 1 + \cfrac{1}{1 + \cfrac{1}{\cfrac{1}{0.3660\cdots}}}.$$

We continue:

$$\frac{1}{0.3660\cdots} = 2 + 0.7320\cdots = 2 + \frac{1}{\frac{1}{0.7320\cdots}}.$$

So

$$\sqrt{3} = 1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{0.7320\cdots}}}.$$

Next,

$$\frac{1}{0.7320\cdots} = 1 + 0.3660\cdots = 1 + \frac{1}{\frac{1}{0.3660\cdots}}.$$

So

$$\sqrt{3} = 1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{0.3660\cdots}}}}.$$

Now our calculations keep cycling and we find that

$$\sqrt{3} = 1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}}$$

So the continued fraction expansion of $\sqrt{3}$ is $[1; \overline{1, 2}]$.

Exercise 30.4 Find the continued fraction expansions of $\sqrt{5}$ and $\sqrt{6}$.

In general, to find the continued fraction of $\alpha = \sqrt{D}$, we can proceed as follows. Set $\alpha_0 = \alpha$ and then recursively define

$$a_n = \lfloor \alpha_n \rfloor \quad \text{and} \quad \alpha_{n+1} = \frac{1}{\alpha_n - a_n} \quad \text{for } n \geq 0.$$

Then

$$\sqrt{D} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots}}}.$$

Let's now explain the connection between continued fraction expansions and rational approximations.

Example 30.5 We can truncate

$$\sqrt{2} = 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{\ddots}}}}$$

to obtain the following rational approximations to $\sqrt{2}$:

$$\sqrt{2} \approx 1$$

$$\sqrt{2} \approx 1 + \frac{1}{2} = \frac{3}{2}$$

$$\sqrt{2} \approx 1 + \cfrac{1}{2 + \cfrac{1}{2}} = \frac{7}{5}$$

$$\sqrt{2} \approx 1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2}}} = \frac{17}{12}$$

In decimal form, the last one is $1.41\overline{6}$, which is rather close to $\sqrt{2} = 1.414\cdots$. If we go deeper into the continued fraction, we get better and better approximations.

Definition 30.6
Convergent

The nth **convergent** of the continued fraction $[a_0; a_1, a_2, \ldots]$ is the rational number

$$\frac{p_n}{q_n} = [a_0; a_1, \ldots, a_n],$$

where p_n and q_n are given in lowest terms (i.e. $\gcd(p_n, q_n) = 1$).

So in the preceding example we computed the first several convergents of the continued fraction $[1; \overline{2}]$.

The first few convergents of $[a_0; a_1, a_2, \ldots]$ are given by

$$\begin{aligned}\frac{p_0}{q_0} &= \frac{a_0}{1}\\ \frac{p_1}{q_1} &= \frac{a_0a_1 + 1}{a_1}\\ \frac{p_2}{q_2} &= \frac{a_0a_1a_2 + a_0 + a_2}{a_1a_2 + 1}.\end{aligned}$$

(You should verify this!) In general, p_n and q_n obey a recursive relationship. We record this relationship, along with several other properties, in the next result.

Proposition 30.7 **(Properties of Convergents)**

Let $a_0, a_1, \ldots \in \mathbb{Z}_{>0}$. Set $p_{-2} = 0, q_{-2} = 1, p_{-1} = 1, q_{-1} = 0$ and define

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

for $n \geq 0$. Then:

(a) $[a_0; a_1, \ldots, a_n] = \dfrac{p_n}{q_n}$ for $n \geq 0$.

(b) $p_{n+1}q_n - p_nq_{n+1} = (-1)^n$ for all $n \geq -2$. (In particular, $\gcd(p_n, q_n) = 1$ for all $n \geq 0$. Thus, $\dfrac{p_n}{q_n}$ is the nth convergent of $[a_0; a_1, a_2, \ldots]$.)

(c) If $[a_0; a_1, a_2, \ldots]$ is the continued fraction expansion of α, then

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_nq_{n+2}} \leq \frac{1}{q_n^2}$$

for all $n \geq 0$. Furthermore,

$$\frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_4}{q_4} < \cdots < \alpha < \cdots < \frac{p_5}{q_5} < \frac{p_3}{q_3} < \frac{p_1}{q_1}.$$

(In particular, $\dfrac{p_n}{q_n}$ converges to α as $n \to \infty$.)

Proof: Parts (a) and (b) can be proved by induction. For part (c), we have

$$\alpha = [a_0; a_1, \ldots, a_n, \alpha_{n+1}],$$

and so

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

In particular, $\alpha > p_n/q_n$ if n is even, and $\alpha < p_n/q_n$ if n is odd. Further, since $\alpha_{n+1} > a_{n+1}$, we see that $\alpha_{n+1}q_n + q_{n-1} > a_{n+1}q_n + q_{n-1} = q_{n+1}$, and therefore

$$\left|\alpha - \frac{p_n}{q_n}\right| < \frac{1}{q_n q_{n+1}} \leq \frac{1}{q_n^2}.$$

Finally,

$$\frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} = \frac{a_{n+2}(p_{n+1}q_n - p_n q_{n+1})}{q_n q_{n+2}} = \frac{a_{n+2}(-1)^n}{q_n q_{n+2}}.$$

So if n is even, the above quantity is positive; while if n is odd, it's negative. ■

Part (c) of the Proposition 30.7 allows us to equate α with its continued fraction expansion $[a_0; a_1, \ldots]$, with the understanding that the continued fraction expansion is the limit of the sequence of convergents. It also tells us that the convergents of the continued fraction expansion of α are good rational approximations of α (remember—our goal is to find good rational approximations to $\sqrt{D}$).

The next example illustrates how we can use the recursion given in Proposition 30.7 to calculate the convergents of a given continued fraction.

Example 30.8 Let's work with $\sqrt{2} = [1, \overline{2}] = [1, 2, 2, 2, \ldots]$. We begin by creating the table

		1	2	2	2	⋯
0	1					
1	0					

Our goal is to populate this table with the convergents:

		1	2	2	2	⋯
0	1	p_0	p_1	p_2	p_3	⋯
1	0	q_0	q_1	q_2	q_3	⋯

To find p_0, we take the entry 1 at the top of its column then multiply it by the entry to the left of p_0 and add to it the entry to the left of that:

$$p_0 = 1 \cdot 1 + 0 = 1.$$

Likewise, to get q_0, we multiply 1 by the entry to the left of q_0 and then add the entry to the left of that:

$$q_0 = 1 \cdot 0 + 1 = 1.$$

We now have

		1	2	2	2	⋯
0	1	1	p_1	p_2	p_3	⋯
1	0	1	q_1	q_2	q_3	⋯

We proceed in a similar manner to fill up the second column. The entry p_1 is $2\times 1+1$, and the entry q_1 is $2\times 1+0$. Continuing on to the next columns, we end up

		1	2	2	2	$\cdots$
0	1	1	3	7	17	$\cdots$
1	0	1	2	5	12	$\cdots$

So the first few convergents of $\sqrt{2}=[1;\overline{2}]$ are

$$1,\ \frac{3}{2},\ \frac{7}{5},\ \frac{17}{12},$$

exactly as we had found in Example 30.5.

Exercise 30.9 Find the first five convergents of $\sqrt{3}$.

Here is our main result.

Theorem 30.10 **(Fundamental Solution of the Pell Equation—First Version)**

Assume $D\in\mathbb{Z}_{>0}$ is not a perfect square and let p_n/q_n be the nth convergent of the continued fraction expansion of $\sqrt{D}$. The fundamental solution of the Pell equation $x^2-Dy^2=1$ is given by $(x,y)=(p_m,q_m)$ where $m\geq 0$ is the smallest index such that $p_m^2-Dq_m^2=1$.

Proof (Sketch): The first key point is that any solution (x,y) with $y\geq 1$ satisfies

$$\left|\frac{x}{y}-\sqrt{D}\right|<\frac{1}{2y^2}.$$

(This follows from $(*)$.) A non-trivial theorem of Lagrange asserts that any $x/y\in\mathbb{Q}_{>0}$ with $\gcd(x,y)=1$ that satisfies this inequality must be a convergent $\sqrt{D}$. In particular, the fundamental solution arises from a convergent.

Next, using the recurrence relation for p_n and q_n (and some work), one can show that

$$p_n^2-Dq_n^2=(-1)^{n+1}k_n$$

for some integer k_n. It can be shown that $k_n\neq -1$ and that there is an $r\in\mathbb{Z}_{>0}$ such that $k_n=1$ if and only of $r\mid n$. Hence all positive solutions to $x^2-Dy^2=1$ must be of the form (p_n,q_n) where n is a multiple of r. So there must be a smallest index m that gives a solution, and this solution will have minimal p_m, hence will be the fundamental solution. ■

So to find the fundamental solution to $x^2-Dy^2=1$, we should first find the continued fraction expansion $[a_0;a_1,\ldots]$ of $\sqrt{D}$ and then make our table of convergents as explained in Example 30.8 but now with an additional row containing $p_n-Dq_n^2$:

		a_0	a_1	a_2	a_3	$\cdots$
0	1	p_0	p_1	p_2	p_3	$\cdots$
1	0	q_0	q_1	q_2	q_3	$\cdots$
		$p_1^2-Dq_1^2$	$p_2^2-Dq_2^2$	$p_3^2-Dq_3^2$	$p_4^2-Dq_4^2$	$\cdots$

The first occurrence of 1 in the bottom row indicates that we have found our fundamental solution.

Example 30.11 Find the fundamental solution of $x^2 - 3y^2 = 1$.

Solution: The continued fraction expansion of $\sqrt{3}$ is $[1; \overline{1, 2}]$. So we have to fill up the table

		1	1	2	1	$\cdots$
0	1	p_0	p_1	p_2	p_3	$\cdots$
1	0	q_0	q_1	q_2	q_3	$\cdots$
		$p_0^2 - 3q_0^2$	$p_1^2 - 3q_1^2$	$p_2^2 - 3q_2^2$	$p_3^2 - 3q_3^2$	$\cdots$

until we find a 1 in the bottom row. Doing so, we end up with

		1	1	2	1	$\cdots$
0	1	1	2	p_2	p_3	$\cdots$
1	0	1	1	q_2	q_3	$\cdots$
		-2	1	$p_2^2 - 3q_2^2$	$p_3^2 - 3q_3^2$	$\cdots$

Okay, that was easy! The fundamental solution in $(x, y) = (2, 1)$.

Example 30.12 Find the fundamental solution of $x^2 - 19y^2 = 1$.

Solution: The continued fraction expansion of $\sqrt{19}$ is $[4; \overline{2, 1, 3, 1, 2, 8}]$ (found using a computer!). The table of convergents is:

		4	2	1	3	1	2	8	$\cdots$
0	1	4	9	13	48	61	170	p_6	$\cdots$
1	0	1	2	3	11	14	39	q_6	$\cdots$
		-3	5	-2	5	-3	1	$\cdots$	

Thus, the fundamental solution is $(x, y) = (170, 39)$.

Example 30.13 Find a solution to $x^2 - 61y^2 = 1$.

Solution: This is Fermat's challenge mentioned last lecture. The continued fraction expansion of $\sqrt{61}$ is

$$[7; \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}].$$

Here is the table of convergents:

		7	1	4	3	1	2	2	1	3	4	1	$\cdots$
0	1	7	8	39	125	164	453	1070	1523	5639	24079	29718	$\cdots$
1	0	1	1	5	16	21	58	137	195	722	3083	3805	$\cdots$
		-4	5	-4	9	-5	5	-9	4	-3	12	-1	$\cdots$

You'll notice that I stopped once I found a -1 in the bottom row. This is because once we have a solution (p, q) to $x^2 - Dy^2 = -1$ then we can "square it" to find a solution to

$x^2 - Dy^2 = 1$. To be more precise, if

$$N(p + q\sqrt{D}) = p^2 - Dq^2 = -1$$

then

$$N((p + q\sqrt{D})^2) = (-1)^2 = 1.$$

Now we compute

$$(p + q\sqrt{D})^2 = (p^2 + Dq^2) + (2pq)\sqrt{D}$$

to conclude that $(p^2 + Dq^2, 2pq)$ is a solution to $x^2 - Dy^2 = 1$. (Refer back to Lemma 29.3.)

In our case, the solution $(29718, 3805)$ to $x^2 - 61y^2 = -1$ yields the solution

$$(1766319049, 226153980)$$

to $x^2 - 61y^2 = 1$. In fact, this is the fundamental solution. It is a theorem that if p/q is the first convergent such that $p^2 - Dq^2 = -1$ then $(p^2 + Dq^2, 2pq)$ is the fundamental solution to $x^2 - Dy^2 = 1$. In our case we can verify this by computing the remaining convergents p_n/q_n in the table above and we'll find that the solution above occurs when $n = 21$.

We can improve Theorem 30.10 by investigating the continued fraction expansion of $\sqrt{D}$ a little more closely. We state the next two results without proof.

Proposition 30.14 **(Continued Fraction Expansion of $\sqrt{D}$)**

Assume $D \in \mathbb{Z}_{>0}$ is not a perfect square. Then the continue fraction expansion of $\sqrt{D}$ periodic:

$$\sqrt{D} = [a_0; \overline{a_1, \ldots, a_\ell}].$$

Assume ℓ is the length of the smallest period. Then:

(a) $a_\ell = 2a_0$.

(b) $a_1, \ldots, a_{\ell-1}$ is a palindromic sequence: $a_1 = a_{\ell-1}$, $a_2 = a_{\ell-2}$, ...

For example, the continued fraction expansions

$$\sqrt{19} = [4; \overline{2, 1, 3, 1, 2, 8}] \quad \text{and} \quad \sqrt{61} = [7; \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}]$$

illustrate this theorem.

Theorem 30.15 **(Fundamental Solution of the Pell Equation—Second Version)**

Assume $D \in \mathbb{Z}_{>0}$ is not a perfect square, and let p_n/q_n and ℓ be, respectively, the nth convergent and the period of the continued fraction expansion of $\sqrt{D}$. The fundamental solution of the Pell equation $x^2 - Dy^2 = 1$ is given by

$$(x, y) = \begin{cases} (p_{\ell-1}, q_{\ell-1}) & \text{if } \ell \text{ is even} \\ (p_{2\ell-1}, q_{2\ell-1}) & \text{if } \ell \text{ is odd.} \end{cases}$$

Furthermore, all positive integer solutions are given by

$$(x, y) = \begin{cases} (p_{n\ell-1}, q_{n\ell-1}) & \text{if } \ell \text{ is even} \\ (p_{2n\ell-1}, q_{2n\ell-1}) & \text{if } \ell \text{ is odd,} \end{cases}$$

where $n = 1, 2, \ldots$.

For instance, we saw above that:

- The fundamental solution of $x^2 - 19y^2 = 1$ is $(x, y) = (p_5, q_5)$, which agrees with the fact that the period of $\sqrt{9}$ is $\ell = 6$.
- The fundamental solution of $x^2 - 61y^2 = 1$ is $(x, y) = (p_{21}, q_{21})$, which agrees with the fact that the period of $\sqrt{61}$ is $\ell = 11$.

Here are the fundamental solutions to $x^2 - Dy^2 = 1$ for all non-square $D \leq 15$.

D	$\sqrt{D}$	Fundamental Solution
2	$[1, \overline{2}]$	$(3, 2)$
3	$[1, \overline{1, 2}]$	$(2, 1)$
5	$[2, \overline{4}]$	$(9, 4)$
6	$[2, \overline{2, 4}]$	$(5, 2)$
7	$[2, \overline{1, 1, 1, 4}]$	$(8, 3)$
8	$[2, \overline{1, 4}]$	$(3, 1)$
10	$[3, \overline{6}]$	$(19, 6)$
11	$[3, \overline{3, 6}]$	$(10, 3)$
12	$[3, \overline{2, 6}]$	$(7, 2)$
13	$[3, \overline{1, 1, 1, 6}]$	$(649, 180)$
14	$[3, \overline{1, 2, 1, 6}]$	$(15, 4)$
15	$[3, \overline{1, 6}]$	$(4, 1)$

Exercise 30.16 Verify some of the entries in the table above.

Lecture 30 Problems

30.1. (a) Determine the real number α whose continued fraction expansion is $[1; \overline{1}]$. [**Hint:** Show that $\alpha = 1 + \frac{1}{\alpha}$ and use this to solve for α.]

(b) Determine the real number β whose continued fraction expansion is $[1; 1, \overline{2, 6}]$.

30.2. Let $D = n^2 + 1$, where $n \geq 2$ is an integer.

(a) Find the continued fraction expansion of $\sqrt{D}$.

(b) Find the fundamental solution of $x^2 - Dy^2 = 1$.

30.3. (a) Find the fundamental solution of $x^2 - 34y^2 = 1$.

(b) Prove that $x^2 - 34y^2 = -1$ has no integer solutions by using the fact that if (p, q) were the smallest such positive solution, then $(p^2 + Dq^2, 2pq)$ would be the fundamental solution to $x^2 - 34y^2 = 1$.

Lecture 31 Fermat's Last Theorem

> *Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.*[20]
>
> – Pierre de Fermat

The story behind Fermat's Last Theorem (FLT) is famous. Fermat left a marginal note in his copy of Diophantus' *Arithmetica* claiming he had a "marvellous proof" of the fact that, for all $n > 2$, the equation

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

has no solution in the positive integers. But no proof of this fact was to be found for 300 years—despite the effort of many great (and not-so-great) mathematicians. In this lecture I will outline—*very broadly*—the strategy that led to the proof of Fermat's Last Theorem.

The mathematics used in the eventual proof was mostly developed in the 20th century and was completely out of Fermat's reach. It's very unlikely that Fermat had a proof. Perhaps the most compelling evidence of this is that he himself never mentioned, in public, that he did. The marginal note was in his personal copy of *Arithmetica*, made available to the public by his son after Fermat had died.

Fermat's way of doing mathematics was through posing problems as challenges to others. If he was in possession of a correct proof of FLT, he would have surely made more noise about it. He did actually pose the case $n = 3$ as a challenge, and we have his proof for $n = 4$ (which I will sketch below)—but that's it. He probably initially believed that his argument for the $n = 4$ case would generalize, but it doesn't.

Here is (essentially[21]) what Fermat proved. FLT for $n = 4$ follows immediately from this.

Theorem 31.1 There is no solution to the equation $x^4 + y^4 = z^2$ with $x, y, z \in \mathbb{Z}_{>0}$.

Proof (Outline): The **strategy** is to use infinite descent: Suppose there is a positive integer solution (a, b, c). Then *somehow* use this solution to construct another positive integer solution (a', b', c') with $c' < c$. Repeat this process to get smaller and smaller positive integer solutions with

$$c > c' > c'' > \cdots$$

which is absurd, since we cannot have an infinite decreasing sequence of positive integers. Thus, our original assumption—the existence of (a, b, c)—must be false.

To **execute** this strategy, we must explain how to construct (a', b', c') from (a, b, c). Here is the outline; I'll let you fill in the details as a fun exercise.

- If (a, b, c) is a positive solution then (a^2, b^2, c) is a Pythagorean triple. Argue that we may assume, wlog, that it is primitive.

[20] It is impossible to separate a cube into two cubes, or a fourth power into two fourth powers, or in general, any power higher than the second, into two like powers. I have discovered a truly marvelous proof of this, which this margin is too narrow to contain.

[21] Fermat actually proved there are no solutions to $x^4 - y^4 = z^2$, but the method of proof is the same.

- Conclude that there exist coprime positive integers $n > m$ with one odd and the other even such that

$$a^2 = n^2 - m^2, \quad b^2 = 2nm \quad \text{and} \quad c = n^2 + m^2.$$

- Observe that $a^2 + m^2 = n^2$. So (a, m, n) is also primitive Pythagorean triple. Also note that m must be even and n must be odd.
- Deduce from this that there are positive integers $u, v \in \mathbb{Z}$ such that $m = 2uv$ and $n = u^2 + v^2$. Hence $b^2 = 4uv(u^2 + v^2)$.
- Argue that u, v and $(u^2 + v^2)$ must be squares, say $u = (a')^2$, $v = (b')^2$ and $u^2 + v^2 = (c')^2$ with $a', b', c' \in \mathbb{Z}_{>0}$.
- Explain why this is our desired smaller solution. ■

Exercise 31.2 Supply the missing details in the above proof.

Having established the $n = 4$ case of FLT, we also get the cases $n = 4k$ for free since

$$x^{4k} + y^{4k} = z^{4k} \quad \Longleftrightarrow \quad (x^k)^4 + (y^k)^4 = (z^k)^4.$$

So now all that remains is to prove FLT in the case where $n = p$ is an odd prime.

Exercise 31.3 Explain why proving FLT in the cases $n = 4$ and $n = p$ an odd prime allows us to deduce FLT for all $n > 2$.

Since the 1700s, mathematicians have tried to tackle FLT one prime at a time. Euler proved $p = 3$ (with a significant gap relating to how number theory in $\mathbb{Z}[\sqrt{-3}]$ works); Dirichlet and Legendre independently proved $p = 5$; Lamé did $p = 7$, and so on, and so forth.

The proofs of these special cases were disjointed and ad hoc. There were some general approaches, most notably due to Sophie Germain and Ernst Kummer, but they all fell far short of the prize.

The Modern Approach

The first real progress occurred in the 1950s (approximately 300 years after Fermat had died), when two Japanese mathematicians, Yutaka Taniyama and Goro Shimura, made the following claim (which was further solidified with evidence by André Weil in 1967).

Conjecture 31.4 **(The Modularity Conjecture)**

Every elliptic curve over $\mathbb{Q}$ is modular.

You are not expected to understand what this means. What is an elliptic curve? What does "modular" mean? I will elaborate next lecture.

The important thing for now: *What does this have to do with FLT?* The short answer is that a positive solution to $x^p + y^p = z^p$ would contradict the Modularity Conjecture.

Indeed, in the 1980s Gerhard Frey noticed that a positive solution (a, b, c) to $x^p + y^p = z^p$ could be used to create an elliptic curve

$$E_{a,b,c}: \quad y^2 = x(x - a^p)(y + b^p)$$

that was very peculiar—specifically, it appeared that $E_{a,b,c}$ was not modular. Ken Ribet proved that Frey's elliptic curve $E_{a,b,c}$ was indeed not modular. This was an amazing achievement—not just because the proof contained many great ideas, but because it meant that now FLT was within reach: all (!) you had to do was prove the Modularity Conjecture.

And this is exactly what happened next. Legend has it that Andrew Wiles holed himself up in his Princeton attic for six years while he secretly tried to find a proof. In 1993, Wiles gave a series of three lectures in Cambridge explaining his proof, and he published a manuscript containing the details shortly thereafter.

Things took a dramatic turn, however, when a flaw was discovered in the proof. The error was not easy to fix. Wiles worked on it for a year, first alone, and then in collaboration with his student Richard Taylor, but seemingly without much success—until one day when the pieces somehow fell into place. In 1994 Taylor and Wiles submitted two manuscripts to the *Annals of Mathematics*, one of the premier journals for mathematics, containing the proof of the following theorem.

Theorem 31.5 **(Semistable Modularity Theorem)**

Every *semistable* elliptic curve over $\mathbb{Q}$ is modular.

This was good enough, because the Frey curve $E_{a,b,c}$ is "semistable". As a corollary, we get:

Theorem 31.6 **(Fermat's Last Theorem)**

For every $n > 2$, the equation

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

has no solution with $x, y, z \in \mathbb{Z}_{>0}$.

Proof: Without loss of generality we may assume that $n = p > 5$ is an odd prime.[22] Suppose (a, b, c) is a positive solution to $a^p + b^p = c^p$. Let $E_{a,b,c}$ be the associated Frey curve. Then, by Ribet, $E_{a,b,c}$ is not modular. However, by Wiles and Taylor–Wiles, $E_{a,b,c}$ must be modular. Contradiction! Thus, there is no such (a, b, c). ■

The full modularity conjecture was proved in 1999 by Breuil, Conrad, Diamond and Taylor.

Lecture 31 Problems

31.1. Prove Fermat's Last Theorem without using the Modularity Theorem.

[22] The restriction to $p > 5$ is needed to make part of the argument work. So to complete this proof, we need to quote the separate proofs for the cases $p = 3$ and $p = 5$ (and $n = 4$, of course).

Lecture 32 Elliptic Curves Over $\mathbb{Q}$

It is possible to write endlessly on elliptic curves. (This is not a threat.)

– Serge Lang

This is going to be a crash course (without proofs) on the basics of elliptic curves. If you're interested in learning more, I highly recommend the book *Rational Points on Elliptic Curves* by Silverman and Tate.

Definition 32.1

Elliptic Curve, Rational Points

The equation

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

where $a, b \in \mathbb{Q}$, is said to define an **elliptic curve** over $\mathbb{Q}$ if $4a^3 + 27b^2 \neq 0$.

Given an elliptic curve E, its set of **rational points** is

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q}^2 \colon y^2 = x^3 + ax + b\}.$$

We similarly define the sets of **integer points** $E(\mathbb{Z})$ and **real points** $E(\mathbb{R})$.

Some comments are in order:

- Elliptic curves are not ellipses. For example, here is the plot of the real points of $y^2 = x^3 + 2$.

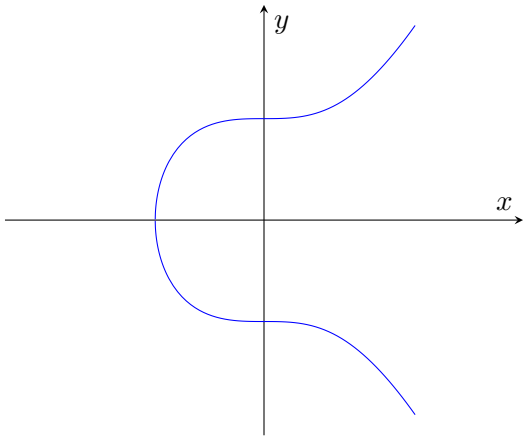


Figure 32.3: Plot of $E(\mathbb{R})$ for $E : y^2 = x^3 + 2$.

- Historically, elliptic curves arose in connection with arc length integrals for calculating lengths of segments of an ellipse. This is a long (and interesting) story. We won't go into it here.
- The condition $4a^3 + 27b^2 \neq 0$ can be equivalently rephrased as: The cubic polynomial $x^3 + ax + b$ does not have repeated roots. Geometrically, this means that the plot of $y^2 = x^3 + ax + b$ in the xy-plane will be *smooth* without cusps or nodes.

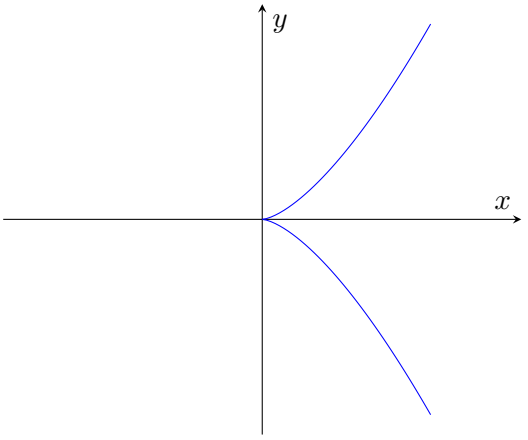


Figure 32.4: $y^2 = x^3$ (cusp)

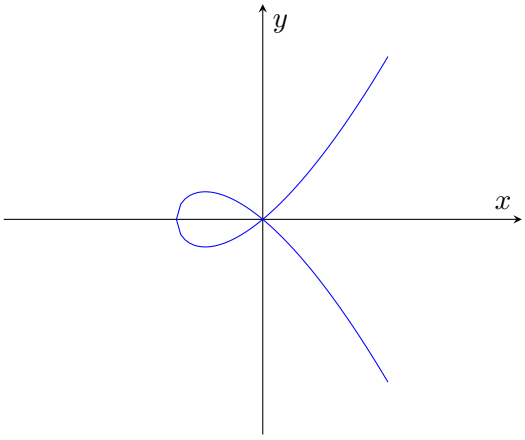


Figure 32.5: $y^2 = x^2(x+1)$ (node)

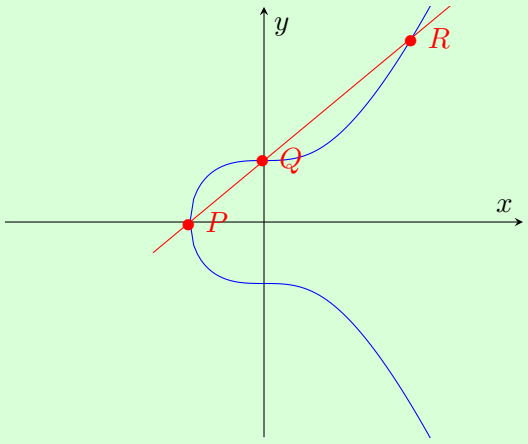
- There is a more general definition of "elliptic curve" that uses equations of the form
$$y^2 + axy + by = x^3 + cx^2 + dx + e,$$
where the coefficients a, b, c, d, e are required to satisfy a similar smoothness condition. However, Definition 32.1 will suffice for our purposes.

The equation defining an elliptic curve E is interesting because the solution set $E(\mathbb{Q})$ of rational points can be turned into a group. That is, there is a way to take two points $P, Q \in E(\mathbb{Q})$ and somehow generate a new point $P \oplus Q \in E(\mathbb{Q})$. This should remind you of the solution set S_D to the Pell equation $x^2 - Dy^2 = 1$ where a similar thing was possible. Let's look at an example.

Example 32.2 **(Group Operation on $y^2 = x^3 + 1$ — First Look)**

Consider the elliptic curve
$$E: \quad y^2 = x^3 + 1.$$
If P and Q are two distinct points in $E(\mathbb{Q})$ then the line through P and Q will intersect the curve E at a third point $R \in E(\mathbb{Q})$.

For instance, consider $P = (-1, 0)$ and $Q = (0, 1)$. The line through these two points has equation $y = x + 1$.



Let $R = (x_R, y_R)$ be the third point of intersection between the line and E. To find the x-coordinate of R, we have to solve the equation

$$(x+1)^2 = x^3 + 1$$

for x. Expanding this out, we end up with

$$x^3 - x^2 - 2x = 0.$$

Now it's easy to factor this, but let's not do that. Observe instead that the sum of the roots of this cubic adds up to (-1) times the coefficient of x^2 (see exercise below). But we already know two roots—namely, the x-coordinates of P and Q! So if x_R is the x-coordinate of R, then we have

$$-1 + 0 + x_R = 1 \quad \Longleftrightarrow \quad x_R = 2.$$

Hence, using the equation of the line $y = x + 1$, the y-coordinate of R must be

$$y_R = x_R + 1 = 3.$$

Thus, $R = (2, 3)$. Notice that $R \in E(\mathbb{Q})$.

Exercise 32.3 Suppose that the cubic equation

$$x^3 + ax^2 + bx + c = 0$$

has roots α, β, γ. Prove that $\alpha + \beta + \gamma = -a$.

Deduce that if the cubic has rational coefficients (i.e. if $a, b, c \in \mathbb{Q}$) and if two of the roots are rational, then the third root must be rational too.

The process in the previous example can be generalized. Suppose E is given by

$$y^2 = x^3 + ax + b, \quad a, b \in \mathbb{Q}.$$

Given $P, Q \in E(\mathbb{Q})$, let R be the third point of intersection with E of the line through P and Q. Suppose the line through P and Q has equation $y = mx + c$; note that both m and c must be rational because P and Q are. Then to find the points of intersection of this line with E, we substitute y into the equation for E to find that

$$(mx + c)^2 = x^3 + ax + b \quad \Longleftrightarrow \quad x^3 - m^2x^2 + (a - 2mc)x + b - c^2 = 0.$$

We know two of the roots, namely x_P and x_Q, so by the preceding exercise we must have

$$x_R + x_P + x_Q = m^2.$$

Using this, we can determine x_R and therefore $y_R = mx_R + c$.

The point R obtained by this process will be denoted by $P * Q$. (**Warning:** This $*$ is *not* a group operation; we will have to modify it. Read ahead.)

The construction of $P * Q$ raises two questions:

1. What do we do if $P = Q$? That is, what is $P * P$?
2. What if the line through P and Q is vertical? Such a line will not intersect $E(\mathbb{Q})$ in a third point.

The first question is easy to deal with. Just use the tangent line to P! This line will intersect E "twice" at P (more precisely, it will intersect with multiplicity equal to 2); we let $P * P$ be the third point of intersection of the tangent line and E.

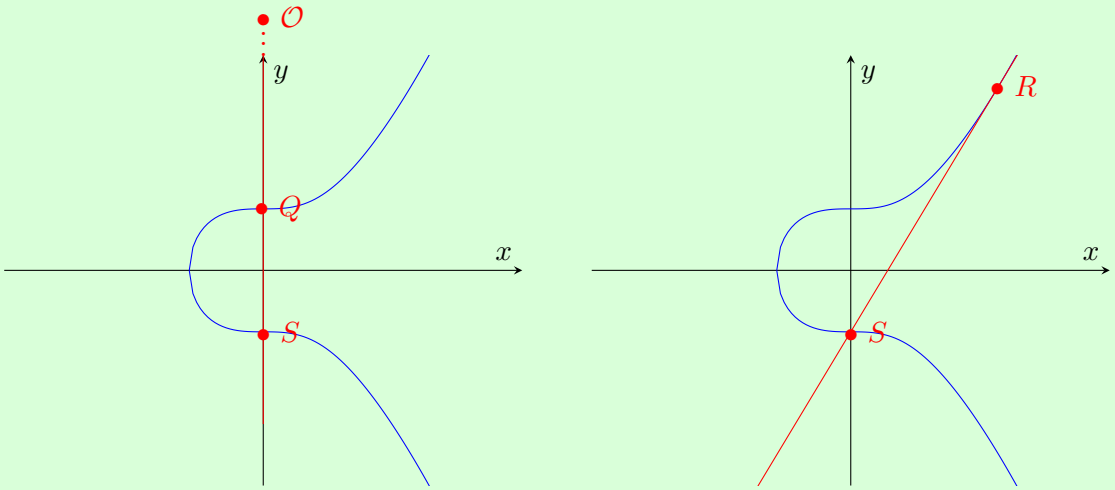
The second question is more tricky. We will weasel our way out of it by pretending that there is a "point at infinity" which we will denote by $\mathcal{O}$. We will pretend that every vertical line passes through $\mathcal{O}$. In particular, if $P, Q \in E(\mathbb{Q})$ are such that the line through them is vertical, then we define $P * Q = \mathcal{O}$.

Example 32.4 **(Group Operation on $y^2 = x^3 + 1$ — Second Look)**

Returning to

$$E: \quad y^2 = x^3 + 1,$$

let $Q = (0, 1)$ as before, and let $S = (0, -1)$. Then the line through Q and S is vertical, so we have $Q * S = \mathcal{O}$.



Suppose now we want to calculate $L = R * R$, where $R = (2, 3)$. We begin by finding the tangent line through E at R, which is $y = 2x - 1$. We plug $y = 2x - 1$ into the equation for E to find

$$(2x - 1)^2 = x^3 + 1 \quad \Longleftrightarrow \quad x^3 - 4x^2 + \cdots = 0.$$

Thus,

$$x_L + x_R + x_R = 4 \quad \Longleftrightarrow \quad x_L = 4 - 2x_R = 0$$

and consequently $y_L = 2x_R - 1 = -1$. This shows that $R * R = (0, -1) = S$.

Exercise 32.5 Determine $Q * Q$, where $Q = (0, 1)$ is as in the preceding example. Can you explain what's going on?

To define a group operation $\oplus$ on $E(\mathbb{Q})$, we need to first agree on what the identity element should be. We choose it to be $\mathcal{O}$, the point at infinity. For this to be OK, we will need to count $\mathcal{O}$ as being in $E(\mathbb{Q})$. Thus, our amended definition of $E(\mathbb{Q})$ is

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q}^2 : y^2 = x^3 + ax + b\} \cup \{\mathcal{O}\}.^{23}$$

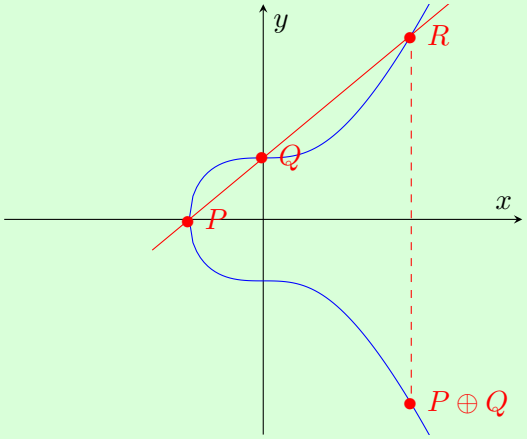
Define $\oplus$ on $E(\mathbb{Q})$ as follows.

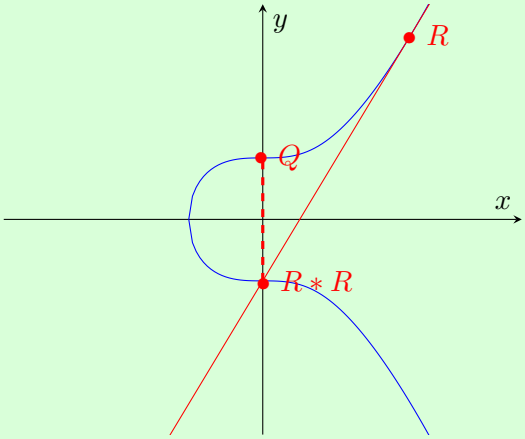
- $P \oplus \mathcal{O} = \mathcal{O} \oplus P = P$ for all $P \in E(\mathbb{Q})$.
- If $P, Q \in E(\mathbb{Q})$ are both not $\mathcal{O}$, then $P \oplus Q = (x_R, -y_R)$, where $R = (x_R, y_R) = P * Q$.

In other words, we obtain $P \oplus Q$ from $P * Q$ by reflecting across the x-axis. That is, $P \oplus Q = (P * Q) * \mathcal{O}$.

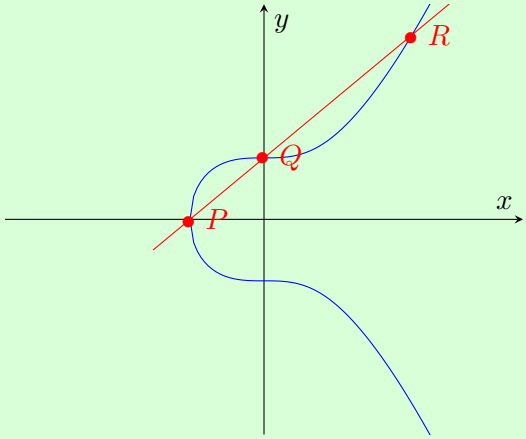
Example 32.6 **(Group Operation on $y^2 = x^3 + 1$ — Third Look)**

Let's determine $P \oplus Q$ where $P = (-1, 0)$ and $Q = (0, 1)$. We've already found that $R = P * Q = (2, 3)$. Thus, $P \oplus Q = (2, -3)$. Similarly, $R \oplus R = (R * R) * \mathcal{O} = (0, -1) * \mathcal{O} = Q$.





What about $R \oplus Q$? The line through R and Q has equation $y = x + 1$, and we can use this to find that $R * Q = P$. Since P is already lying on the x-axis, it follows that $R \oplus Q = P$ too.



[23] We view $\mathcal{O}$ as also being in $E(\mathbb{R})$ but not in $E(\mathbb{Z})$.

Exercise 32.7 Determine $Q \oplus Q$, with $Q = (0, 1)$ as in the preceding example.

To summarize:

Theorem 32.8 **($E(\mathbb{Q})$ is a Group)**

Let E be an elliptic curve over $\mathbb{Q}$. Then $\oplus$ as defined above turns $E(\mathbb{Q})$ into a commutative group. The identity element is $\mathcal{O}$ and the additive inverse of $P = (x, y)$ is $-P = (x, -y)$.

REMARK ($E(\mathbb{R})$ and $E(\mathbb{Z})$)

The same definition of $\oplus$ turns $E(\mathbb{R})$ into a group. However, $E(\mathbb{Z})$ is not a group under $\oplus$, and not just because $\mathcal{O} \notin E(\mathbb{Z})$; see Example 32.12 for an elliptic curve E with a point $P \in E(\mathbb{Z})$ such that $P \oplus P \notin E(\mathbb{Z})$.

Theorem 32.8 is tricky to prove. Specifically, the fact that $\oplus$ is associative is not so obvious. (Draw a picture to see what is involved in showing that $(P \oplus Q) \oplus R = P \oplus (Q \oplus R)$.) The other group axioms are easy to check.

Exercise 32.9 Show that if $P = (x, y) \in E(\mathbb{Q})$ then $-P = (x, -y)$ is the additive inverse of P.

Example 32.10 **(Group Operation on $y^2 = x^3 + 1$ — Final Look)**

It can be shown that if (x, y) is a rational solution to $y^2 = x^3 + 1$ then (x, y) must be one of $(-1, 0)$, $(0, \pm 1)$ or $(2, \pm 3)$. Thus,

$$E(\mathbb{Q}) = \{(-1, 0), (0, \pm 1), (2, \pm 3), \mathcal{O}\}.$$

Let's take $R = (2, 3)$ and see what happens if we calculate

$$2R = R \oplus R, \quad 3R = R \oplus R \oplus R, \quad \ldots.$$

We already saw that $2R = (0, 1)$. Our calculations in the last example also give us

$$3R = R \oplus 2R = R \oplus (0, 1) = (-1, 0).$$

I will leave the following for you to verify:

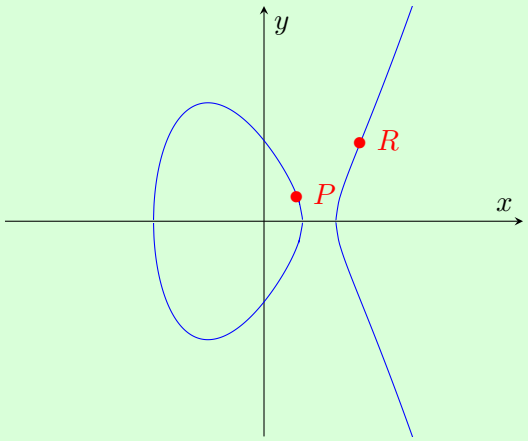
$$\begin{aligned} 4R &= (0, -1) \\ 5R &= (2, -3) \\ 6R &= \mathcal{O}. \end{aligned}$$

Thus, R *generates* the entire group $E(\mathbb{Q})$!

Exercise 32.11 Confirm the calculations of $4R$, $5R$ and $6R$ above.

The preceding example should remind you of how Pell equations have fundamental solutions that generate the entire solution set. Perhaps elliptic curves do as well? Before we get too carried away, let's look at another example.

Example 32.12 Let E be the elliptic curve $y^2 = x^3 - 9x + 9$ and consider the points $P = (1, 1)$ and $R = (3, 3)$ in $E(\mathbb{Q})$.



We have

$$\begin{aligned} 2P &= (7, 17) \\ 3P &= (8/9, 109/27) \\ 4P &= (715/289, 6731/4913) \\ 5P &= (-17195/5041, 38413/357911) \\ 6P &= (1048753/427716, -361921823/279726264). \end{aligned}$$

Continuing this way, it appears that we can generate infinitely many rational points in $E(\mathbb{Q})$ by repeatedly adding P to itself. However, notice that R is seemingly missing from this list. Indeed, it can be proved that $R \neq nP$ for all $n \in \mathbb{Z}$. So P does not generate all of $E(\mathbb{Q})$.

Turning to R, we find that

$$\begin{aligned} 2R &= (3, -3) \\ 3R &= \mathcal{O}. \end{aligned}$$

So unlike P, repeatedly adding R to itself will only produce the points R, $2R$ and $\mathcal{O}$. We have discovered that R has finite order in $E(\mathbb{Q})$—specifically, $\operatorname{ord}(R) = 3$. On the other hand, P has infinite order though we won't prove that here.

Examining E more closely, we can find three other integer points, namely $Q = (-3, 3)$, $T = (0, 3)$ and $S = (15, 57)$. Remarkably, these points can be generated using P and R:

$$P \oplus R = Q, \quad -P \oplus R = T \quad \text{and} \quad -2P \oplus R = S.$$

In fact, P and R together generate the entirety of $E(\mathbb{Q})$, in the sense that

$$E(\mathbb{Q}) = \{aP + bR \colon a, b \in \mathbb{Z}\}.$$

So, unlike the Pell equation where a single fundamental solution generates the entire solution set (up to sign), with an elliptic curve E we might require several "fundamental solutions" to generate all of $E(\mathbb{Q})$. The amazing result here is that we will only ever need finitely many such solutions. That is, $E(\mathbb{Q})$ is a *finitely generated* group.

Theorem 32.13 **(Mordell's Theorem)**

Let E be an elliptic curve over $\mathbb{Q}$. There exist $P_1, \dots, P_k \in E(\mathbb{Q})$ such that

$$E(\mathbb{Q}) = \{a_1P_1 \oplus \cdots \oplus a_kP_k \colon a_i \in \mathbb{Z}\}.$$

This theorem has been generalized by André Weil, and so is sometimes called the Mordell–Weil theorem. Because of this, $E(\mathbb{Q})$ is called the **Mordell–Weil group** of E.

In Example 32.10, we saw that the Modell–Weil group of $E: \ y^2 = x^3 + 1$ is

$$E(\mathbb{Q}) = \{a(2,3) \colon a \in \mathbb{Z}\}.$$

In fact, we also saw that $(2,3)$ has finite order in $E(\mathbb{Q})$—specifically, its order is 6—so we actually have

$$E(\mathbb{Q}) = \{a(2,3) \colon a = 1, 2, \dots, 6\}.$$

If we squint, this group looks a lot like $\mathbb{Z}/6\mathbb{Z}$. In particular, it's finite.

On the other hand, the elliptic curve $E': \ y^2 = x^3 - 9x + 9$ of Example 32.12 has Mordell–Weil group

$$E'(\mathbb{Q}) = \{a(1,1) + b(3,3) \colon a \in \mathbb{Z}, b = 1, 2, 3\}.$$

The point $(1,1)$ has infinite order and $(3,3)$ has order 3. So, in this case, the Mordell–Weil group looks like $\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$. In particular, it's infinite.

Lecture 32 Problems

32.1. Show that the change of variables

$$x = Y + 36, \quad y = -Y + 36 \quad \text{and} \quad z = 6X$$

turns the Fermat cubic $x^3 + y^3 = z^3$ into an elliptic curve E of the form $Y^2 = X^3 + aX + b$. Assuming Fermat's Last Theorem, determine $E(\mathbb{Q})$.

32.2. Let E be the elliptic curve $y^2 = x^3 + ax + b$, and let $P_i = (x_i, y_i) \in E(\mathbb{Q})$ for $i = 1, 2$. Assume that $P_1 \neq -P_2$ and let $P_3 = (x_3, y_3) = P_1 \oplus P_2$. Prove:

(a) If $P_1 = P_2$ then

$$(x_3, y_3) = \begin{cases} \mathcal{O} & \text{if } y_1 = 0 \\ (m^2 - 2x_1, m(x_1 - x_3) - y_1) & \text{if } y_1 \neq 0, \end{cases}$$

where $m = (3x_1^2 + a)/2y_1$.

(b) If $P_1 \neq P_2$, then

$$(x_3, y_3) = (m^2 - x_1 - x_2, m(x_1 - x_3) - y_1),$$

where $m = (y_1 - y_2)/(x_1 - x_2)$.

Lecture 33 Elliptic Curves Over $\mathbb{Z}/p\mathbb{Z}$

I would like to now (attempt to) explain what it means for an elliptic curve to be *modular*. Recall that Taylor and Wiles proved that all (semistable) elliptic curves over $\mathbb{Q}$ are modular, and this was the final piece of puzzle in the proof of Fermat's Last Theorem. The definition of modularity involves working with elliptic curves modulo p.

Definition 33.1
Elliptic Curve Over $\mathbb{Z}/p\mathbb{Z}$

The equation $y^2 = x^3 + ax + b$, where $a, b \in \mathbb{Z}$, will be said to define an **elliptic curve** over $\mathbb{Z}/p\mathbb{Z}$ if $4a^3 + 27b^2 \not\equiv 0 \pmod{p}$.

This is meant to mirror Definition 32.1 except with $\mathbb{Z}/p\mathbb{Z}$ taking the place of $\mathbb{Q}$.[24] We similarly define the set of $\mathbb{Z}/p\mathbb{Z}$ points of E to be

$$E(\mathbb{Z}/p\mathbb{Z}) = \{(x, y) \in (\mathbb{Z}/p\mathbb{Z})^2 : y^2 \equiv x^3 + ax + b \pmod{p}\} \cup \{\mathcal{O}\},$$

Technically, we should be working with congruence classes $([x]_p, [y]_p)$ but I will drop the $[\,]_p$ just to keep the notation manageable. The point $\mathcal{O}$ is again a "point at infinity" that we include in $E(\mathbb{Z}/p\mathbb{Z})$ for things to work out nicely. In particular, we find that $E(\mathbb{Z}/p\mathbb{Z})$ is a group with the same definition of $\oplus$ as in the previous lecture. In fact, the explicit formulas for $P \oplus Q$ given in Problem 32.2 work just as well over $\mathbb{Z}/p\mathbb{Z}$ as they do over $\mathbb{Q}$.[25]

Example 33.2

Let E be defined by $y^2 = x^3 + 1$. We have

$$4a^3 + 27b^2 = 4(0)^3 + 27(1)^2 = 27 \not\equiv 0 \pmod{p}$$

if $p \neq 3$. Thus, E is an elliptic curve over $\mathbb{Z}/p\mathbb{Z}$ for all $p > 3$.

Let's determine $E(\mathbb{Z}/p\mathbb{Z})$ for $p = 5$ and 7. This is just a matter of plugging $x = 0, 1, \ldots, p-1$ into the equation for E and seeing if we can solve for y.

If $p = 5$, we find:

x	$x^3 + 1$	Square mod 5?	y
0	1	Yes	± 1
1	2	No	-
2	4	Yes	± 2
3	3	No	-
4	0	Yes	0

Thus,

$$E(\mathbb{Z}/5\mathbb{Z}) = \{(0, \pm 1), (2, \pm 2), (4, 0), \mathcal{O}\}.$$

[24] This definition needs to be modified if $p = 2$ or 3. The fix involves working with the more general equation $y^2 + axy + by = x^3 + cx^2 + dx + e$. We will sidestep this and assume that $p > 3$ in this lecture.

[25] The astute reader will have noticed that the formula for $P \oplus P$ involves dividing by 2, so it doesn't make sense mod p if $p = 2$. This is related to the previous footnote.

If $p = 7$, we find:

x	$x^3 + 1$	Square mod 7?	y
0	1	Yes	± 1
1	2	Yes	± 3
2	2	Yes	± 3
3	0	Yes	0
4	2	Yes	± 3
5	0	Yes	0
6	0	Yes	0

Thus,

$$E(\mathbb{Z}/7\mathbb{Z}) = \{(0, \pm 1), (1, \pm 3), (2, \pm 3), (3, 0), (4, \pm 3), (5, 0), (6, 0), \mathcal{O}\}.$$

Exercise 33.3 Determine $E(\mathbb{Z}/11\mathbb{Z})$ for the elliptic curve E given in the preceding example.

One thing that is immediately obvious is that $E(\mathbb{Z}/p\mathbb{Z})$ will always be a finite group, unlike $E(\mathbb{Q})$ which could contain elements of infinite order. In fact, since there are p choices for each $x, y \in \mathbb{Z}/p\mathbb{Z}$, and since $\mathcal{O} \in E(\mathbb{Z}/p\mathbb{Z})$, we see that

$$|E(\mathbb{Z}/p\mathbb{Z})| \leq p^2 + 1.$$

We can do quite a bit better, as the following heuristic argument shows. The congruence

$$y^2 \equiv x^3 + ax + b \pmod{p}$$

will have a solution if and only if $x^3 + ax + b$ is a quadratic residue mod p. We expect this to occur about half the time, since half the non-zero elements in $\mathbb{Z}/p\mathbb{Z}$ are quadratic residues. Whenever $x^3 + ax + b$ is a (non-zero) quadratic residue, we get two solutions for y. If we remember to include $\mathcal{O}$ into this count, we expect to find that

$$|E(\mathbb{Z}/p\mathbb{Z})| \approx 2 \times \frac{p}{2} + 1 = p + 1.$$

on average. The error in this approximation has a name.

Definition 33.4 Let E be an elliptic curve over $\mathbb{Z}/p\mathbb{Z}$, The **p-defect** of E is defined to be

$$a_p = (p + 1) - |E(\mathbb{Z}/p\mathbb{Z})|.$$

p-defect, a_p

Our heuristic argument above suggests that a_p should be fairly small relative to p. This is indeed the case.

Theorem 33.5 **(Hasse Bound)**

Let E be an elliptic curve over $\mathbb{Z}/p\mathbb{Z}$. Then $|a_p| \leq 2\sqrt{p}$.

For example, if E is given by $y^2 = x^3 + 1$, our calculations in Example 33.2 and Exercise 33.3 show that

$$\begin{aligned} a_5 &= (5+1) - |E(\mathbb{Z}/5\mathbb{Z})| = 6 - 6 = 0 \\ a_7 &= (7+1) - |E(\mathbb{Z}/7\mathbb{Z})| = 8 - 12 = -4 \\ a_{11} &= (11+1) - |E(\mathbb{Z}/11\mathbb{Z})| = 12 - 12 = 0 \end{aligned}$$

all of which satisfy the Hasse bound.

33.1 Modularity

Talent hits a target no one else can hit; genius hits a target no one else can see.

– A. Schopenhauer

Let's dive deeper into the elliptic curve defined by $y^2 = x^3 + 1$. Here is a table containing the values of a_p for all $p < 50$:

p	a_p
5	0
7	−4
11	0
13	2
17	0
19	8
23	0
29	0
31	−4
37	−10
41	0
43	8
47	0

Do you notice any patterns? It appears that a_p is nonzero if and only if $p \equiv 1 \pmod 6$. In general, when mathematicians want to study a sequence of numbers (a_n), they tend to package the sequence into a device called a generating function

$$f(q) = \sum_{n=1}^{\infty} a_n q^n = a_1 q + a_2 q^2 + \cdots .$$

The idea is that doing algebra with the generating function will often reveal hidden patterns in the sequence of coefficients.

An elliptic curve is said to be **modular** if the generating function built out of its p-defects exhibits specific behaviour. To be a little more accurate, $f(q)$ should be a *modular form*. This is definitely not the time or place to define what a modular form is. In broad strokes, a modular form is an analytic function that obeys interesting transformation rules.

It's instructive to think about a function like $\sin x$, which has the series expansion

$$\sin x = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1}.$$

The series expansion tells us that $\sin x$ is analytic (i.e. infinitely differentiable). What is not obvious from the series expansion is that $\sin x$ obeys certain identities such as

$$\sin(x + 2\pi) = \sin x.$$

A modular form is kind of like this, but more complicated. (To be clear: $\sin x$ is not a modular form.)

To see what this looks like for our elliptic curve $y^2 = x^3 + 1$, I first have to tell you how to define a_n for composite n. We begin by setting $a_1 = 1$ and $a_2 = a_3 = 0$. Then, for $k \geq 2$, we define

$$a_{p^k} = \begin{cases} (a_p)^k & \text{if } p = 2, 3 \\ a_p a_{p^{k-1}} - p a_{p^{k-2}} & \text{if } p > 3. \end{cases}$$

(There is good reason for this, I promise.) Finally, we let $a_{mn} = a_m a_n$ if $\gcd(m, n) = 1$.

So, for example,

$$a_{14} = a_2 a_7 = 0 \cdot (-4) = 0$$

and

$$a_{5^2} = (a_5)^2 - 5a_1 = -5.$$

The first few terms of the generating function are therefore

$$f(q) = q - 4q^7 + 2q^{13} + 8q^{19} - 5q^{25} - 4q^{31} - 10q^{37} + 8q^{43} + 9q^{49} + \cdots . \qquad (*)$$

The next step is to set $q = e^{2\pi i z}$, where $z \in \mathbb{C}$ is a complex number. Thus, our generating function becomes a Fourier series

$$F(z) = f(e^{2\pi i z}) = e^{2\pi i z} - 4e^{14\pi i z} + 2e^{26\pi i z} + \cdots .$$

This series converges for all z with positive imaginary part because $e^{2\pi i n z}$ will decay exponentially fast as $n \to \infty$. To say that it's a modular form amounts to saying that $F(z)$ is an analytic function (if $\mathrm{Im}(z) > 0$) and that F satisfies certain identities, such as

$$F\left(\frac{-1}{z}\right) = z^2 F(z) \quad \text{and} \quad F\left(\frac{z}{36z+1}\right) = (36z+1)^2 F(z).$$

I'll leave it at that!

Wiles proved that the generating function built out of the p-defects of *any* (semistable) elliptic curve over $\mathbb{Q}$ is a modular form. Thus, in a sense, he was able to bridge two worlds:

$$\text{elliptic curves} \quad \rightsquigarrow \quad \text{modular forms.}$$

This is remarkable because elliptic curves live in the world of arithmetic and algebra, whereas modular forms live in the world of analysis.

In closing, let me mention one (sort of) consequence of the modularity of the generating series $f(q)$ given above. It admits the following infinite product expansion:

$$f(q) = q \prod_{n=1}^{\infty} (1 - q^{6n})^4 = q(1 - q^6)^4 (1 - q^{12})^4 (1 - q^{18})^4 \cdots .$$

Try expanding out a little bit of this, and confirm that it matches $(*)$.

33.2 Integer Factorization Using Elliptic Curves

We now turn to the one remaining question in Section 14.1 that we have yet to address: How do we factor large integers? We can use addition of points on elliptic curves to help find divisors! It's best to start with an example.

Example 33.6 Suppose we want to factor $N = 65$. Let E be the elliptic curve $y^2 = x^3 - 9x + 9 \bmod 65$. (This is technically undefined, since 65 is not prime. But we will pretend as though it were.) Let $P = (1, 1) \in E(\mathbb{Z}/65\mathbb{Z})$, and let's try to calculate $9P$.

Using our formula for $2P$ in Problem 32.2(a), we find that

$$\begin{aligned} 2P &= (7, 17) \\ 4P &= (50, 57) \\ 8P &= (31, 54). \end{aligned}$$

In the above we had to calculate the tangent line slope $m = (3x^2 - 9)/2y$, where $1/2y$ is to be interpreted as $(2y)^{-1} \bmod 65$. In each case $2y$ ended up being coprime to 65, so this was no problem.

However, when we try to calculate $9P = P + 8P$, the slope

$$m = \frac{54 - 1}{31 - 1} = \frac{53}{30}$$

has denominator that isn't invertible mod 65. Specifically, $\gcd(65, 30) = 5$. We have found a non-trivial divisor of $N = 65$!

What happened in the preceding example is the following. The point $P = (1, 1)$ has order 9 modulo 5, in the sense that

$$9P = \mathcal{O} \quad \text{in } E(\mathbb{Z}/5\mathbb{Z}).$$

But $9P$ is the finite point $(10, 10)$ in $E(\mathbb{Z}/13\mathbb{Z})$. This kind of clash resulted in us discovering 5 in the slope denominator of $9P$. (If the order of P were 9 in mod 13 too, then we would have found a $5 \cdot 13 = 65 = N$ in the denominator, so we would not have discovered a nontrivial divisor of N.)

Hendrik Lenstra used this observation as the basis of his **Elliptic Curve Method (ECM)** for integer factorization. The basic idea is that a point P in $E(\mathbb{Z}/N\mathbb{Z})$ may be viewed as a point in $E(\mathbb{Z}/p\mathbb{Z})$ for each prime $p \mid N$. If E is randomly chosen and if k is a reasonably large integer, then we will likely have $kP = \mathcal{O}$ in $E(\mathbb{Z}/p\mathbb{Z})$ (if not, then choose another E). If $kP = \mathcal{O}$ in $E(\mathbb{Z}/p\mathbb{Z})$ it will be unlikely that $kP = \mathcal{O}$ in $E(\mathbb{Z}/q\mathbb{Z})$ for a different prime $q \mid N$. (This is because, by the Hasse bound, the orders of $E(\mathbb{Z}/p\mathbb{Z})$ and $E(\mathbb{Z}/q\mathbb{Z})$ are close to $p + 1$ and $q + 1$, resp., so will unlikely share many divisors if $p \neq q$. On the other hand, if $kP = \mathcal{O}$ in E then k divides the order of E.) Once we find such a P, then $\gcd(k, N) > 1$ will be a divisor of N.

ALGORITHM (Lenstra's ECM)

To factor an odd integer N:

1. Choose a large integer k (such as $k = B!$ for $B \approx 10^8$) and randomly select several elliptic curves E_i and points $P_i \in E_i(\mathbb{Z}/N\mathbb{Z})$.
2. Compute kP_i in $E_i(\mathbb{Z}/N\mathbb{Z})$.
3. If Step 2 fails because some slope has a denominator d not coprime to n, calculate $\gcd(d, n)$ to find a divisor of n.
4. If $\gcd(d, n) < n$ in Step 3 then you have found a nontrivial divisor of n. Stop.
5. If Step 2 succeeds or if $\gcd(d, n) = n$, repeat Step 1 with a larger k and/or new random E_i and P_i.

This is currently the most efficient algorithm for finding divisors not exceeding 50 or so digits. In practice, it's used in a first pass-through to detect all small divisors. For the remaining divisors, other algorithms (such as the number field sieve) must be employed.

Lecture 33 Problems

33.1. Go to the LMFDB page for the elliptic curve $y^2 = x^3 + 1$. Poke around and see if you can spot anything interesting there. Can you find the associated modular form?

33.2. Look up the Goldwasser–Killian Elliptic Curve Primality Proving (ECPP) algorithm. How does it compare to the primality tests we learned about in Lecture 23?

Lecture 34 What's Next?

We are at the end of our course—but there's plenty more number theory left to explore! Here are some suggestions if you're interested in learning more.

- Are you curious about the proofs of the Prime Number Theorem or Dirichlet's theorem on primes in arithmetic progression? Consider PMATH 440 - Analytic Number Theory. You will need to beef up your complex analysis (PMATH 352).
- Want to learn how to do number theory in number systems like $\mathbb{Z}[\sqrt{-D}\,]$ or even more exotic ones like $\mathbb{Z}[e^{2\pi i/n}]$ (without making fallacious arguments like we did in Lecture 28)? Then PMATH 441 - Algebraic Number Theory is the course for you.
- Did you enjoy our little preview of groups? Do you want to learn what a "ring" is so that you don't keep saying "number system"? Then consider a course in abstract algebra, such as PMATH 334, 336, 347 and/or 348. This stuff is a pre-req to PMATH 441.
- Were you intrigued by elliptic curves and want to investigate them further? For an introduction, it's hard to beat the book by Silverman and Tate that I mentioned in Lecture 32. Eventually you will want to read Silverman's *The Arithmetic of Elliptic Curves*, and for this you will want to know a thing or two about algebraic geometry; PMATH 464 has you covered.
- Finally, if you want to dive more deeply into the mathematics of cryptography, then you should check out CO 485 and CO 487.

Good luck!

Appendix A

Solutions to Exercises and Practice Problems

Lecture 1

Exercises

1.2 (a) Simply note that

$$(na)^2 + (nb)^2 = n^2a^2 + n^2b^2 = n^2(a^2 + b^2) = n^2c^2 = (nc)^2.$$

(b) Let's plug $(x, y) = (3a + 4b, 2a + 3b)$ into $x^2 - 2y^2$ and confirm that we get 1:

$$\begin{aligned} x^2 - 2y^2 &= (3a + 4b)^2 - 2(2a + 3b)^2 \\ &= 9a^2 + 24ab + 16b^2 - 2(4a^2 + 12ab + 9b^2) \\ &= a^2 - 2b^2 \\ &= 1, \end{aligned}$$

where the last equality holds because $(x, y) = (a, b)$ is a solution to $x^2 - 2y^2 = 1$.

1.3 *No solution provided.*

Practice Problems

1.1. *No solution provided. Assignment problem.*

1.2. Let's treat y as a constant and view this as an equation in x. The divisors of the constant term are ± 1 and ± 7. Let's try these one at a time:

- $x = 1$ gives $2 + y - 7 = 0$ hence $y = 5$. So $(x, y) = (1, 5)$ is a solution.
- $x = -1$ gives $-2 - y - 7 = 0$ hence $y = -9$. So $(x, y) = (-1, 9)$ is a solution.
- $x = 7$ gives $2 \cdot 7^3 + 7y - 7 = 0$ hence $y = -97$. So $(x, y) = (7, -97)$ is a solution.
- $x = -7$ gives $-2 \cdot 7^3 - 7y - 7 = 0$ hence $y = -99$. So $(x, y) = (-7, -99)$ is a solution.

The above four solutions are the only solutions.

1.3. The solutions are given by the quadratic formula:

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

So if $x \in \mathbb{Z}$ then $2ax + b \in \mathbb{Z}$ and hence $\pm\sqrt{b^2 - 4ac} = n$ is an integer. Squaring both sides, we get that $b^2 - 4ac = n^2$ is a perfect square. This proves the first part.

The converse does not hold. Consider, for example, the equation $4x^2 - 1 = 0$. Here, $b^2 - 4ac = 0 - 4(-1)(4) = 16$ is a perfect square but the roots $\pm\frac{1}{2}$ are not integers.

To guarantee integer solutions, we need $b^2 - 4ac$ to be a perfect square n^2 and we need $-b \pm n$ to be divisible by $2a$. (This can be proved by examining the quadratic formula.)

1.4. If the n cannonballs can be arranged into a square, then that means $n = y^2$ for some $y \in \mathbb{Z}$.

A square pyramid with x layers will consist of 1 cannonball at the top, 2^2 cannonballs at the next layer, 3^2 cannonballs at the layer after next, and so on, ending with x^2 cannonballs at the bottom layer. Thus, $n = 1 + 2^2 + 3^2 + \cdots + x^2$.

So our desired Diophantine equation is

$$y^2 = 1^2 + 2^2 + 3^2 + \cdots + x^2$$

or equivalently

$$y^2 = \frac{x(x+1)(2x+1)}{6}.$$

1.5. (a) *Omitted — messy but straightforward algebra.*

(b) If $y = 28$, then the equation turns into the univariate Diophantine equation

$$x^3 + 109x^2 + 224x - 28^2 = 0.$$

Now we just have to try all divisors of 28^2. Doing so, we discover that there is a solution when $x = -4$. Thus, $(x, y) = (-4, 28)$ is a solution to the equation in part (a). Using the formulas given there for a, b, and c, we can deduce that $(a, b, c) = (11, 4, -1)$ is a solution to the original equation.

Lecture 2

Exercises

2.2 We can write $b = ka$ and $c = la$ with $k, l \in \mathbb{Z}$. Then $xb+yc = x(ka)+y(la) = (xk+yl)a$. Since $xk + yl \in \mathbb{Z}$, it follows that $a \mid xb + yc$, as required.

For the converse, observe that if $a \mid xb + yc$ for *all* choices of $x, y \in \mathbb{Z}$, the a will divide $xb + yc$ for the choices $(x, y) = (1, 0)$ and $(x, y) = (0, 1)$. This gives us $a \mid b$ and $a \mid c$, respectively.

2.5 We have $\lfloor x \rfloor \leq x$ by definition. Next, notice that there must be an integer between $x - 1$ and x, so we get $x - 1 \leq \lfloor x \rfloor$. If this were an equality, then $x - 1$ would be an integer, and hence x would be an integer—which would mean that we actually have $x = \lfloor x \rfloor = x - 1$, which is absurd! So the inequality is strict: $x - 1 < \lfloor x \rfloor$.

2.7 The quotient is $q = \lfloor -75/6 \rfloor = \lfloor -12.5 \rfloor = -13$ and the remainder is $r = -75 - 6 \cdot (-13) = 3$.

2.9 The remainder theorem shows that every $a \in \mathbb{Z}$ can be written in the form $a = 3q$, $a = 3q + 1$ or $a = 3q + 2$. All we have to do is observe that every integer of the form $3q + 2$ can be re-expressed in the form $3k - 1$:

$$3q + 2 = 3(q + 1 - 1) + 2 = 3(q + 1) - 3 + 2 = 3(q + 1) - 1.$$

2.12 In Example 2.11 we saw that if a and b are both odd then $a^2 + b^2$ leaves a remainder of 2 after division by 4. If $a = 2k$ and $b = 2l$ are both even then

$$a^2 + b^2 = 4k^2 + 4l^2 = 4(k^2 + l^2)$$

is divisible by 4 hence leaves a remainder of zero. Finally, if only one of a and b is odd, say a, then

$$a^2 + b^2 = (2k + 1)^2 + (2l)^2 = 4k^2 + 4k + 1 + 4l^2 = 4(k^2 + k + l^2) + 1$$

leaves a remainder of 1.

2.13 If $x, y \in \mathbb{Z}$ satisfy $x^2 + y^2 = 6$ then

$$x^2 = 6 - y^2 \leq 6 \implies |x| \leq \sqrt{6} \implies |x| \leq 2.$$

So let's plug in $x = 0, \pm 1, \pm 2$ into $6 - x^2$ and see if we get a square (since this should be equal to y^2):

- $x = 0$: $6 - x^2 = 0$ — not a square.
- $x = \pm 1$: $6 - x^2 = 5$ — not a square.
- $x = \pm 2$: $6 - x^2 = 2$ — not a square.

Thus, none of the possible integer values of x satisfy the equation, and so there can be no integer solutions.

Practice Problems

2.1. (a) $a = a \cdot 1$.

(b) $0 = 0 \cdot a$.

(c) If $0 \mid a$ then $a = n0 = 0$. The converse is part (b).

(d) If $a = 0$ then this follows from (c). So assume $a \neq 0$. If $a \mid b$ then $b = na$ for some $n \in \mathbb{Z}$. If $b \mid a$ then $a = mb$ for some $m \in \mathbb{Z}$. Combining both, we get

$$a = mb = m(na) = (mn)a.$$

Since $a \neq 0$, it follows that $mn = 1$. Since $m, n \in \mathbb{Z}$, it follows that either $m = n = 1$ or $m = n = -1$. In either case, we get that $a = mb = \pm b$, as desired.

(e) If $a \mid b$ then $b = ma$ with $m \in \mathbb{Z}$ hence $b^n = m^n a^n$ so $a^n \mid b^n$ since $m^n \in \mathbb{Z}$.

2.2. (a) True. If $b = na$ and $d = mc$ then $bd = (nm)ac$.

(b) False. For example, $1 \mid 2$ and $1 \mid 3$ but $1 + 1 \nmid 2 + 3$.

(c) False. $4 \mid 2 \cdot 2$ but $4 \nmid 2$ and $4 \nmid 2$.

(d) False. Same example in (c) works here.

2.3. *No solution provided. Assignment problem.*

2.4. If $b < 0$ then $-b > 0$ so we can apply The Remainder Theorem to a and $-b$. Doing so, we can write $a = (-b)q + r$, where $0 \leq r < -b$. This is the same as $a = b(-q) + r$ and $0 \leq r < |b|$, since $|b| = -b$ if $b < 0$.

2.5. (a) Let $S = \{a - nb \colon n \in \mathbb{Z}\} \cap \mathbb{Z}_{\geq 0}$. If $a \geq 0$ then $a - 0b \in S$.
If $a < 0$ then, since $b \geq 1$, $a - ab = a(1 - b) \geq 0$ is in S.

(b) By the definition of S, $r = a - qb$ for some q and $r \geq 0$. Suppose for a contradiction that $r \geq b$. Then $r - b \geq 0$. However, $r - b = a - qb - b = a - (q + 1)b$ is in S and clearly $r - b < r$. This contradicts the minimality of r. So $r < b$, as desired.

(c) From (b), we have $r = a - qb$ with $0 \leq r < b$. This is what Theorem 2.3 (The Remainder Theorem) asserts, since the first equation can be re-written as $a = qb + r$.

Lecture 3

Exercises

3.2 If $a = 0$ then this is true by definition, since $\gcd(0, 0) = 0$. So we may assume that $a \neq 0$.

First, since $|a|$ divides both a and 0, it's a common divisor of both. Second, of all the divisors of a, $|a|$ is the largest (by Proposition 2.1(c)). It follows that $|a|$ is the greatest common divisor of a and 0.

3.4 If $d \mid a$ and $d \mid b$, then d will divide the linear combination $a - qb$. Hence every common divisor of a and b is a common divisor of b and $a - qb$.

Conversely, if $d \mid b$ and $d \mid a - qb$, d will divide the linear combination $q \cdot b + 1 \cdot (a - qb) = a$. So every common divisor of b and $a - qb$ is a common divisor of a and b.

3.6 Since $\gcd(1234, 5678) = \gcd(5678, 1234)$, we can run the Euclidean algorithm with $a = 5678$ and $b = 1234$:

$$\begin{aligned} 5678 &= 4 \cdot 1234 + 742 \\ 1234 &= 1 \cdot 742 + 492 \\ 742 &= 1 \cdot 492 + 250 \\ 492 &= 1 \cdot 250 + 242 \\ 250 &= 1 \cdot 242 + 8 \\ 242 &= 30 \cdot 8 + 2 \\ 8 &= 4 \cdot 2 + 0. \end{aligned}$$

We've reached a zero remainder! So what we want is the last non-zero remainder, that is,

$$\gcd(1234, 5678) = 2.$$

3.10 In the preceding exercise, we computed $\gcd(1234, 5678)$ to be 2. We can run our Euclidean algorithm backwards to find that

$$\begin{aligned} 2 &= 242 - 30 \cdot 8 \\ &= 242 - 30 \cdot (250 - 242) \\ &= -30 \cdot 250 + 31 \cdot 242 \\ &= -30 \cdot 250 + 31(492 - 250) \\ &= -61 \cdot 250 + 31 \cdot 492 \\ &= -61 \cdot (742 - 492) + 31 \cdot 492 \\ &= -61 \cdot 742 + 92 \cdot 492 \\ &= -61 \cdot 742 + 92(1234 - 742) \\ &= -153 \cdot 742 + 92 \cdot 1234 \\ &= -153(5678 - 4 \cdot 1234) + 92 \cdot 1234 \\ &= 704 \cdot 1234 + (-153) \cdot 5678. \end{aligned}$$

Practice Problems

3.1. If $c \mid \gcd(a,b)$ then since $\gcd(a,b)$ divides both a and b, c must too, by transitivity of division.

For the converse, apply Bézout's lemma to write $\gcd(a,b) = ax + by$ for some $x, y \in \mathbb{Z}$. Then if $c \mid a$ and $c \mid b$, it follows that $c \mid ax + by = \gcd(a,b)$.

3.2. (a) Let $g = \gcd(na, nb)$ and $h = \gcd(a,b)$. We wish to show that $g = |n|h$. We will show that $g \mid |n|h$ and $|n|h \mid g$. It will follow then that $g = |n|h$ since both are ≥ 0. (Recall that if $a \mid b$ and $b \mid a$ then $b = \pm a$.)

Proof that g divides $|n|h$: By Bézout's Lemma, we can write $h = ax + by$ with $x, y \in \mathbb{Z}$. Then $|n|h = (\pm n)h = na \cdot (\pm x) + nb \cdot (\pm y)$. So, since $g \mid na$ and $g \mid nb$, it follows that $g \mid |n|h$.

Proof that $|n|h$ divides g: Note that $|n|h = \pm nh$. Since $h \mid a$ it follows that $nh \mid na$ and hence $|n|h = \pm nh \mid a$. Similarly, since $h \mid b$, we find that $|n|h \mid nb$. Hence, by Problem 3.1 above, $|n|h$ must divide $\gcd(na, nb) = g$.

(b) Note that a/d and b/d are integers. So, by part (a),

$$\gcd(a,b) = \gcd\left(d\frac{a}{d}, d\frac{b}{d}\right) = |d| \gcd\left(\frac{a}{d}, \frac{b}{d}\right).$$

Now divide both sides by $|d|$.

3.3. *No solution provided. Assignment problem.*

3.4. (a) If d is a common divisor of $a_1, \ldots, a_n$, then d is in particular a common divisor of $\gcd(a_1, a_2)$, hence a common divisor of $\gcd(a_1, a_2), a_3, \ldots, a_n$.

Conversely, if d is a common divisor of $\gcd(a_1, a_2), a_3, \ldots, a_n$, then d is a divisor of $\gcd(a_1, a_2)$ and hence d is also a common divisor of a_1 and a_2.

This shows that the every common divisors of $a_1, \ldots, a_n$ is a common divisor of $\gcd(a_1, a_2), a_3, \ldots, a_n$, and vice versa. So it must be the case that their gcd's are the same.

(b) We apply part (a) repeatedly:

$$\begin{aligned}
\gcd(20, 28, 100, 36) &= \gcd(\gcd(20, 28), 100, 36) \\
&= \gcd(4, 100, 36) \\
&= \gcd(\gcd(4, 100), 36) \\
&= \gcd(4, 36) \\
&= 4.
\end{aligned}$$

Lecture 4

Exercises

4.4 **Proof 1 (using Bézout):** Since a and b are coprime, we have $1 = ax + by$ for some $x, y \in \mathbb{Z}$. Now multiply this by c to get $c = cax + cby$. Since $a \mid c$ and $b \mid c$, we can write $c = ak$ and $c = bl$ for some $k, l \in \mathbb{Z}$. Therefore,

$$c = cax + cby = (bl)ax + (ak)by = ab(lx + ky).$$

So $ab \mid c$, as required.

Proof 2 (using Proposition 4.3(a)): Since $a \mid c$ and $b \mid c$, we can write $c = ak$ and $c = bl$ for some $k, l \in \mathbb{Z}$. Therefore, $ak = bl$. From this we see that $a \mid bl$ and thus $a \mid l$ since a and b are coprime. So we can write $l = ar$ for some $r \in \mathbb{Z}$. But then $c = bl = (ab)r$ and hence $ab \mid c$.

4.8 We can solve this problem by inspection, but let's run through our algorithm. We begin by applying the Euclidean algorithm to determine $\gcd(5, 7)$:

$$\begin{aligned} 7 &= 1 \cdot 5 + 2 \\ 5 &= 2 \cdot 2 + 1 \\ 2 &= 2 \cdot 1 + 0. \end{aligned}$$

So $\gcd(5, 7) = 1$. Reversing the Euclidean algorithm, we get:

$$1 = 5 - 2 \cdot 2 = 5 - 2(7 - 1 \cdot 5) = 5 \cdot 3 + 7 \cdot (-2).$$

(The above could have been determined by inspection.) Multiplying through by 23, we find that

$$23 = 5 \cdot 69 + 7 \cdot (-46).$$

So a particular solution is $(x, y) = (69, -46)$ and the general solution is therefore

$$(x, y) = (69, -46) + n(-7, 5), \quad n \in \mathbb{Z}.$$

Practice Problems

4.1. If $d \mid a$ and $d \mid b$ then $d \mid ra + sb = 2$ and $d \mid ta + tb = 5$, so $d \mid \gcd(2, 5) = 1$. Thus $d = \pm 1$. So the only common divisors of a and b are ± 1, hence a and b must be coprime.

4.2. (a) By Lemma 3.3, we have $\gcd(2a - 1, 2a + 1) = \gcd((2a - 1) - (2a + 1), 2a + 1) = \gcd(-2, 2a + 1)$. Since $2a + 1$ is odd, the last gcd is 1, so $2a - 1$ and $2a + 1$ are coprime.

(b) Note that $(a + 1)(a! + 1) = (a + 1)! + (a + 1)$. So, by Lemma 3.3,

$$\gcd(a!+1, (a+1)!+a) = \gcd(a!+1, (a+1)!+a-(a+1)(a!+1)) = \gcd(a!+1, -1) = 1.$$

4.3. *No solution provided. Assignment problem.*

4.4. (a) Any solution $(u, z) = (u_0, z_0)$ to (♣) gives us a potential value for z and a two-variable equation

$$\frac{a}{d}x + \frac{b}{d}y = u_0.$$

Since $\gcd(a/d, b/d) = 1$, this equation always has solutions $(x, y) = (x_0, y_0)$. Then $(x, y, z) = (x_0, y_0, z_0)$ are solutions to (♢).

Conversely, any solution (x_0, y_0, z_0) to (♢) gives a solution to (♣) with $(u, z) = ((a_1/d)x_0 + (a_2/d)y_0, z_0)$.

(b) If there are integers $x_1, x_2, x_3 \in \mathbb{Z}$ that satisfy (♢), then since $g \mid a_1x_1+a_2x_2+a_3x_3$, we must have that $g \mid b$.

Conversely, assume that $g \mid b$. Since $g = \gcd(\gcd(a, b), c) = \gcd(d, c)$, it follows that (♣) has integer solutions by Theorem 4.6(a). By part (a), such solutions give integer solutions to (♢).

4.5. *No solution provided. Assignment problem.*

Lecture 5

Exercises

5.4 Induction on n plus Euclid's lemma. The base case $n = 1$ is immediate. So assume the result holds for $n = k$ and suppose that $p \mid a_1 \cdots a_k a_{k+1}$. If we view $a_1 \cdots a_{k+1}$ as ab, where $a = a_1 \cdots a_k$ and $b = a_{k+1}$, Euclid's lemma tells us that $p \mid a$ or $p \mid b$. If the latter holds, we're done. If the former holds, apply the inductive hypothesis to conclude that $p \mid a_i$ for some $i \leq k$. We're done in either case.

5.6 Since $100 = 2^2 \cdot 5^2$, we have $v_2(100) = v_5(100) = 2$, and $v_p(100) = 0$ for all $p \neq 2, 5$.

5.8 We have

$$\begin{aligned} v_3(1000!) &= \left\lfloor \frac{1000}{3} \right\rfloor + \left\lfloor \frac{1000}{3^2} \right\rfloor + \left\lfloor \frac{1000}{3^3} \right\rfloor + \cdots \\ &= 333 + 111 + 37 + 12 + 4 + 1 + 0 + \cdots \\ &= 498. \end{aligned}$$

5.9 One is obvious: $9 = 3^2$. To find the other one, notice that

$$9 = 4 + 5 = 4 - (-5) = 2^2 - (\sqrt{-5})^2$$

hence

$$9 = (2 - \sqrt{-5})(2 + \sqrt{-5}).$$

Note that the given factorization of 6 comes from

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

Note: There remains the issue of checking that $2 \pm \sqrt{-5}$ and $1 \pm \sqrt{-5}$ (and 2 and 3) are "prime" in $\mathbb{Z}[\sqrt{-5}]$. But take this for granted for now. This would require a more thorough examination of what it means for a number to be prime.

Practice Problems

5.1. Suppose that q is not prime, so that $q = ab$ for some $a, b \in \mathbb{Z}$ with $1 < a < q$ and $1 < b < q$. Then certainly $q \mid ab$ but $q \nmid a$ and $q \nmid b$ since $a, b < q$.

5.2. Suppose that $a = \prod_i p^{a_i}$ and $b = \prod_i p_i^{b_i}$ are the prime factorizations of a and b, where $a_i \geq 0$ and $b_i \geq 0$ to allow for the same set of primes to occur in both factorizations. Then $a^5 = \prod_i p^{5a_i}$ and $b^2 = \prod_i p^{2b_i}$. So, since $a^5 \mid b^2$, Euclid's lemma tells us that $p^{5a_i} \mid p^{2b_i}$ for all i. Thus, $5a_i \leq 2b_i$, and consequently $b_i \geq (5/2)a_i \geq a_i$ for all i. This implies that $p_i^{a_i} \mid p^{b_i}$ for all i and therefore $a \mid b$. (For the last step, we can either appeal to Proposition 4.3(b) or observe that

$$b = \prod_i p_i^{b_i} = \prod_i p_i^{a_i} p_i^{b_i - a_i} = \prod_i p_i^{a_i} \prod_i p_i^{b_i - a_i} = ac$$

where $c = \prod_i p_i^{b_i - a_i}$ is an integer since $b_i - a_i \geq 0$ for all i.)

5.3. (a) Since $a = \prod_p p^{v_p(a)}$ and $b = \prod_p p^{v_p(b)}$, it follows that

$$ab = \prod_p p^{v_p(a)} p^{v_p(b)} = \prod_p p^{v_p(a)+v_p(b)}.$$

Thus, $v_p(ab) = v_p(a) + v_p(b)$ by definition.

(b) Repeated application of part (a) with $a = b$. For example,

$$v_p(a^2) = v_p(aa) = v_p(a) + v_p(a) = 2v_p(a).$$

(c) Let $k = v_p(a)$ and $l = v_p(b)$ and assume, without loss of generality that $k \leq l$ so that $\min(v_p(a), v_p(b)) = k$. We must prove that $v_p(a + b) \geq k$ with equality of $l > k$.

Write $a = p^k a'$ and $b = p^l b'$. So, in particular, $p \nmid a'$ and $p \nmid b$. Thus,

$$a + b = p^k a' + p^l b' = p^k(a' + p^{l-k} b').$$

Note that $a' + p^{l-k}b'$ is an integer since $l \geq k$. This shows that $p^k \mid a + b$ so $v_p(a + b) \geq k$.

Suppose now that $v_p(a) \neq v_p(b)$ so that $l > k$. Then $a' + p^{l-k}b'$ is not divisible by p (because if it were then since p divides $p^{l-k}b'$, then p would divide $a' + p^{l-k}b' - p^{l-k}b' = a'$). Thus, p^k is the highest power of p that divides $a+b$. So $v_p(a+b) = k$.

(d) If $a \mid b$ then since $p^{v_p(a)} \mid a$, it follows that $p^{v_p(a)} \mid b$. So $v_p(b)$ is at least $v_p(a)$.

Conversely, assume that $v_p(b) \geq v_p(a)$ for all p. We have

$$\begin{aligned} b = \prod_p p^{v_p(b)} &= \prod_p p^{v_p(b)-v_p(a)} p^{v_p(a)} \\ &= \prod_p p^{v_p(b)-v_p(a)} \prod_p p^{v_p(a)} \\ &= \left(\prod_p p^{v_p(b)-v_p(a)} \right) a. \end{aligned}$$

The first factor on the right is an integer, since $v_p(a) - v_p(a) \geq 0$, so it follows that $a \mid b$.

(e) If $k \mid v_p(a)$ for all p, so that $v_p(a) = n_p k$ for some $n_p \in \mathbb{Z}$, we find that

$$a = \prod_p p^{v_p(a)} = \prod_p p^{n_p k} = \left(\prod_p p^{n_p} \right)^k$$

so a is a kth power of an integer.

Conversely, if $a = b^k$, then using the prime factorization of b we get

$$a = \left(\prod_p p^{v_p(b)} \right)^k = \prod_p p^{k v_p(b)}$$

from which it follows that $v_p(a) = k v_p(b)$, and therefore $k \mid v_p(a)$.

5.4. Any positive common divisor d of a and b will be of the form $d = \prod_i p_i^{n_i}$ where $n_i \leq a_i$ and $n_i \leq b_i$, since any prime divisor of d must divide both a and b. The way to maximize this divisor is to maximize n_i. This is achieved when $n_i = \min(a_i, b_i)$.

5.5. (a) The divisors of p^k are $\pm 1, \pm p, \ldots, \pm p^k$. So there are $2(k+1)$ divisors.

(b) The divisors of $p^k q^l$ are of the form $\pm p^i q^j$ where $0 \le i \le k$ and $0 \le j \le l$. There are $2(k+1)(l+1)$ such numbers.

5.6. (a) We first observe that if $b \mid a$, then $v_p(a/b) = v_p(a) - v_p(b)$. (We can prove this using problem 3(a): $v_p(a) = v_p((a/b)b) = v_p(a/b) + v_p(b)$.)
Consequently,

$$v_p\left(\binom{n}{k}\right) = v_p(n!) - v_p(k!) - v_p((n-k)!).$$

Now apply Legendre's formula.

(b) We have

$$\begin{aligned} v_2\left(\binom{600}{300}\right) &= \sum_{i=1}^{\infty} \left\lfloor \frac{600}{p^i} \right\rfloor - \left\lfloor \frac{300}{p^i} \right\rfloor - \left\lfloor \frac{300}{p^i} \right\rfloor \\ &= \sum_{i=1}^{\infty} \left\lfloor \frac{600}{p^i} \right\rfloor - 2\sum_{i=1}^{\infty} \left\lfloor \frac{300}{p^i} \right\rfloor \\ &= (300 + 150 + 75 + 37 + 18 + 9 + 4 + 2 + 1) \\ &\quad - 2(150 + 75 + 37 + 18 + 9 + 4 + 2 + 1) \\ &= 4. \end{aligned}$$

5.7. What must be proved is that $v_p\left(\binom{p}{k}\right) = 1$. From the previous exercise, we have

$$v_p\left(\binom{p}{k}\right) = \sum_{i=1}^{\infty} \left\lfloor \frac{p}{p^i} \right\rfloor - \left\lfloor \frac{p-k}{p^i} \right\rfloor - \left\lfloor \frac{k}{p^i} \right\rfloor.$$

If $k < p$, then $\left\lfloor \frac{k}{p^i} \right\rfloor = 0$ for all $i \ge 1$. Likewise, since $k > 0$, we have that $p - k < p$, and so $\left\lfloor \frac{p-k}{p^i} \right\rfloor = 0$ for all $i \ge 1$. Also, $\left\lfloor \frac{p}{p^i} \right\rfloor = 0$ for all $i > 1$. Consequently,

$$v_p\left(\binom{p}{k}\right) = \sum_{i=1}^{\infty} \left\lfloor \frac{p}{p^i} \right\rfloor - 0 - 0 = \left\lfloor \frac{p}{p} \right\rfloor = 1,$$

as required.

Lecture 6

Exercises

6.7. *Answer:* $7 + 30q$, $0 \leq q \leq 5$.

Practice Problems

6.1. (a) Euclid's proof shows that $p_1 \cdots p_n + 1$ must have a prime divisor q that isn't among the first n primes $p_1, \ldots, p_n$. Whatever q is, it's not among the first n primes, so it's one of $p_{n+1}, p_{n+2}, \ldots$. In particular then, $p_{n+1} \leq q$. Since $q \mid N$, it follows that $q \leq N$ and hence $p_{n+1} \leq N$, as desired.

(b) The base case is obvious ($p_1 = 2$ and $2 = 2^{2^{1-1}}$). So assume $p_k \leq 2^{2^{k-1}}$ for all $k = 1, \ldots, n$. Then, by part (a),

$$p_{n+1} \leq p_1 \cdots p_n + 1 \leq 2^{2^{1-1}} 2^{2^{2-1}} \cdots 2^{2^{n-1}} + 1 = 2^{2^0 + 2^1 + \cdots 2^{n-1}} + 1.$$

Now,

$$2^0 + 2^1 + \cdots 2^{n-1} = \frac{2^n - 1}{2 - 1} = 2^n - 1.$$

Finally, since $a + 1 \leq 2a$ for $a \geq 1$, we have

$$p_{n+1} \leq 2^{2^n - 1} + 1 \leq 2(2^{2^n - 1}) = 2^{2^n},$$

as desired.

6.2. (a) Induction. Base case is easy, and here is the inductive step:

$$\begin{aligned} F_0 \cdots F_{n+1} &= (F_0 \cdots F_n) F_{n+1} \\ &= (F_{n+1} - 2) F_{n+1} \\ &= (2^{2^{n+1}} - 1)(2^{2^{n+1}} + 1) \\ &= 2^{2^{n+2}} - 1 \\ &= F_{n+2} - 2. \end{aligned}$$

(b) Suppose that $m > n$ and that $d \mid F_n$ and $d \mid F_m$. Then by part (a) $F_m = F_0 \cdots F_n \cdots F_{m-1} + 2$, so we must have that $d \mid 2$. So the only possibilities for d are ± 1 and ± 2. The latter are impossible since $d \mid F_n$ and F_n is odd. So $d = \pm 1$, and therefore $\gcd(F_n, F_m) = 1$.

(c) If p_n is a prime dividing F_n (and there must be such a prime since $F_n > 1$), then $p_1, p_2, \ldots$ is an infinite sequence of primes since by part (b) we have $p_i \neq p_j$ for all $i \neq j$.

6.3. *No solution provided. Assignment problem.*

6.4. *No solution provided. Assignment problem.*

6.5. The ath term in the progression will necessarily be composite. Indeed, if $q = a$ then $a \mid a + qb$ and $a \neq a + qa$.

Lecture 7

Exercises

7.3. *Answer:* 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97.

7.5. We have to check divisibility by primes $\leq \lfloor\sqrt{1891}\rfloor = 43$. Doing so, we find that $31 \mid 1891$, so 1891 is not prime.

7.10. First off, we have

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

Taking logs of both sides (and passing the log through the limit, which is valid because log is continuous), we have

$$0 = \log(1) = \lim_{n\to\infty} \left(\log n + \log(\log p_n) - \log p_n\right).$$

Thus,

$$\lim_{n\to\infty} \frac{\log n}{\log p_n} = \lim_{n\to\infty} \frac{\log p_n - \log(\log p_n)}{\log p_n} = \lim_{n\to\infty} 1 - \frac{\log(\log p_n)}{\log p_n} = 1 - 0 = 1,$$

where we used the fact that $p_n \to \infty$ and $\lim_{x\to\infty} \log(\log x)/\log x = 0$ (use, e.g., L'Hopital's rule).

So, finally,

$$\lim_{n\to\infty} \frac{n/\log n}{p_n} = \lim_{n\to\infty} \frac{n}{p_n \log n} \cdot \frac{\log p_n}{\log p_n} = \lim_{n\to\infty} \underbrace{\frac{n\log p_n}{p_n}}_{\to 1} \underbrace{\frac{\log n}{\log p_n}}_{\to 1} = 1,$$

which is exactly what we want to prove.

7.13. If $\lfloor x\rfloor = a$ then

$$a \leq x < a + 1.$$

So

$$2a \leq 2x < 2a + 2.$$

Thus, $\lfloor 2x\rfloor$ could be either $2a$ or $2a{+}1$. Correspondingly, $\lfloor 2x\rfloor - 2\lfloor x\rfloor$ will be $2a - 2a = 0$ or $2a + 1 - 2a = 1$.

Practice Problems

7.1. *No solution provided. Assignment problem.*

7.2. There are x/g intervals of length g between 0 and x. Each of these must contain at least one prime by assumption. So $\pi(x) \geq x/g$. Consequently,

$$\frac{\pi(x)}{(x/\log x)} \geq \frac{\log x}{g}.$$

As $x \to \infty$, the right side goes to ∞ but the left-side goes to 1 by the Prime Number Theorem. Contradiction!

7.3. The PNT implies that $p_n \sim n \log n$, meaning

$$\lim_{n\to\infty} \frac{p_n}{n \log n} = 1$$

and likewise

$$\lim_{n\to\infty} \frac{(n+1)\log(n+1)}{p_{n+1}} = 1.$$

Thus,

$$\begin{aligned}
\lim_{n\to\infty} \frac{p_{n+1}}{p_n} &= \lim_{n\to\infty} \frac{p_{n+1}}{p_n} \frac{(n+1)\log(n+1)}{(n+1)\log(n+1)} \frac{n\log n}{n\log n} \\
&= \lim_{n\to\infty} \underbrace{\frac{p_{n+1}}{(n+1)\log(n+1)}}_{\to 1} \underbrace{\frac{n\log n}{p_n}}_{\to 1} \underbrace{\frac{(n+1)\log(n+1)}{n\log n}}_{\to 1} \\
&= 1.
\end{aligned}$$

7.4. When $x = 1$, we find that $y = 1$, so $(x, y) = (1, 1)$ is a solution. By inspection we can easily see that there are no solutions when $x = 2,\ 3$ or 4.

So assume that $x > 4$. By Bertrand's postulate, there is a prime p such that $x/2 < p < x$, so $p \mid x!$. Observe that $p^2 > x^2/4$ and $x^2/4 > x$ if $x > 4$. So $p^2 \nmid x!$. It follows that $x!$ cannot be a perfect square since $v_p(x!) = 1$ is odd. Thus there are no solutions to $x! = y^2$ when $x > 4$.

Lecture 8

Exercises

8.3. $n \mid a-a$ is true since all integers divide 0. If $n \mid a-b$ then $n \mid -(a-b) = b-a$. Finally, if $n \mid b-a$ and $n \mid c-b$ then $n \mid (c-b)-(b-a) = c-a$.

8.9. One possible set is $\{0, 6, 2, 8, 4\}$. Another is $\{10, 26, -8, 18, -6\}$. There are infinitely many possibilities here.

8.11. We want to prove that $[a+b]_n = [c+d]_n$, or equivalently, that $a+b \equiv c+d \pmod{n}$. However, since $a \equiv c \pmod{n}$ and $b \equiv d \pmod{n}$, this is precisely what Proposition 8.10 guarantees.

Practice Problems

8.1. By the remainder theorem, we have $a = nq + r$ and $b = nq' + r'$ where r and r' are the remainders of a and b modulo n, resp. If $a \equiv b \pmod{n}$ then by definition $r = r'$ and therefore $a - b = n(q - q')$ is divisible by n. Conversely, if $n \mid a - b$, then $n \mid n(q - q') + (r - r')$ and hence $n \mid r - r'$. Since $r - r'$ is strictly between $-n$ and n, it follows that $r - r' = 0$. So $r = r'$ and therefore $a \equiv b \pmod{n}$.

8.2. Week days cycle every 7 days. Since $365 \equiv 1 \pmod{7}$, that means the day we want will be one day away from today. So it's also a Tuesday.

8.3. *No solution provided. Assignment problem.*

8.4. This is impossible. We've already seen that perfect squares are congruent to either 0 or 1 modulo 4, so there can be no perfect square that represents the congruence class $[2]_4$ for example.

Lecture 9

Exercises

9.2. Observe that $n! \equiv 0 \pmod 5$ for all $n \geq 5$. So

$$1! + 2! + 3! + \cdots + 100! \equiv 1! + 2! + 3! + 4! \equiv 33 \equiv 3 \pmod 5.$$

9.4. What we *can* do is plug in $0, 1, \ldots, 29$ for a and confirm that $a^5 - a \equiv 0 \pmod{30}$. However, this is tedious!

Here is a better approach. First observe that

$$a^5 - a = a(a^4 - 1) = a(a^2 - 1)(a^2 + 1) = (a^3 - n)(a^2 + 1).$$

We proved that $6 \mid a^3 - a$ in example preceding this exercise. So $6 \mid a^5 - a$. It therefore suffices to prove that $5 \mid a^5 - a$, since then $30 = 5 \cdot 6$ will divide $a^5 - a$ by Proposition 4.3(b). For this, we can plug in $a = 0, 1, 2, 3, 4$ and reduce modulo 5. However, it will be slightly quicker to instead use the representatives $0, \pm 1, \pm 2$ modulo 5. We have

$$\begin{aligned}
&0^5 - 0 \equiv 0 \pmod 5 \\
&1^5 - 1 \equiv 0 \pmod 5 \\
&2^5 - 2 \equiv 30 \equiv 0 \pmod 5 \\
&3^5 - 3 \equiv (-2)^5 - (-2) \equiv -(2^5 - 2) \equiv 0 \pmod 5 && \text{(by the } a = 2 \text{ calculation)} \\
&4^5 - 4 \equiv (-1)^5 - (-1) \equiv -(1^5 - 1) \equiv 0 \pmod 5. && \text{(by the } a = 1 \text{ calculation)}
\end{aligned}$$

This completes the proof.

9.6. Repeat the same proof given in the example preceding this exercise, except now $10^k \equiv (-1)^k \pmod{11}$.

Practice Problems

9.1. Suppose that $f(x) = c_0 + c_1 x + \cdots + c_m x^m$, with $c_i \in \mathbb{Z}$.

If $a \equiv b \pmod n$ then $a^k \equiv b^k \pmod n$ for all $k \in \mathbb{Z}_{\geq 0}$ (by repeatedly using the fact that $a \equiv b$ and $a \equiv b$ implies $a \cdot a \equiv b \cdot b$). Similarly, $c_i a^k \equiv c_i b^k \pmod m$ for all i. Combining both of these observations, we get $f(a) \equiv f(b) \pmod m$.

9.2. This is false. Counterexample: Consider $n = 6$. Then $[2] \neq [0]$ and $[3] \neq [0]$ but $[2][3] = [6] = [0]$.

9.3. Using the binomial theorem and the fact that $p \mid \binom{p}{k}$ for $0 < k < p$, we have

$$\begin{aligned}
(a+b)^p &= a^p + \binom{p}{1} a^{p-1} b + \binom{p}{2} a^{p-2} b + \cdots + \binom{p}{p-1} a b^{p-1} + b^p \\
&\equiv a^p + 0 + \cdots + 0 + b^p \pmod p,
\end{aligned}$$

as desired.

9.4. Any k consecutive integers will reduce to give, after re-ordering if necessary, the k remainders $0, 1, \ldots, k-1$ modulo k. So the product of k consecutive integers is congruent to

$$0 \cdot 1 \cdots (k-1) \equiv 0 \pmod{k}$$

hence is divisible by k.

For the stronger result, we can observe that

$$\frac{a(a+1)\cdots(a+k-1)}{k!} = \frac{(a+k-1)!}{(a-1)!k!} = \binom{a+k-1}{k}$$

is a binomial coefficient if $a \geq 0$, so is an integer. It follows that $k! \mid a(a+1)\cdots(a+k-1)$. If $a < 0 \leq a+k-1$, then the product is zero, so is again an integer. Finally, if the k consecutive integers are all negative, then their product is $(-1)^k$ times the product of k consecutive positive integers, which is an integer by what we have already proved.

I do not know of a proof of the stronger result that uses simple modular arithmetic arguments. Can you find one?

9.5. *No solution provided. Assignment problem.*

9.6. *No solution provided. Assignment problem.*

9.7. (a) We saw in Exercise 9.2 that

$$1! + 2! + \cdots + 100! \equiv 3 \pmod{5}.$$

But squares can only be congruent to 0, 1 or 4 modulo 5.

(b) The same argument as in Exercise 9.2 (reduce modulo 5) shows that if $x \geq 4$ then

$$1! + 2! + \cdots + x! \equiv 3 \pmod{5}$$

and hence can never be a perfect square. So we only need to consider the cases where $x < 4$:

- $x = 1$: $1! = 1$ is a perfect square. This gives the solution $(x, y) = (1, 1)$.
- $x = 2$: $1! + 2! = 3$ is not a perfect square.
- $x = 3$: $1!+2!+3! = 9$ is a perfect square. This gives the solution $(x, y) = (3, 3)$.

So $(x, y) = (1, 1)$ and $(3, 3)$ are the only solutions with $x, y \in \mathbb{Z}_{>0}$.

Lecture 10

Exercises

10.4. First, suppose that $0 \leq k, k' \leq g-1$ and that

$$x_0 + k\frac{n}{g} \equiv x_0 + k'\frac{n}{g} \pmod{n}.$$

This implies that

$$(k-k')\frac{n}{g} \equiv 0 \pmod{n}$$

and hence

$$(k-k')\frac{n}{g} = tn \quad \text{for some } t \in \mathbb{Z}.$$

But then $k - k' = tg$, which implies that $g \mid k - k'$. However, since $|k - k'| < g$, it follows that $k = k'$.

Next, given $l \in \mathbb{Z}$, apply the remainder theorem to write $l = qg + r$, where $0 \leq r \leq g-1$. Then

$$x_0 + l\frac{n}{g} = x_0 + (qg+r)\frac{n}{g} = x_0 + qn + r\frac{n}{g} \equiv x_0 + r\frac{n}{g} \pmod{n}.$$

This shows that $[x_0 + l(n/g)] = [x_0 + r(n/g)]$ with $0 \leq r \leq g-1$.

10.6. Since $g = \gcd(15, 35) = 5$ divides 25, there are solutions to this congruence. It's not immediately obvious (to me) what a particular solution is, so let's run the Euclidean algorithm on $15x - 35y = 25$:

$$35 = 2 \cdot 15 + 5$$
$$15 = 3 \cdot 5 + 0.$$

So $5 = 15 \cdot (-2) + 35$ and therefore, by multiplying this through by 5, we get

$$25 = 15 \cdot (-10) + 35 \cdot 5.$$

Hence $(x, y) = (-10, 5)$ is a particular solution to $15x - 35y = 25$, and therefore $x \equiv -10$ is a particular solution to $15x \equiv 25 \pmod{35}$. Consequently, the full solution set to this congruence modulo 35 is given by

$$\left\{ \left[-10 + k\frac{35}{5} \right] : 0 \leq k \leq 4 \right\} = \{[-10],\ [-3],\ [4],\ [11],\ [18]\}.$$

10.10. Looking at the solution set in the previous exercise, we see that all congruence classes modulo 35 reduce to just $[4]_7$ modulo 7.

10.13. The table is:

$[x]$	$[0]$	$[1]$	$[2]$	$[3]$	$[4]$	$[5]$	$[6]$	$[7]$	$[8]$	$[9]$	$[10]$	$[11]$
$[x]^{-1}$	–	$[1]$	–	–	–	$[5]$	–	$[7]$	–	–	–	$[11]$

Practice Problems

10.1. Since $[a]_n = [b]_n$, we have $a = b + kn$ for some $k \in \mathbb{Z}$, and therefore

$$\gcd(a, n) = \gcd(b + kn, n) = \gcd(b, n)$$

where the last step follows from Lemma 3.3. So, $\gcd(a, n) = 1$ if and only if $\gcd(b, n) = 1$.

10.2. For $[2]^{-1}$ to exist in $\mathbb{Z}/n\mathbb{Z}$, we need $\gcd(2, n) = 1$, that is, we need n to be odd.

So assume now that $n = 2k + 1$ is odd. To find the inverse of 2 modulo n, we need to solve the congruence $2x \equiv 1 \pmod{n}$. This is equivalent to solving the Diophantine equation

$$2x - ny = 1.$$

Since $n = 2k + 1$, a particular solution is given by $(x, y) = (k + 1, 1)$. Thus,

$$[2]^{-1} = [k + 1] = \left[\frac{n-1}{2} + 1\right] = \left[\frac{n+1}{2}\right].$$

[**Sanity checks:** $\dfrac{n+1}{2}$ is an integer since n is odd. Also, $2 \cdot \dfrac{n+1}{2} = n + 1 \equiv 1 \pmod{n}$.]

Alternative solution: From $n = 2k + 1$ we find that $2k + 1 \equiv 0 \pmod{n}$. Hence $2k + 2 \equiv 1 \pmod{n}$ and therefore

$$2(k + 1) \equiv 1 \pmod{n}.$$

So $[2]^{-1} = [k + 1] = \left[\frac{n+1}{2}\right]$.

10.3. (a) If $x \equiv a_1 \pmod{n_1}$ then we can write $x = a_1 + n_1 y$ for some $y \in \mathbb{Z}$. Therefore, the congruence $x \equiv a_2 \pmod{n_2}$ becomes

$$a_1 + n_1 y \equiv a_2 \pmod{n_2}$$

or equivalently

$$n_1 y \equiv a_2 - a_1 \pmod{n_2}.$$

This is a linear congruence in y. Since $\gcd(n_1, n_2) = 1$, this congruence has a unique solution modulo n_2, say given by $y \equiv y_0 \pmod{n_2}$. That is, $y = y_0 + n_2 z$ for some $z \in \mathbb{Z}$.

Substituting this back into $x = a_1 + n_1 y$, we obtain

$$x = a_1 + n_1(y_0 + n_2 z) = a_1 + n_1 y_0 + n_1 n_2 z_0.$$

Note that

$$x \equiv a_1 + n_1 y_0 \pmod{n_1 n_2}.$$

This shows that there is a solution to the pair of congruences modulo $n_1 n_2$. To prove uniqueness, we can argue as follows. If we have two solutions

$$x \equiv r \pmod{n_1 n_2} \quad \text{and} \quad x \equiv r' \pmod{n_1 n_2}$$

to the pair of congruences modulo n_1 and n_2, then by reducing these solutions modulo n_1 we must have

$$r \equiv r' \pmod{n_1}$$

since both must be congruent to a_1 modulo n_1. Likewise, we obtain

$$r \equiv r' \pmod{n_2}.$$

Thus,

$$n_1 \mid (r - r') \quad \text{and} \quad n_2 \mid (r - r').$$

Since n_1 and n_2 are coprime, it follows that

$$n_1 n_2 \mid (r - r')$$

and therefore $r \equiv r' \pmod{n_1 n_2}$, which is what we wanted to prove.

(b) By part (a), the congruences modulo n_1 and n_2 have a unique solution modulo $n_1 n_2$. Then this congruence paired with the congruence modulo n_3 has a unique solution modulo $n_1 n_2 n_3$. Continuing in this manner, we obtain a unique solution modulo $n_1 n_2 \cdots n_k$.

10.4. We run through the proof of the CRT presented in the previous problem.

Starting from $x \equiv 2 \pmod 3$, we write this as $x = 2 + 3y$ with $y \in \mathbb{Z}$. Then we plug this into the second congruence, obtaining

$$2 + 3y \equiv 3 \pmod 5 \iff 3y \equiv 1 \pmod 5.$$

This has the unique solution $y \equiv 2 \pmod 5$. Writing this as $y = 2 + 5z$ and substituting it into $x = 2 + 3y$, we obtain $x = 8 + 15z$. If we plug this into the third congruence, we obtain

$$8 + 15z \equiv 2 \pmod 7 \iff z \equiv 1 \pmod 7.$$

Thus, $z = 1 + 7w$ and therefore $x = 8 + 15z = 8 + 15(1 + 7w) = 23 + 105w$.

To summarize: The solutions to the system of linear congruences are given by all x such that

$$x \equiv 23 \pmod{105}.$$

Lecture 11

Exercises

11.1. We want to solve $7x \equiv 1 \pmod{11}$. We can either do this by inspection or by using the Euclidean algorithm. Let's do it by inspection, but also using a trick.

If we try $x = 1, 2, \ldots$ one by one, we'll find when $x = 3$ that $7x = 21 \equiv -1 \pmod{11}$. This means that if $x = -3$ then $7x \equiv -(-1) \pmod{11}$. So $-3 \equiv 8$ is our desired inverse modulo 11.

11.5. The identity element in $\mathbb{Z}$ under multiplication would have to be 1 but then there are integers without inverses (e.g. 0 or 2). Likewise, the identity element in $\mathbb{Z}/n\mathbb{Z}$ would have to be $[1]$ but then there are congruence classes without inverses (e.g. $[0]_n$).

For $(\mathbb{Z}/n\mathbb{Z})^\times$ with addition, there is no identity element since $[x] + [e] = [x]$ forces $[e]$ to be $[0]$ which is not in $(\mathbb{Z}/n\mathbb{Z})^\times$.

Practice Problems

11.1. *No solution provided.* (However, if you've taken a course in linear algebra, then you should recognize that a vector space is an example of a group with $\star$ being vector addition. Scalar multiplication is "extra structure" on top of the group structure.)

11.2. (a) Since the product of two invertible matrices is invertible, we have that $A \star B \in G$ whenever $A, B \in G$. The identity element is the identity matrix $I = \begin{bmatrix} [1] & [0] \\ [0] & [1] \end{bmatrix}$. The inverse of A is the matrix inverse A^{-1}. Finally, the associative law is known to hold for matrix multiplication. Thus, G is a group under $\star$.

(b) There are a few ways we can proceed. The first way is to list all $3^4 = 81$ matrices with entries in $\mathbb{Z}/3\mathbb{Z}$ and check which are invertible. This is awful but can be done if we're desperate enough.

So let's use the hint! To construct an invertible 2×2 matrix, we just two linearly independent vectors with entries in $\mathbb{Z}/3\mathbb{Z}$. We can pick any non-zero vector $\begin{bmatrix} [a] \\ [b] \end{bmatrix}$ and then any vector that isn't a scalar multiple of this. There are 3^2 vectors with entries in $\mathbb{Z}/3\mathbb{Z}$. So we have $3^2 - 1$ choices for the first vector (since anything other that $\begin{bmatrix} [0] \\ [0] \end{bmatrix}$ is OK). Once we've picked the first vector, it will have 3 scalar multiples (since the scalars are coming from $\mathbb{Z}/3\mathbb{Z}$), leaving us with $3^2 - 3$ choices for the second vector. So in total, there are $(3^2 - 1)(3^2 - 3) = 48$ matrices in G.

11.3. We basically proved this in the course of proving Lagrange's theorem! If we can show that the congruence classes $[ua_1], \ldots, [ua_n]$ are all distinct, then since there are n of them, they must be all of the congruence classes in $\mathbb{Z}/n\mathbb{Z}$. Now simply note that if $[ua_i] = [ua_j]$ then

$$[u][a_i] = [u][a_j]$$

and so by multiplying through by $[u]^{-1}$ we find that $[a_i] = [a_j]$ and hence $i = j$ since S is a complete set of representatives. So the $[ua_i]$ are all distinct, as desired.

11.4. By Lagrange's theorem, $g^{|G|} = e$. Hence $g \star g^{|G|-1} = g^{|G|} = e$, so $g^{|G|-1}$ must be the inverse of g (by uniqueness).

11.5. (a) Note that $[a] \in (\mathbb{Z}/p\mathbb{Z})^\times$ if and only if a and p are coprime. Since all of the integers $1 \leq a \leq p-1$ are coprime to p, we have that

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

hence $|(\mathbb{Z}/p\mathbb{Z})^\times| = p - 1$.

(b) Note that $[a] \in (\mathbb{Z}/p^2\mathbb{Z})^\times$ if and only if a and p^2 are coprime. Which of the integers a in the interval $1 \leq a \leq p^2 - 1$ are coprime to p^2? It'll be easier to determine those that aren't coprime to p^2. These are precisely the multiples of p that are $< p^2$—namely: p, $2p$, ..., $(p-1)p^2$. There are $p-1$ such multiples, and so there are

$$(p^2 - 1) - (p - 1) = p^2 - p$$

integers in the interval $1 \leq a \leq p^2 - 1$ that are coprime to p^2. Thus,

$$|(\mathbb{Z}/p^2\mathbb{Z})^\times| = p^2 - p.$$

(c) Note that $[a] \in (\mathbb{Z}/(pq)\mathbb{Z})^\times$ if and only if a and pq are coprime. As in part (b), we'll begin determining by determining the integers a in the interval $1 \leq a \leq pq - 1$ that aren't coprime to pq. These are precisely the integers that are multiples of p or q. The multiples of p are p, $2p$, ..., $(q-1)p$ — so there are $q-1$ of them. Likewise, there are $p-1$ multiples of q. So that leaves us with

$$(pq - 1) - (p - 1) - (q - 1) = pq - p - q - 1 = (p - 1)(q - 1)$$

integers in the interval $1 \leq a \leq pq - 1$ that are coprime to pq. Thus,

$$|(\mathbb{Z}/(pq)\mathbb{Z})^\times| = (p - 1)(q - 1).$$

Lecture 12

Exercises

12.6. Suppose $a \in \mathbb{Z}$ is negative. Since we know the theorem holds for $-a > 0$, we have $(-a)^p \equiv (-a) \pmod{p}$. If p is odd then this becomes $(-1)a^p \equiv (-1)a$, and we're done since we can divide both sides by the unit -1 modulo p. If p is even then $p = 2$ and so $-1 \equiv 1$, so $-a \equiv a$, and we're done again.

12.10. We have

$$\begin{aligned}
173 &= 2 \cdot 86 + 1 \\
&= 2(2 \cdot 43 + 0) + 1 \\
&= 2^2 \cdot 43 + 2 \cdot 0 + 1 \\
&= 2^2(2 \cdot 21 + 1) + 2 \cdot 0 + 1 \\
&= 2^3 \cdot 21 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^3 \cdot (2 \cdot 20 + 1) + 2^2 + 2 + 1 \\
&= 2^4 \cdot 20 + 2^3 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^4(2 \cdot 10 + 0) + 2^2 + 2 \cdot 0 + 1 \\
&= 2^5 \cdot 10 + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^5(2 \cdot 5 + 0) + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^6 \cdot 5 + 2^5 \cdot 0 + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^6 \cdot (2 \cdot 2 + 1) + 2^5 \cdot 0 + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1 \\
&= 2^7 + 2^6 + 2^5 \cdot 0 + 2^4 \cdot 0 + 2^2 + 2 \cdot 0 + 1
\end{aligned}$$

12.1. Since 13 is prime, $3^{12} \equiv 1 \pmod{13}$ by Fermat. Since $155 = 12 \cdot 12 + 11$, we have

$$3^{155} = (3^{12})^{12} 3^{11} \equiv 1^{12} 3^{11} \pmod{13}.$$

In binary form, $11 = 2^3 + 2 + 1$. So $3^{11} = 3^{2^3} 3^2 3$. Now,

$$3^2 = 9$$

hence

$$3^{2^2} = (3^2)^2 = 9^2 \equiv 3 \pmod{13}$$

hence

$$3^{2^3} = (3^{2^2})^2 \equiv 3^2 \equiv 9 \pmod{13}.$$

So, putting all this together, we arrive at

$$3^{155} \equiv 3^{11} \equiv 3^{2^3} 3^2 3 \equiv 9 \cdot 9 \cdot 3 \equiv 3^4 \cdot 3 \equiv 3 \cdot 3 \equiv 9 \pmod{13}.$$

Alternative Solution: Since $3^{12} \equiv 1 \pmod{13}$, we have

$$3^{11} \equiv 3^{-1} \pmod{13}$$

where 3^{-1} is the inverse of 3 modulo 13. Since $3 \cdot 9 = 27 \equiv 1 \pmod{13}$, it follows that $[3]^{-1} = [9]$. Thus,

$$3^{11} \equiv 9 \pmod{13}.$$

Now we conclude as in the above solution:

$$3^{155} = (3^{12})^{12} 3^{11} \equiv 1^{12} \cdot 3^{11} \equiv 9 \pmod{13}.$$

Practice Problems

12.1. This can be done fairly easily by repeated squaring. However, here is a shortcut.

Proving that $2^{340} \equiv 1 \pmod{341}$ means proving that $341 \mid 2^{340} - 1$. Since the prime factorization of 341 is $341 = 11 \times 31$, it suffices to prove that $11 \mid 2^{340} - 1$ and $31 \mid 2^{340} - 1$. That is, it suffices to prove that

$$2^{340} \equiv 1 \pmod{11} \quad \text{and} \quad 2^{340} \equiv 1 \pmod{31}.$$

The first of these an immediate consequence of Fermat's Little Theorem, since $2^{10} \equiv 1 \pmod{11}$. For the second one, observe that $340 = 30 \times 11 + 10$. Since $2^{30} \equiv 1 \pmod{31}$ by Fermat again, we have

$$2^{340} = (2^{30})^{11} 2^{10} \equiv 1^{11} 2^{10} \pmod{31}.$$

Now it's just a matter of proving that $2^{10} \equiv 1 \pmod{31}$, and this can be done quickly via repeated squaring.

12.2. If $k \equiv l \pmod{\varphi(n)}$ then we can write $k = l + m\varphi(n)$ for some $m \in \mathbb{Z}$. Therefore,

$$a^k = a^{l+m\varphi(n)} = a^l (\underbrace{a^{\varphi(n)}}_{\equiv 1})^m \equiv a^l \pmod{n},$$

where Euler's theorem was used in the last step.

12.3. This is a special case of Problem 4 in Lecture 11.

12.4. This is false. For a counterexample, take $a = 2$ and $n = 4$. Then $\varphi(n) = \varphi(4) = 2$ (why?) and

$$2^{\varphi(4)+1} = 8 \not\equiv 2 \pmod{4}.$$

It's an interesting exercise to determine conditions on a and n that make this result true. It's true if a and n are coprime, but can we do better?

Lecture 13

Exercises

13.2. *Answer:* `TVKBSHY HYPAOTLAPJ`.

13.4. *Key:* $k = 8$. *Plaintext:* `NUMBER THEORY IS FUN`.

13.6. (a) *Answer:* `RWHO TZS W`.

(b) *Answer:* `IS INVERTIBLE`. (The decryption key is $(k_1^{-1}, k_2) = (9, 9)$.)

13.8. The shift cipher satisfies requirements 1 and 2 but not 3 or 4. It fails requirement 3 because brute-force attacks and frequency attacks can easily determine the key k. Since it fails 3, it also fails 4. But it fails 4 in a more extreme way! Knowledge of a single pair (m, c) allows us to determine the key since

$$c \equiv m + k \pmod{26} \implies k = m - c \pmod{26}.$$

The affine cipher satisfies requirements 1 and 2 but fails 3 for the same reasons. It fails 4 since it fails 3, but it's an interesting exercise to analyze requirement 4 separately in this case. Assuming $d_k(c)$ is difficult to compute without knowing k, is the affine cipher resistant to known-plaintext attacks?

13.10. (a) *Answer:* 205028 64933 214281 242631.

(b) *Answer:* `BEZOUT LEMMA`. The decryption function uses $d \equiv 143279 \pmod{p}$.

Practice Problems

13.1. *No solution provided. Assignment problem.*

13.2. This cipher is not secure against known-plaintext attacks. All it takes is for Eve to obtain a single plaintext-ciphertext pair $(m, c) \neq (0, 0)$ and she can break the cipher.

The mathematical problem is as follows. Eve knows c, m and p and wants to solve the equation $c \equiv em \pmod{p}$ for e. (Once she has e she can determine d via the Euclidean algorithm.) If Eve knows a pair $(m, c) \neq (0, 0)$, then since m will be a unit mod p, Eve can determine $e \equiv cm^{-1} \pmod{p}$. She can find m^{-1} mod p using the Euclidean algorithm.

13.3. (a) Calculating powers of 3 modulo 7, we find:

e	1	2	3	4	5	6
$3^e \pmod 7$	3	2	6	4	5	1

Thus, as e runs from 1 to 6, 3^e runs over all congruence classes in $(\mathbb{Z}/7\mathbb{Z})^\times$. So, given any $c \in (\mathbb{Z}/7\mathbb{Z})^\times$, we can find an e such that $3^e \equiv c \pmod{7}$.

(b) If $3^e \equiv 7 \equiv 3^{e'} \pmod{7}$, we get

$$1 \equiv 3^e (3^{e'})^{-1} \equiv 3^{e-e'} \pmod{7}.$$

If $e - e' \equiv r \pmod{6}$, then $3^{e-e'} \equiv 3^r \pmod{7}$ (by Problem 12.2). From our table in part (a), we see that the only r mod 6 for which $3^r \equiv 1 \pmod{7}$ is $r \equiv 6 \equiv 0 \pmod{6}$. It follows that $e - e' \equiv 0 \pmod{6}$ hence $e \equiv e' \pmod{6}$.

(c) (i) If $3^e \equiv 1 \pmod 7$, then since $3^0 \equiv 1 \pmod 7$, we have from part (b) that $e \equiv 0 \pmod 6$. Thus, $\log_3(1) \equiv 1$.

(ii) If $3^e \equiv ab \pmod 7$, $3^{e_1} \equiv a \pmod 7$ and $3^{e_2} \equiv b \pmod 7$, then

$$3^{e_1+e_2} = 3^{e_1}3^{e_2} \equiv ab \equiv 3^e \pmod 7.$$

It follows, by part (b), that $e \equiv e_1 + e_2 \pmod 6$.

Lecture 14

Exercises

14.2. (a) $n = 6319 = 71 \times 89$ so $\varphi(n) = 70 \times 88 = 6160$. The private d is the inverse of 3 mod 6160. Solving $3d \equiv 1 \pmod{6160}$, we find that $d = 4107$.

(b) `ZETA` is 2504 1925. This gets encrypted into 2173 and 4914.

(c) Notice that 2504 is as in part (b), so we know it decrypts to `ZE`. (This type of behaviour is *bad*! It makes the cryptosystem vulnerable to attacks. When RSA is used in practice, messages get padded with random bits in order to avoid such "coincidences.") The second half 5047 gets decrypted into 1714, which is `RO`. Thus, the original message was `ZERO`.

14.3. p and q will be roots of the quadratic equation

$$x^2 - 1038x + 239777 = 0.$$

Using the quadratic formula, we find that

$$x = \frac{1038 \pm \sqrt{1038^2 - 4 \cdot 239777}}{2} = \frac{1038 \pm 344}{2}.$$

Thus, $p = 691$ and $q = 347$.

Practice Problems

14.1. (a) $A^b \equiv (g^a)^b = (g^b)^a \equiv B^a \pmod{p}$.

(b) If Eve can solve the DLP then she can find a from $A \equiv g^a \pmod{p}$ and hence she can determine the key K by computing $B^a \pmod{p}$.

14.2. (a) Since $ed \equiv 1 \pmod{\varphi(pq)}$, and since $\varphi(pq) = (p-1)(q-1)$, we can write $ed = 1 + k(p-1)(q-1)$ for some $k \in \mathbb{Z}$. Thus,

$$m^{ed} = m(m^{p-1})^{k(q-1)}.$$

If $p \nmid m$ then using Fermat's Little Theorem we get

$$m^{ed} \equiv m(1)^{k(q-1)} \equiv m \pmod{p}.$$

If $p \mid m$ we again get $m^{ed} \equiv m \pmod{p}$ since both sides are congruent to 0. Thus, $m^{ed} \equiv m \pmod{p}$ for all m. The exact same proof works mod q.

(b) By part (a), $x = m^{ed}$ is a solution to the system

$$\begin{aligned} x &\equiv m \pmod{p} \\ x &\equiv m \pmod{q} \end{aligned}$$

Since $x = m$ is also a solution, it follows from the uniqueness assertion in the Chinese Remainder Theorem that

$$m^{ed} \equiv m \pmod{pq}.$$

14.3. *No solution provided. Assignment problem.*

Lecture 15

Exercises

15.3. The answer, for each, is neither! Here's why:

- $v_p(n)$: If p and q are distinct primes then $v_p(pq) = 1$ while $v_p(p)v_p(q) = 1 \cdot 0 = 0$.
- $\pi(n)$: For example $\pi(1) = 0$ and $\pi(2) = 1$ so $\pi(2) = \pi(2 \cdot 1) \neq \pi(2)\pi(1)$.
- $\omega(n)$: For example, $\omega(6) = 2$ while $\omega(2)\omega(3) = 1$.
- $\Omega(n)$: Same counterexample as for $\omega(n)$.
- $r_2(n)$: For example, $r_2(1) = 4$ so $r_2(1 \cdot 1) \neq r_2(1)r_2(1)$.

15.6. No. For example, $\sigma_k(4) = 1^k + 2^k + 4^k$ while $\sigma_k(2) = 1^k + 2^k$ so $\sigma_k(2)\sigma_k(2) = 1^k + 2^{k+1} + 4^k \neq \sigma_k(2^2)$.

Practice Problems

15.1. First off, $f(1) = f(p^0) = g(p^0) = g(1)$. Now assume $n > 1$ and that $n = \prod_i p_i^{a_i}$ is the prime factorization of n. Then since f is multiplicative we have

$$f(n) = \prod_i f(p_i^{a_i}) = \prod_i g(p_i^{a_i}).$$

This product is also equal to $g(n)$ since g is multiplicative. Thus, $f(n) = g(n)$.

15.2. We have $f(1) = f(1 \cdot 1) = f(1)f(1)$. So either $f(1) = 1$ or else $f(1) = 0$. If $f(1) = 0$ then $f(n) = f(n \cdot 1) = f(n)f(1) = 0$ for all n.

15.3. *No solution provided. Assignment problem.*

15.4. (a) Assume that $n, m \in \mathbb{Z}_{>0}$ are coprime. If either is even, then nm is even and $\chi(nm) = \chi(n)\chi(m) = 0$. So we may assume that both n and m are odd. If they are both 1 mod 4 or both 3 mod 4, then nm is 1 mod 4, so in either case we get $\chi(nm) = 1$ and $\chi(n)\chi(m) = 1$. If one is 1 mod 4 and the other is 3 mod 4, then nm is 3 mod 4, and in this case we get $\chi(nm) = -1$ and $\chi(n)\chi(m) = -1$. So $\chi(nm) = \chi(n)\chi(m)$ in all cases.

(b) Since X is the summatory function of χ, and since χ is (completely) multiplicative, we know that X must be multiplicative. On the other hand, X is not completely multiplicative. For instance, $X(3) = 0$ while $X(3^2) = 1$ so $X(3^2) \neq X(3)^2$.

(c) To compute $r_2(n)$, we must count the solutions to $x^2 + y^2 = n$ where $x, y \in \mathbb{Z}$.

- If $n = 1$, the only solutions are $(x, y) = (\pm 1, 0)$ and $(0, \pm 1)$. So $r_2(1) = 4$.
  On the other hand, $X(1) = \chi(1) = 1$. So $r_2(1) = 4X(1)$.
- If $n = 2$, the only solutions are $(x, y) = \pm(1, 1)$ and $\pm(1, -1)$. So $r_2(2) = 4$.
  On the other hand, $X(2) = \chi(1) + \chi(2) = 1$. So $r_2(2) = 4X(2)$.
- If $n = 3$, there are no solutions. So $r_2(3) = 0$.
  On the other hand, $X(3) = \chi(1) + \chi(3) = 1 - 1 = 0$. So $r_2(3) = 4X(3)$.
- If $n = 4$, the only solutions are $(x, y) = (\pm 2, 0)$ and $(x, y) = (0, \pm 2)$. So $r_2(4) = 4$.
  On the other hand, $X(4) = \chi(1) + \chi(2) + \chi(4) = 1 + 0 + 0$. So $r_2(4) = 4X(4)$.
- If $n = 5$, then the only solutions are $(x, y) = (\pm 2, \pm 1), (\pm 1, \pm 2)$. So $r_2(5) = 8$.
  On the other hand, $X(5) = \chi(1) + \chi(5) = 1 + 1 = 2$. So $r_2(5) = 4X(5)$.

I'll let yo do the rest!

15.5. First of all, there are the constant functions $f(n) = 1$ and $f(n) = 0$. These are the only possible constant functions (why?) so let's assume that f is non-constant. In particular, $f(1) = 1$ (since otherwise $f = 0$ is constant). Once we determine the possible values $f(2)$ and $f(3)$, we will have determined f completely. Note that $f(2)f(3) = f(6) = f(3)$ since $6 \equiv 3 \pmod 3$. If $f(3) \neq 0$ then $f(2) = 1$. This gives the function

$$f(n) = \begin{cases} a & \text{if } n \equiv 0 \pmod 3 \\ 1 & \text{otherwise} \end{cases}$$

where a is an arbitrary non-zero number. It's easy to check that f is multiplicative.

On the other hand, if $f(3) = 0$, then $f(2)$ can potentially be arbitrary. This gives the function

$$f(n) = \begin{cases} 0 & \text{if } n \equiv 0 \pmod 3 \\ 1 & \text{if } n \equiv 1 \pmod 3 \\ b & \text{if } n \equiv 2 \pmod 3 \end{cases}$$

where a is an arbitrary number. For f to be multiplicative, we need to check that if $n, m \in \mathbb{Z}_{>0}$ are coprime then $f(nm) = f(n)f(m)$. This is clear if either of n or m is 0 mod 3, if both are 1 mod 3, or if one is 1 mod 3 and the other is 2 mod 3. However, if both are 2 mod 3, then $nm = 1 \pmod 3$, so we need to have $b^2 = f(n)f(m) = f(nm) = 1$. Thus, we must have that $b = \pm 1$.

Lecture 16

Exercises

16.2. (a) Here is the table of values:

n	1	2	3	4	5	6	7	8	9	10
$\mu(n)$	1	-1	-1	0	-1	1	-1	0	0	1

(b) Note that $\mu(2^2) = 0 \neq 1 = \mu(2)\mu(2)$. So μ is not completely multiplicative.

Let's now prove that $\mu(n)$ is multiplicative. Suppose that $n, m \in \mathbb{Z}_{>0}$ are coprime. We wish to prove that $\mu(nm) = \mu(n)\mu(m)$. We may assume that $n, m > 1$ since the result is trivial otherwise. If either n or m is divisible by p^2 for some prime p then so is nm and hence $\mu(n) = \mu(m) = \mu(nm) = 0$ and therefore $\mu(nm) = \mu(n)\mu(m)$. So all that remains is the case where $n = p_1 \cdots p_k$ and $m = q_1 \cdots q_l$ are products of distinct primes. Note that since n and m are coprime then $p_i \neq q_j$ for any i, j. Thus, $nm = p_1 \cdots p_k q_1 \cdots q_l$ is a product of distinct primes. So the definition of μ gives

$$\mu(nm) = (-1)^{k+l} = (-1)^k(-1)^l = \mu(n)\mu(m),$$

as desired.

16.9. The prime factorization of 2024 is $2024 = 2^3 \cdot 11 \cdot 23$. So

$$\begin{aligned}\varphi(2024) &= \varphi(2^3 \cdot 11 \cdot 23)\\ &= (2^3 - 2^2)(11 - 1)(23 - 1)\\ &= 880.\end{aligned}$$

Practice Problems

16.1. If $n = p_1^{a_1} \cdots p_k^{a_k}$ then

$$\begin{aligned}n\prod_{p|n}\left(1 - \frac{1}{p}\right) &= p_1^{a_1} \cdots p_k^{a_k}\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_k}\right)\\ &= p_1^{a_1}\left(1 - \frac{1}{p_1}\right) \cdots p_k^{a_k}\left(1 - \frac{1}{p_k}\right)\\ &= \left(p_1^{a_1} - \frac{p_1^{a_1}}{p_1}\right) \cdots \left(p_k^{a_k} - \frac{p_k^{a_1}}{p_k}\right)\\ &= \varphi(p_1^{a_1}) \cdots \varphi(p_k^{a_k})\\ &= \varphi(n).\end{aligned}$$

16.2. (a) Using the previous problem, we have

$$\frac{\varphi(nm)}{nm} = \prod_{p|nm}\left(1 - \frac{1}{p}\right) = \frac{\prod_{p|n}\left(1 - \frac{1}{p}\right)\prod_{p|m}\left(1 - \frac{1}{p}\right)}{\prod_{p|d}\left(1 - \frac{1}{p}\right)} = \frac{\frac{\varphi(n)}{n}\frac{\varphi(m)}{m}}{\frac{\varphi(d)}{d}}.$$

Now multiply through by nm to get the desired result.

(b) We will proceed by induction on m. If $m = 1$ then $n = 1$ and there is nothing to prove. So assume the result holds for all integers $< m$. Since $n \mid n$ we can write $m = kn$ where $1 \leq k < m$. By part (a), we have $\varphi(m) = \varphi(k)\varphi(n)(d/\varphi(d))$, where $d = \gcd(k, n)$. Note this doesn't prove that $\varphi(n) \mid \varphi(m)$ since $d/\varphi(d)$ is not necessarily an integer. However, $d \mid k$ and since $k < m$ the inductive hypothesis tells us that $\varphi(d) \mid \varphi(k)$. So $l = \varphi(k)/\varphi(d)$ is an integer. Since

$$\varphi(m) = ld\varphi(n)$$

it follows that $\varphi(n) \mid \varphi(m)$, as required.

(c) If n has an odd prime divisor p, then we can write

$$\varphi(n) = (p^a - p^{a-1})m$$

for some integers $a, m \geq 1$. Note that $p^a - p^{a-1}$ is even, so $\varphi(n)$ must be even too. If n doesn't have any odd prime divisors, then $n = 2^a$ for some $a \geq 2$ (since $n \geq 3$). In this case, $\varphi(n) = 2^a - 2^{a-1} = 2^{a-1}$ is even.

16.3. (a) If $n = \prod_{i=1}^{k} p_i^{a_i}$, then the only divisors $d \mid n$ for which $\Lambda(d)$ is non-zero are those of the form $d = p_i^{b_i}$ for $1 \leq b_i \leq a_i$. Thus,

$$\begin{aligned}
\sum_{d|n} \Lambda(n) &= \sum_{i=1}^{k} \sum_{j=1}^{a_i} \Lambda(p_i^j) \\
&= \sum_{i=1}^{k} \sum_{j=1}^{a_i} \log p_i \\
&= \sum_{i=1}^{k} a_i \log p_i \\
&= \sum_{i=1}^{k} \log(p_i^{a_i}) \\
&= \log\left(\prod_{i=1}^{k} p_i^{a_i}\right) \\
&= \log n.
\end{aligned}$$

(b) This is clear if $n = 1$, so let's assume that $n > 1$. Möbius inversion gives

$$\begin{aligned}
\Lambda(n) &= \sum_{d|n} \mu(d) \log(n/d) \\
&= \sum_{d|n} \mu(d)(\log n - \log d) \\
&= \log n \sum_{d|n} \mu(d) - \sum_{d|n} \mu(d) \log d \\
&= -\sum_{d|n} \mu(d) \log d
\end{aligned}$$

since $\sum_{d|n} \mu(d) = 0$ for $n > 1$ by by Theorem 16.3.

16.4. *No solution provided.*

16.5. (a) By definition, $(u * \mu)(n) = \sum_{d|n} u(d)\mu(n/d) = \sum_{d|n} \mu(n/d) = \sum_{d|n} \mu(d)$. So $u * \mu$ is the summatory function of μ, which is e by Theorem 16.3.

(b) The computation in part (a) shows that $f * u$ is the summatory function of f, that is, $F = f \times u$. Therefore, $F * \mu = (f * u) * \mu = f * e = f$.

Lecture 17

Exercises

17.3. The powers of 4 mod 11 in order from 4^1 to 4^5 are:

$$4,\ 5,\ 9,\ 3,\ 1.$$

Here $e = 5$ is the smallest exponents where $4^e \equiv 1$. So the powers of 4 mod 11 will repeat in cycles of length 5.
Since $4^2 \equiv 5$, it follows that the solutions are all given by $x = 2 + 5k$ where $k \in \mathbb{Z}$.

17.5. Here is the table of elements in $(\mathbb{Z}/11\mathbb{Z})^\times$ and their orders:

a	values a^e	$\operatorname{ord}_{11}(a)$
1	$\{1\}$	1
2	$\{2,4,8,5,10,9,7,3,6,1\}$	10
3	$\{3,9,5,4,1\}$	5
4	$\{4,5,9,3,1\}$	5
5	$\{5,3,4,9,1\}$	5
6	$\{6,3,7,9,10,5,8,4,2,1\}$	10
7	$\{7,5,2,3,10,4,6,9,8,1\}$	10
8	$\{8,9,6,4,10,3,2,5,7,1\}$	10
9	$\{9,4,3,5,1\}$	5
10	$\{10,1\}$	2

17.9. Since $\varphi(23) = 22$, the only possibilities for $\operatorname{ord}_{23}(5)$ are 1, 2, 11 and 22. We can compute

$$5^2 \equiv 2 \quad \text{and} \quad 5^{11} \equiv 22 \pmod{23}.$$

It follows that 5^{22} must be congruent to 1 mod 23, and thus $\operatorname{ord}_{23}(5) = 22$.

17.11. From our table in the solution to Exercise 17.5, we see that the primitive roots mod 11 are all integers congruent to 2, 6, 7 or 8 mod 11.

17.13. A complete set of representatives for $(\mathbb{Z}/8\mathbb{Z})^\times$ is $\{1,3,5,7\}$. We can quickly compute that $\operatorname{ord}(3) = \operatorname{ord}(5) = \operatorname{ord}(7) = 2$ and of course $\operatorname{ord}(1) = 1$. Since $\varphi(8) = 2^3 - 2^2 = 4$, there are no primitive roots mod 8.

Practice Problems

17.1. If $\operatorname{ord}(a) = 2$ then that means $a^2 \equiv 1 \pmod 2$ but $a \not\equiv 1 \pmod p$. Since the only roots of $x^2 - 1 = (x-1)(x+1)$ mod p are ± 1, and since $a \not\equiv 1$, it follows that $a \equiv -1$.

17.2. *No solution provided. Assignment problem.*

17.3. (a) Note that

$$(ab)^{\operatorname{ord}(a)\operatorname{ord}(b)} = (a^{\operatorname{ord}(a)})^{\operatorname{ord}(b)}(b^{\operatorname{ord}(b)})^{\operatorname{ord}(a)} \equiv 1 \pmod n.$$

So $\operatorname{ord}(ab) \mid \operatorname{ord}(a)\operatorname{ord}(b)$. On the other hand,

$$1 \equiv (ab)^{\operatorname{ord}(ab)\operatorname{ord}(a)} \equiv a^{\operatorname{ord} ab \operatorname{ord}(a)} b^{\operatorname{ord}(ab)\operatorname{ord}(a)} \equiv b^{\operatorname{ord}(ab)\operatorname{ord}(a)} \pmod n.$$

So $\operatorname{ord}(b) \mid \operatorname{ord}(ab)\operatorname{ord}(a)$. Since $\operatorname{ord}(b)$ and $\operatorname{ord}(a)$ are coprime, it follows that $\operatorname{ord}(b) \mid \operatorname{ord}(ab)$. A similar argument shows that $\operatorname{ord}(a) \mid \operatorname{ord}(ab)$. Thus, $\operatorname{ord}(a)\operatorname{ord}(b) \mid \operatorname{ord}(ab)$ by coprimality again.
Since $\operatorname{ord}(a)\operatorname{ord}(b)$ divides and is divisible by $\operatorname{ord}(ab)$, it follows that $\operatorname{ord}(a)\operatorname{ord}(b) = \operatorname{ord}(ab)$, as required.

(b) Mod 5, $\operatorname{ord}(2) = 4$ and $\operatorname{ord}(4) = 2$. So $\operatorname{ord}(4) \neq \operatorname{ord}(2)\operatorname{ord}(2)$.

17.4. (a) $a^x \equiv a^y \pmod{n}$ if and only if $a^{x-y} \equiv 1 \pmod{n}$ if and only if $\operatorname{ord}(a) \mid x - y$ if and only if $x \equiv y \pmod{\operatorname{ord}(a)}$.

(b) This is a special case of (a) since $\operatorname{ord}(g) = p - 1$.

17.5. (a) This follows from the fact that $g^0 \equiv 1 \pmod{p}$.

(b) We have $g^{\log(ab)} \equiv ab \pmod{p}$ and $g^{\log a + \log b} = g^{\log a} g^{\log b} \equiv ab \pmod{p}$. Thus,

$$g^{\log(ab)} \equiv g^{\log a + \log b} \pmod{p}.$$

The previous problem now implies that

$$\log(ab) \equiv \log a + \log b \pmod{p-1}.$$

(c) We have $g^{\log(a^k)} \equiv a^k \pmod{p}$ and $g^{k \log a} = (g^{\log a})^k \equiv a^k \pmod{p}$. So $g^{\log(a^k)} \equiv g^{k \log a} \pmod{p}$, and we can finish things off just as we did in part (b).

17.6. *No solution provided.*

Lecture 18

Exercises

18.6. Since $\varphi(19) = 18$, the possible orders in this case are $1, 2, 3, 6, 9$ and 18. Any integer a of order 9 will be a root of $x^9 - 1 \bmod 19$. The nine roots of this polynomial will then be $1, a, a^2, \ldots, a^8$. Since $\operatorname{ord}(a^i) = \operatorname{gcd}(\operatorname{ord}(a))/\operatorname{gcd}(i, \operatorname{ord}(a)) = 9/\operatorname{gcd}(i, 9)$, the only roots a^i that have order 9 are the ones with $\operatorname{gcd}(i, 9) = 1$. There are $\varphi(9) = 6$ of these. Similarly, any integer of order $d \mid 18$ will be a root of $x^d - 1 \bmod 19$, and by mimicking the above argument, we see that there are at most $\varphi(d)$ of these. In general we see that there are at most

$$\varphi(1) + \varphi(2) + \varphi(3) + \varphi(6) + \varphi(9) = 1 + 1 + 2 + 2 + 6 = 12$$

classes of order < 18. So there must be at least $18 - 12 = 6$ classes of order 18. These are each primitive roots mod 19.

18.11. We have $p - 1 = 2 \cdot 3 \cdot 5$. Let's try $a = 2$. We compute

$$2^{(p-1)/2} = 2^{15} \equiv 1 \pmod{31}.$$

So $a = 2$ isn't a primitive root mod 31. (In retrospect this is obvious, since $2^5 = 32 \equiv 1 \pmod{31}$.) Let's try $a = 3$. We compute

$$\begin{aligned} 2^{(p-1)/2} &= 3^{15} \equiv 30 \pmod{31} \\ 2^{(p-1)/3} &= 3^{10} \equiv 25 \pmod{31} \\ 2^{(p-1)/5} &= 3^6 \equiv 16 \pmod{31}. \end{aligned}$$

Since none of these is congruent to 1, 3 must be a primitive root mod 31.

Practice Problems

18.1. (a) Since g is a primitive root mod p, $a \in \mathbb{Z}$ coprime to p is congruent to g^i for some i. Since $\operatorname{ord}(g^i) = \operatorname{ord}(g)/\operatorname{gcd}(\operatorname{ord}(g), i) = p - 1/\operatorname{gcd}(p - 1, i)$, it follows that g^i is a primitive root if and only if $\operatorname{gcd}(p - 1, i) = 1$.

(b) Since $p - 1 = 2 \cdot 9$, to show that 2 is a primitive root mod 19, we just have to check that $2^{18/2}$ and $2^{18/9}$ are not congruent to 1 mod 19. And indeed,

$$\begin{aligned} 2^9 &\equiv -1 \pmod{19} \\ 2^2 &\equiv 4 \pmod{19}. \end{aligned}$$

So 2 is a primitive root mod 19, and all the other ones are congruent to some 2^i with $\operatorname{gcd}(i, 18) = 1$. Explicitly, here is a set of representatives for the primitive roots mod 19:

$$\{2, 3, 10, 13, 14, 15\}.$$

18.2. *No solution provided. Assignment problem.*

18.3. *No solution provided. Assignment problem.*

18.4. (a) $p - 2$, since the x^{p-1} terms cancel off.

(b) Since $a^{-1} - 1 \equiv 0 \pmod{p}$ by Fermat's Little Theorem, we are left with

$$f(a) \equiv (a-1)(a-2)\cdots(a-(p-1)) \pmod{p}.$$

The product on the right is zero since a occurs among 1, 2, ..., $p-1$.

(c) Part (b) shows that $f(x)$ has $p-1$ distinct roots mod p. Since $\deg f = p-2 < p-1$, it follows that all of the coefficients of f are zero mod p since otherwise we would contradict Theorem 18.2.

(d) The constant term of f is $f(0) = (-1)^{p-1}(p-1)! - (0-1)$. This is zero mod p by part (c). Wilson's Theorem follows immediately upon noting that if p is odd then $(-1)^{p-1} = 1$ while if $p = 2$ then the signs don't matter.

18.5. *No solution provided. Assignment problem.*

Lecture 19

Exercises

19.3. The powers of 3 are:

$$3^1 \equiv 3, \quad 3^2 \equiv 2, \quad 3^3 \equiv 6, \quad 3^4 \equiv 4, \quad 3^5 \equiv 5.$$

So our desired table is:

x	1	2	3	4	5	6
$\log_3(x)$	0	2	1	4	5	3

19.5. Since $\log_3(5) = 5$ and $\log_3(6) = 3$ according the above table, we have

$$\log_3(5) + 46\log_3(x) \equiv \log_3(6) \pmod 6 \quad \Longrightarrow \quad 4\log_3(x) \equiv -2 \pmod 6.$$

Hence

$$2\log_3(x) \equiv -1 \pmod 3$$

and therefore

$$\log_3(x) \equiv 1 \pmod 3.$$

So this gives

$$\log_3(x) \equiv 1 \pmod 6 \quad \text{and} \quad \log_3(x) \equiv 4 \pmod 6$$

from which we find that

$$x \equiv 3^1 \equiv 3 \pmod 7 \quad \text{and} \quad x \equiv 3^4 \equiv 4 \pmod 7.$$

19.8. The cubic residues are 1 and 6, since

$$1^3 \equiv 1, \quad 2^3 \equiv 1, \quad 3^3 \equiv 6, \quad 4^3 \equiv (-3)^3 \equiv 1, \quad 5^3 \equiv (-2)^3 \equiv 6, \quad 6^3 \equiv (-1)^3 \equiv 6.$$

Consequently, the non-cubic residues are 2, 3, 4 and 5.

Practice Problems

19.1. By the kth Power Residue Criterion, the kth power residues are precisely the roots of the polynomial $x^{(p-1)/d} - 1 \bmod p$. By Theorem 18.3, this polynomial has exactly $(p-1)/d$ roots.

19.2. For a to be a cubic residue mod p, we need to have $a^{(p-1)/d} \equiv 1 \pmod p$, where $d = \gcd(p-1, 3)$. If $p \equiv 2 \pmod 3$, then $d = 1$ and so the aforementioned congruence is true for all a coprime to p by Fermat's Little Theorem.

19.3. Let $d = \gcd(p-1, 4)$. If -1 is a 4th power residue mod p then $(-1)^{(p-1)/d} \equiv 1 \pmod p$ so $(p-1)/d$ must be even. Note that $p-1$ is even so we can write $p-1 = 2^v u$ with u odd and $v \geq 1$. If $v \leq 2$ then $d = \gcd(p-1, 4) = 2^v$ so $(p-1)/d = u$ is odd—contradiction. So $v \geq 3$ and thus $p - 1 \equiv 0 \pmod 8$. [**Alternatively:** We can note that if -1 is a 4th power mod p then it must be a square mod p, so putting $d' = \gcd(p-1, 2) = 2$, we would have $(-1)^{(p-1)/2} \equiv 1 \pmod p$. Thus, $(p-1)/2$ must be even, so $4 \mid (p-1)$. This implies that $d = \gcd(p-1, 4) = 4$. Since we had observed that $(p-1)/d$ is even, it follows that $(p-1)/4 = 2k$ hence $p = 8k+1$.]

19.4. If $p_1, \ldots, p_n$ are all such primes then $N = (2p_1 \cdots p_n)^4 + 1$ is odd and is not divisible by any prime of the form $8k+1$. Since $N > 1$, N must have an odd prime divisor q. But then $N \equiv 0 \pmod q$ implies that $(2p_1 \cdots p_n)^4 \equiv -1 \pmod q$ meaning -1 is a quartic residue mod q. By the previous problem, this forces q to be congruent to 1 mod 8 which is a contradiction since no such prime divides N.

Lecture 20

Exercises

20.2. We have

$$\left(\frac{a}{7}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod 7 \\ 1 & \text{if } a \equiv 1, 2, 4 \pmod 7 \\ -1 & \text{if } a \equiv 3, 5, 6 \pmod 7. \end{cases}$$

20.5. Since $8^{(17-1)/2} \equiv 1 \pmod{17}$, it follows that $\left(\frac{8}{17}\right) = 1$.

20.8. The same argument as in Example 20.7 allows us to conclude that $x \equiv 1 \pmod 4$. Then, by adding 16 to the given equation we can re-write it as

$$y^2 + 16 = x^3 + 27 = (x+3)(x^2 - 3x + 9).$$

Now, $x^2 - 3x + 9 \equiv 3 \pmod 4$ since $x \equiv 1 \pmod 4$. Also, $x^2 - 3x + 9$ is > 1 so by the argument in Example 20.7 we conclude that $x^2 - 3x + 9$ has a prime divisor $p \equiv 3 \pmod 4$. From this we get that $y^2 + 16 \equiv 0 \pmod p$ and hence that $-1 \equiv (-4^{-1}y)^2 \pmod 4$. Thus, $\left(\frac{-1}{p}\right) = 1$, a contradiction since $p \equiv 3 \pmod 4$.

20.9. *No solution provided.*

Practice Problems

20.1. Straightforward using Legendre symbols.

20.2. This is a special case of Problem 19.1.

20.3. (a) The congruence has either 0, 1 or 2 solutions depending respectively on whether a is a quadratic nonresidue mod p, $a \equiv 0 \pmod p$ or a is a quadratic residue mod p. These counts match up with $1 + \left(\frac{a}{p}\right)$.

(b) Multiplying by $4a$ (which is invertible mod p) and completing the square, the congruence is equivalent to

$$(2ax + b)^2 - (b^2 - 4ac) \equiv 0 \pmod p.$$

Letting $X = 2ax + b$, this becomes $X^2 \equiv b^2 - 4ac \pmod p$. Now apply part (a).

20.4. Since $p - 1 = 2^2 q$ is the prime factorization of $p - 1$, Proposition 18.9 tells us that we need only check that neither of $2^{(p-1)/2}$ and $2^{(p-1)/q}$ is congruent to 1 mod p. To this end, we have

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

by Euler's criterion. Since q is odd, $p = 4q + 1 = 4(2k + 1) + 1 = 8k + 5$ is congruent to -3 mod 8, so $\left(\frac{2}{p}\right) = -1$. Thus, $2^{(p-1)/2} \not\equiv 1 \pmod p$.

Next,

$$2^{(p-1)/q} = 2^4 = 16.$$

If this were 1 mod p, then p would divide $15 = 3 \cdot 5$ implying that either $p = 3$ or $p = 5$ neither of which is of the form $4q + 1$ with q a prime. So $2^{(p-1)/q} \not\equiv 1 \pmod p$, and we're done.

20.5. *No solution provided. Assignment problem.*

20.6. First note that

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

Now, since half of the congruence classes in $(\mathbb{Z}/p\mathbb{Z})^\times$ are quadratic residues and the other half are nonresidues, half the terms in the sum $\sum_{k=0}^{p-1}\left(\frac{k}{p}\right)$ are $+1$ and the other half are -1. So the sum is 0.

20.7. *No solution provided. Assignment problem.*

Lecture 21

Exercises

21.2. If $p = 8k \pm 1$ then $(p^2-1)/8 = 8k^2 \pm 2k$ is even, so $(-1)^{(p^2-1)/2} = 1$. The other cases, where $p = 8k \pm 3$, follow in the same way.
Similarly, $(p-1)(q-1)/4$ is odd if and only if both p and q are congruent to 3 mod 4. This proves part (b). So $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = -1$ if and only if $p \equiv q \equiv 3 \pmod 4$. Multiplying both sides by $\left(\frac{q}{p}\right)$ (and noting that $\left(\frac{q}{p}\right)^2 = (\pm 1)^2 = 1$ in any case), we get half of the statement in the theorem. The other half follows similarly.

21.4. We have

$$\left(\frac{30}{61}\right) = \left(\frac{2}{61}\right)\left(\frac{3}{61}\right)\left(\frac{5}{61}\right).$$

Now,

$$\begin{aligned}\left(\frac{2}{61}\right) &= -1\\ \left(\frac{3}{61}\right) &= \left(\frac{61}{3}\right)\\ &= \left(\frac{1}{3}\right)\\ &= 1\\ \left(\frac{5}{61}\right) &= \left(\frac{61}{5}\right)\\ &= \left(\frac{1}{5}\right)\\ &= 1.\end{aligned}$$

Thus,

$$\left(\frac{30}{61}\right) = (-1)(1)(1) = -1.$$

21.6. Since $5 \equiv 1 \pmod 4$, we end up with (for $p \neq 5$):

$$\left(\frac{5}{p}\right) = \left(\frac{p}{5}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 4 \pmod 5 \\ -1 & \text{if } p \equiv 2, 3 \pmod 5. \end{cases}$$

On the other hand, since $7 \equiv 3 \pmod 4$, then we will need to consider two cases: one for $p \equiv 1 \pmod 4$ and one for $p \equiv 3 \pmod 4$. We can then express our end results as conditions on $p \bmod 4 \cdot 7 = 28$. This is what we get (for $p \neq 7$):

$$\left(\frac{7}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 3, 9, 19, 25, 27 \pmod{28} \\ -1 & \text{if } p \equiv 5, 11, 13, 15, 17, 23 \pmod{28}. \end{cases}$$

21.9. (a) If say $d = \gcd(x, y)$, we can write $x = du$ and $y = dv$, where $\gcd(u, v) = 1$, and then the equation becomes

$$d^4u^4 - 17d^4v^4 = 2z^2.$$

This shows that $d^4 \mid 2z^2$. Thus, d^2 must divide z, and so we can write $z = d^2t$ and then we can cancel off d from all sides of the equation obtaining

$$u^4 - 17v^4 = 2t^2.$$

This is the same as our starting equation, but now $\gcd(u, v) = 1$. So we may as well have assumed that $\gcd(x, y) = 1$. A similar analysis with $\gcd(x, z)$ and $\gcd(y, z)$ shows that we can assume that they are both equal to 1.

(b) We can, of course, just compute $1^4, 2^4, \ldots, 16^4$ and check if any of these are congruent to 2 mod 17. However, a (much) better approach is to use Theorem 19.9 (kth Power Residue Criterion). We simply compute

$$2^{16/\gcd(16,4)} = 2^4 = 16 \equiv -1 \pmod{17}.$$

Since this is $\not\equiv 1 \pmod{17}$, it follows that 2 is not a 4th power residue mod 17.

(c) If $z = 0$ then $x^4 - 17y^4 = 0$ or equivalently $x^4 = 17y^4$. At this point we can argue that $x = y = 0$ since 17 is irrational, but an argument is also possible using techniques we've developed in the course. Consider the 17-adic valuation. We get

$$v_{17}(x^4) = v_{17}(17y^4) \quad \Longrightarrow \quad 4v_{17}(x) = 1 + 4v_{17}(y).$$

This is impossible since the left side is divisible by 4 but the right side is congruent to 1 mod 4.

Practice Problems

21.1. (a) If $p \mid N$ then $5(p_1 \cdots p_n)^2 \equiv 1 \pmod{p}$. Note that $p \neq p_i$ since otherwise p would not divide N. Thus, each p_i is invertible mod p, giving us

$$5 \equiv (p_1^{-1} \cdots p_n^{-1})^2 \pmod{p}.$$

So $\left(\frac{5}{p}\right) = 1$ (since $p \neq 5$ because $5 \nmid N$) and hence $\left(\frac{p}{5}\right) = 1$ by quadratic reciprocity.

(b) If $\left(\frac{p}{5}\right) = 1$ then $p \equiv 1, 4 \pmod{5}$. If all the primes dividing N were congruent to 1 mod 5, then N (being their product) would also be congruent to 1 mod 5. But N is congruent to -1 mod 5. So at least one of the primes dividing N must be congruent to 4 mod 5.

(c) By part (b), we know that there must be some prime of the form $5k + 4$ that divides N. This prime cannot be any of the p_i. So we have found a new such prime. Repeating this argument, we can produce infinitely many primes of the form $5k + 4$.

21.2. (a) Note that $p \equiv 1 \pmod{4}$ and also $p \equiv (-1)^{2^n} + 1 \equiv 2 \pmod{3}$. So by quadratic reciprocity,

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

(b) By part (a) and Euler's criterion, we have

$$3^{(p-1)/2} \equiv \left(\frac{3}{p}\right) \equiv -1 \pmod{p}.$$

Since 2 is the only prime divisor of $p-1$, it follows that 3 must be a primitive root mod p by Proposition 18.9.

21.3. *No solution provided. Assignment problem.*

21.4. *No solution provided. Assignment problem.*

21.5. By completing the square, we can re-write the equation

$$x^2 + 10xy - 6y^2 = 17$$

as

$$(x + 5y)^2 - 31y^2 = 17.$$

By reducing mod 31 (which is prime), we see that 17 must be a quadratic residue mod 31, that is,

$$\left(\frac{17}{31}\right) = 1.$$

However, we have

$$\left(\frac{17}{31}\right) = \left(\frac{31}{17}\right) = \left(\frac{-3}{17}\right) = \left(\frac{-1}{17}\right)\left(\frac{17}{3}\right) = (+1)\left(\frac{2}{3}\right) = -1.$$

Lecture 22

Exercises

22.2. We have

$$5 \cdot 1 \equiv -2, \quad 5 \cdot 2 \equiv 3 \quad \text{and} \quad 5 \cdot 3 \equiv 1.$$

So $n(5) = 1$ and therefore $\left(\frac{5}{7}\right) = -1$.

Practice Problems

22.1. We have that $n(-1) = |S| = (p-1)/2$ since $(-1)S \in -S$ for all s in S. Thus, by Gauss's Lemma,

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

This is equivalent to what must be proved.

Lecture 23

Exercises

23.5. Since $561 = 3 \times 11 \times 17$, it suffices, by the Chinese Remainder Theorem, to show that

$$\begin{aligned} a^{560} &\equiv 1 \pmod 3 \\ a^{560} &\equiv 1 \pmod{11} \\ a^{560} &\equiv 1 \pmod{17} \end{aligned}$$

for all a coprime to 561 (hence coprime to 3, 11 and 17). Each of these follows immediately from Fermat's Little Theorem. Indeed: $p - 1 \mid 560$ for $p = 3, 11, 17$.

23.9. Here, $n = 2^5 \times 495 + 1$. Let's try $a = 2$. We have

$$2^{495} \equiv 1 \pmod{15841}.$$

So n passes the Miller–Rabin test for the base $a = 2$.

Let's try $a = 3$.

$$\begin{aligned} 3^{495} &\equiv 12802 \pmod{15841} \\ 3^{2\cdot 495} &\equiv 218 \pmod{15841} \\ 3^{2^2\cdot 221} &\equiv 1 \pmod{15841}. \end{aligned}$$

So n fails the Miller–Rabin test. Thus, n is composite.

23.14. *No solution provided.*

23.16. We have

$$\left(\frac{655}{719}\right) = -\left(\frac{719}{655}\right) = -\left(\frac{64}{655}\right) = \left(\frac{8^2}{655}\right) = -1.$$

Practice Problems

23.1. (a) Suppose $d \mid n$ is a proper divisor (i.e. $d \neq n$). Then $d^{n-1} \equiv 1 \pmod n$ implies that $n \mid d^{n-1} - 1$ and hence $d \mid d^{n-1} - 1$ since $d \mid n$. On the other hand, since $d \mid d^{n-1}$, it follows that $d \mid 1 = (d^{n-1} - 1) - d^{n-1}$. So $d = \pm 1$. Thus, the only proper divisors of n are ± 1, so n is prime.

(b) There's not contradiction because n is a Carmichael number if $a^{n-1} \equiv 0 \pmod n$ for a coprime to n, not for all $a \not\equiv 0 \pmod n$.

23.2. We wish to compute $2^{F_n - 1} = 2^{2^{2^n}}$ modulo F_n. Note that

$$F_n \equiv 0 \pmod{F_n} \quad \iff \quad 2^{2^n} \equiv -1 \pmod{F_n}.$$

Raising both sides to $2^{2^n - n}$ (which is even), we get

$$2^{2^n \cdot 2^{2^n - n}} \equiv (-1)^{2^{2^n - n}} \pmod{F_n} \quad \iff \quad 2^{2^{2^n}} \equiv 1 \pmod{F_n}.$$

So, if F_n is composite, it's a base-2 pseudoprime.

23.3. (a) This is true. We know n is composite and that a and b are coprime to n. Since $a^{n-1} \equiv 1 \pmod n$ and $b^{n-1} \equiv 1 \pmod n$, we also have that $(ab)^{n-1} = a^{n-1}b^{n-1} \equiv 1 \pmod n$. So n a base-ab pseudoprime.

(b) This is also true, since

$$a^{n-1} \equiv 1 \pmod{n} \implies (a^{-1})^{n-1} \equiv (a^{n-1})^{-1} \equiv 1^{-1} \equiv 1 \pmod{n}.$$

23.4. If n is a base-a Euler pseudoprime, then n is composite, coprime to a and we have

$$\left(\frac{a}{n}\right) \equiv a^{(n-1)/2} \pmod{n}.$$

By squaring both sides and using the fact that $\left(\frac{a}{n}\right)^2 = 1$, we see that

$$1 = a^{n-1} \pmod{n}.$$

So n is a Fermat pseudoprime.

To see that the converse is false, note that $n = 341$ is a Fermat base-2 pseudoprime, but

$$\left(\frac{2}{341}\right) = \left(\frac{2}{11}\right)\left(\frac{2}{31}\right) = (-1)(1) = -1$$

while

$$2^{(341-1)/2} = 2^{170} \equiv 1 \pmod{3)41}.$$

So n is not an Euler pseudoprime.

Lecture 24

Exercises

24.4. Let $\alpha = 3 + 4i$ ad $\beta = 1 + 2i$. Then $\beta \nmid \alpha$ but $N(\beta) = 5$ divides $N(\alpha) = 25$.

24.6. Let $\beta = 2$ and $\alpha = 1 + i$. (So α is in the center of the square with vertices at 0, β, $i\beta$ and $(1 + i)\beta$. Then

$$\alpha = 0\beta + (1 + i) \quad \text{and} \quad \alpha = (1)\beta + (-1 + i)$$

are decompositions of α with

$$N(1 + i) = N(-1 + i) = 2$$

strictly less than $N(\beta) = N(2) = 4$.

Practice Problems

24.1. We have $\beta = \alpha\tau$ and $\gamma = \alpha\rho$ for some $\tau, \rho \in \mathbb{Z}[i]$, so $\beta x + \gamma y = \alpha(\tau x + \rho y)$ is divisible by α.

24.2. If $a \mid b$ in $\mathbb{Z}[i]$ then $b = a(c + di)$ for some $c, d \in \mathbb{Z}$. Thus,

$$b = ac + (ad)i.$$

By equating real parts, we see that $b = ac$, so $a \mid b$ in $\mathbb{Z}$.

24.3. Let $\alpha = x + iy$. Observe that if $N(\alpha) = n \iff x^2 + y^2 = n$, so we must have $|x| \leq \sqrt{n}$ and $|y| \leq \sqrt{n}$, because if either $|x|$ or $|y|$ is $> \sqrt{n}$ then $x^2 + y^2 > n$.
Now let's consider each equation separately.

- $N(\alpha) = 1 \iff x^2 + y^2 = 1$. In this case $|x| \leq 1$ and $|y| \leq 1$. From this we quickly see that $\alpha \in \{\pm 1, \pm i\}$.
- $N(\alpha) = 2 \iff x^2 + y^2 = 2$. In this case $|x| \leq 1$ and $|y| \leq 1$, too, since $x, y \in \mathbb{Z}$. Running through the cases, we end up finding that $\alpha \in \{\pm(1 + i), \pm(1 - i)\}$.
- $N(\alpha) = 3 \iff x^2 + y^2 = 3$. There are no solutions (consider the equation mod 4; or do a case analysis like above).
- $N(\alpha) = 4 \iff x^2 + y^2 = 4$. This time $|x| \leq 2$ and $|y| \leq 2$. Running through the cases, we end up finding that $\alpha \in \{\pm 2, \pm 2i\}$.
- $N(\alpha) = 5 \iff x^2 + y^2 = 5$. So $|x| \leq 2$ and $|y| \leq 2$ again. Running through the cases this time, we find $\alpha \in \{\pm(1 + 2i), \pm(1 - 2i), \pm(2 + i), \pm(2 - i)\}$.

24.4. This is false. For example, $i \mid 1$ because $i(-i) = 1$ and $1 \mid i$ because $1 \cdot i = i$, but $1 \neq \pm i$.

24.5. If $\alpha \mid 2$ then $N(\alpha) \mid N(2) = 4$. So the possibilities for $N(\alpha)$ are $1, 2$ and 4. From the previous problem, we see therefore that the possibilities for α are $\pm 1, \pm i, \pm(1 + i), \pm(1 - i), \pm 2$ and $\pm 2i$. We must now check that each of these does in fact divide 2. And indeed:

$$\begin{aligned} 2 &= (\pm 2)(\pm 1) \\ &= (\mp 2i)(\pm i) \\ &= \pm(1 + i)(\mp(1 - i)) \\ &= \pm(1 - i)(\mp(1 + i)) \\ &= \pm 2(\pm 1) \\ &= (\pm 2i)(\mp i). \end{aligned}$$

Lecture 25

Exercises

25.7. If $\gcd(\alpha, \beta) = 1$, then every common divisor of α and β divides 1 hence is a unit. The converse is obvious.

25.11. Write $1 = \alpha x + \beta y$ by Bézout. Then multiply by γ to get $\gamma = \alpha\gamma x + \beta\gamma y$. Since α divides $\alpha\gamma x$ and $\beta\gamma y$, it must also divide their sum, which is γ.

Practice Problems

25.1. By Bézout's Lemma, we can write $\gcd(\alpha, \beta) = \alpha x + \beta y$ with $x, y \in \mathbb{Z}[i]$. If α and β are coprime, then their gcd is a unit, say u. So $u = \alpha x + \beta y \iff 1 = \alpha(u^{-1}x) + \beta(u^{-1}y)$.

25.2. Let g and γ denote the gcd of a and b in $\mathbb{Z}$ and $\mathbb{Z}[i]$, respectively. By Bézout's Lemma in $\mathbb{Z}$, we can write $g = ax + by$ for some $x, y \in \mathbb{Z}$. Since γ divides both a and b, it follows that $\gamma \mid g$. On the other hand, since $g \in \mathbb{Z}[i]$ is a common divisor of a and b, then $g \mid \gamma$. So $\gamma = ug$ for some unit $u \in \mathbb{Z}[i]$.

25.3. If δ is a common divisor of α and β then $\delta \mid \alpha - \beta q = r$. So δ is a common divisor of β and r.

Conversely, if δ is a common divisor of β and r, then $\delta \mid \beta q + r = \alpha$. So δ is a common divisor of β and α.

So the pairs (α, β) and (β, r) have the same common divisors, and therefore they have the same gcd.

25.4. By Problem 1, we can write $1 = \alpha x + \beta y$ with $x, y \in \mathbb{Z}[i]$. Therefore,

$$\gamma = \gamma \cdot 1 = \gamma(\alpha x + \beta y) = \alpha\gamma x + \beta\gamma x.$$

On the other hand, since $\alpha \mid \gamma$ and $\beta \mid \gamma$, we can write $\gamma = \alpha z$ and $\gamma = \beta w$ for some $z, w \in \mathbb{Z}[i]$. So the above becomes

$$\gamma = \alpha(\beta w)x + \beta(\alpha z)y = \alpha\beta(wx + zw).$$

This shows that $\alpha\beta \mid \gamma$, as required.

Lecture 26

Exercises

26.6. (a) If $\beta \mid \alpha$ then $N(\beta) \mid N(\alpha)$ so either $N(\beta) = 1$, in which case β is a unit, or $N(\beta) = p$ in which case β is an associate of α (since $\alpha = \beta\gamma$ and we must have $N(\gamma) = 1$ if $N(\alpha) = N(\beta)$).

(b) If $p = \alpha\beta$ non-trivially then $p^2 = N(p) = N(\alpha)N(\beta)$ implies that $N(\alpha) = N(\beta) = p$. If $\alpha = a + ib$ then $p = N(\alpha) = a^2 + b^2$ is a sum of two integer squares, so p cannot be 3 mod 4.

26.11. We have $\alpha = 2(2 - 9i)$ so we must factor 2 and $2 - 9i$. We know how to deal with 2:

$$2 = i(1 - i)^2.$$

To factor $2 - 9i$ into Gaussian prime, we begin by computing

$$N(2 - 9i) = 4 + 91 = 85 = 5 \times 17 = (1 + 2i)(1 - 2i)(1 + 4i)(1 - 4i),$$

where we used $5 = 1^2 + 2^2$ and $17 = 1^2 + 4^2$ to factor 5 and 17. Now we just have to multiply two of these factors to get $2 - 9i$ (perhaps up to a unit). By inspection, we find that

$$2 - 9i = -i(1 - 2i)(1 + 4i).$$

Thus,

$$4 - 18i = i(1 - i)^2(-i)(1 - 2i)(1 + 4i) = (1 - i)^2(1 - 2i)(1 + 4i).$$

Practice Problems

26.1. $\implies$: This is Euclid's Lemma.

$\impliedby$: Suppose that $\pi = \alpha\beta$. Then $\pi \mid \alpha\beta$ and so, wlog, $\pi \mid \alpha$, hence we can write $\alpha = \pi\gamma$ for some $\gamma \in \mathbb{Z}[i]$. But then $\pi = \alpha\beta = \pi\gamma\beta \implies 1 = \gamma\beta$, so β is a unit. Thus, he only way to write π as a product of two Gaussian integers is if one of the two is a unit. So π is a Gaussian prime.

26.2. If $p = a^2 + b^2$ then letting $\pi = a + ib$ we have

$$p = (a + ib)(a - ib) = \pi\overline{\pi}.$$

We know that π and $\overline{\pi}$ are Gaussian primes by Theorem 26.7. Thus, every decomposition of p into a sum of two squares gives rise into a factorization of p into Gaussian primes. Since such a factorization is unique up to units, and since multiplying $a \pm ib$ by a unit $u \in \{-1, \pm i\}$ swaps a sign and/or swaps the positions of a and b, the result follows.

26.3. If $n = x^2 + y^2$ and $m = u^2 + v^2$ then

$$n = N(x + iy) \quad \text{and} \quad m = N(u + iv)$$

so

$$nm = N((x + iy)(u + iv)) = N((xy - yv) + (xv + yu)i) = (xy - yv)^2 + (xv + yu)^2.$$

26.4. *Sketch:* Factor γ into Gaussian primes, say $\gamma = \prod_\pi \pi^a$. Then $\alpha\beta = \gamma^n = \prod \pi^{an}$. Since α and β are coprime, their prime factorizations do not share a common Gaussian prime. The product of the prime factorizations of α and β gives a prime factorization of γ^n. By unique factorization, α must be equal to a unit times a collection of some π^{an}; likewise for β. But now observe that

$$u \prod \pi^{an} = u(\prod \pi^a)^n$$

is a unit times an nth power.

26.5. Since there are infinitely many rational primes $p \equiv 3 \pmod 4$, and since these are all Gaussian primes, there are infinitely many Gaussian primes.

Alternative Solution: Each rational prime p has a Gaussian prime divisor. We claim that if $p \neq q$ are distinct rational primes, then they cannot share a Gaussian prime divisor. Indeed, since $\gcd(p, q) = 1$ in $\mathbb{Z}$, we must also have $\gcd(p, q) = 1$ in $\mathbb{Z}[i]$ by an earlier problem, so the claim follows. Now since there are infinitely many rational primes, and since we can select a Gaussian prime divisor of each and no two of these are the same, it follows that there are infinitely many Gaussian primes.

Lecture 27

Exercises

27.2. We take our cue from

$$N(a+b\sqrt{-D}) = a^2 + Db^2.$$

We have

$$\begin{aligned}(a^2+Db^2)(u^2+Dv^2) &= N(a+b\sqrt{-D})N(u+v\sqrt{-D}) \\ &= N((a+b\sqrt{-D})(u+v\sqrt{-D})) \\ &= (au-Dbv)^2 + D(av+bu)^2.\end{aligned}$$

27.4. Let $q = q_j$. If $q \mid x$ then since $q \mid n$, we would have that $q \mid n - x^2 = y^2$. But then $q^2 \mid x^2 + y^2 = n$. So we may divide both sides of the equation by q^2. Repeating this argument sufficiently often, we are reduced to a case where $q \mid n$ (there will always be a q left in n because $v_q(n)$ is odd) but where $q \nmid x$.

27.6. We have $r_2(100) = r_2(2^2 5^2) = 4(1+2) = 12$. Now, by inspection we find

$$\begin{aligned}100 &= 0^2 + (\pm 10)^2 = (\pm 10)^2 + 0^2 && \text{(4 solutions)} \\ &= (\pm 6)^2 + (\pm 8)^2 = (\pm 8)^2 + (\pm 6)^2. && \text{(8 solutions)}\end{aligned}$$

27.6. Any common divisor d of m and n will divide each of a, b and c. Since $\gcd(a,b,c) = 1$, it follows that $d = \pm 1$. So $\gcd(m,n) = 1$. In particular, one of m and n must be odd. If they were both odd then $a = m^2 - n^2$ would be even, which contradicts Lemma 27.7.

Practice Problems

27.1. Note that $r_2(n)$ counts the number of integer pairs (x,y) such that $x^2 + y^2 = n$. If n is not a perfect square, then neither x nor y can be zero. So, for every such solution (x,y), precisely one of

$$(x,y),\ (x,-y),\ (-x,y) \text{ or } (-x,-y)$$

is positive. Thus, $r_2^+(n) = \frac{1}{4} r_2(n)$.

If n is a perfect square, say $n = N^2$, then there are four solutions to $n = x^2 + y^2$ with either x or y equal to 0, namely: $(x,y) = (\pm N, 0)$ and $(x,y) = (0, \pm N)$. Excluding these solutions, an argument as in the preceding paragraph shows that

$$r_2^+(n) = \frac{1}{4}(r_2(n) - 4) = \frac{1}{4} r_2(n) - 1.$$

27.2. A positive integer n can be written as $n = x^2 - y^2$ if and only if $n \not\equiv 2 \pmod 4$.

Indeed, if $n = x^2 - y^2$ then since squares are 0 or 1 mod 4, we see that n must be either $0, 1$ or 3 mod 4. Equivalently, n must be odd or divisible by 4.

Conversely, assume $n = 2k+1$ is odd. Then

$$2k+1 = (k+1)^2 - k^2.$$

If $n = 4k$ is divisible by 4, then

$$4k = (k+1)^2 - (k-1)^2.$$

27.3. *No solution provided. Assignment problem.*

27.4. *No solution provided. Assignment problem.*

Lecture 28

Exercises

28.3. *No solution provided. Refer to the discussion following the exercise.*

28.4. *Ditto.*

Practice Problems

28.1. If $k < 0$, say $k = -l$, then $\sqrt{k} = li$ is not real. So

$$a + b\sqrt{k} = c + d\sqrt{k} \iff a + bli = c + dli.$$

Equating real and imaginary parts, we get

$$a = c \quad \text{and} \quad bl = dl \iff b = d.$$

If $k > 0$ and not square, then $\sqrt{k}$ is irrational. We can re-arrange the equation into

$$a - c = (d - b)\sqrt{k}.$$

If $d \neq b$ then we get a contradiction since this implies that

$$\sqrt{k} = \frac{a - c}{d - b}.$$

So $d = b$. Consequently $a - c = 0$, so $a = c$.

28.2. We will assume that:

(i) The only units in $\mathbb{Z}[\sqrt{-2}\,]$ are ± 1.

(ii) There is a version of the FTA asserting that numbers in $\mathbb{Z}[\sqrt{-2}\,]$ can be uniquely factored into "primes" in $\mathbb{Z}[\sqrt{-2}\,]$.

(iii) The norm function $N(a+b\sqrt{-2}) = a^2+2b^2$ satisfies Proposition 24.3. In particular, if $N(a + b\sqrt{-2})$ is a rational prime then $a + b\sqrt{-2}$ is a prime in $\mathbb{Z}[\sqrt{-2}\,]$.

[**Note:** (i) and (iii) are easy to prove; (ii) is true but requires more work.]

Starting from

$$(y - \sqrt{-2})(y + \sqrt{-2}) = x^3$$

we first claim that $y - \sqrt{-2}$ and $y + \sqrt{-2}$ are coprime in $\mathbb{Z}[\sqrt{-2}\,]$. To see why, suppose $\delta \in \mathbb{Z}[\sqrt{-2}\,]$ is a common divisor. Then

$$\delta \mid (y + \sqrt{-2}) - (y - \sqrt{-2}) = 2\sqrt{-2} = (\sqrt{-2})^3.$$

Now, note that $\sqrt{-2}$ is a prime in $\mathbb{Z}[\sqrt{-2}\,]$. (Indeed, if $N(\sqrt{-2}) = 2$ is prime.) So, by unique factorization, $\delta = \pm(\sqrt{-2})^i$ for some $i \in \{1, 2, 3\}$ hence $N(\delta) = 2^i$.

On the other hand, since $\delta \mid y + \sqrt{-2}$ we get that $N(\delta) \mid N(y + \sqrt{-2}) \iff 2^i \mid y^2 + 2$. But if $y^2 = x^3 - 2$ then it's easy to see that y must be odd. Thus, we must have $i = 0$ so $\delta = \pm(\sqrt{-2})^i = \pm 1$ is a unit. This completes the proof that $\gcd(y - \sqrt{-2}, y + \sqrt{-2}) = 1$.

By unique factorization, we can therefore conclude that $y \pm \sqrt{-2}$ are each a unit times a cube in $\mathbb{Z}[\sqrt{-2}\,]$, and since the units $\pm 1 = (\pm 1)^3$ are themselves cubes, we can simply assume that $y \pm \sqrt{-2}$ are cubes. In particular,

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3 = a(a^2 - 6b^2) + b(3a^2 - 2b^2)\sqrt{-2}.$$

Equating coefficients of $\sqrt{-2}$ as usual, we eventually find that $a = \pm 1$ and $b = 1$, so $y = a(a^2 - 6b^2) = \pm 5$ and therefore $x = 3$.

28.3. *No solution provided. Assignment problem.*

Lecture 29

Exercises

29.2. If $D = 0$ then the equation is just $x^2 = 1$ so $S_0 = \{(\pm 1, y) \colon y \in \mathbb{Z}\}$.
If $D = d^2$ with $d \neq 0$ then the equation reduces to

$$x^2 - (dy)^2 = 1 \quad \Longleftrightarrow \quad = (x - dy)(x + dy) = 1.$$

From this we get that either $x - dy = x + dy = 1$ or $x - dy = x + dy = -1$. In either case we get that $y = 0$ and hence $x = \pm 1$. So $S_{d^2} = \{(\pm 1, 0)\}$.

29.4. Part (a) Follows from the fact that norm is multiplicative. Parts (b) and (c) are routine calculations. For part (d), we claim that $(a, b)^{-1} = (a, -b)$. To confirm, we just calculate $(a, b) * (a, b)$ and verify that this results in $(1, 0)$.

29.8. It suffices to confirm that there are no solutions when $x = 1$ and $x = 2$. And indeed, in this case the equation $x^2 - 2y^2 = 1$ reduces, respectively, to $1 - 2y^2 = 1$ which has $y = 0$ (not positive) and $4 - 2y^2 = 1$ which has no integer solutions.

29.10. It's clear that $\pm(a, b)^n \in S_D$ for all $n \in \mathbb{Z}$. Conversely, let $(x, y) \in S_D$. If both x and y are positive, then $(x, y) \in S_D^+$ so $(x, y) = (a, b)^n$ for some $n \geq 1$. Assume now that $x \leq 0$. The case $x = 0$ is impossible, so we must have $x < 0$. If $y < 0$ too, then $(-x, -y) = -(x, y)$ is in S_D^+ so $(x, y) = -(a, b)^n$ for some $n \geq 1$. If $y = 0$ then $(x, y) = \pm(1, 0) = \pm(a, b)^0$. Finally, if $y > 0$ then $(-x, y) \in S_D^+$ so $(-x, y) = (a, b)^m$ for some $m \geq 1$ hence

Practice Problems

29.1. *Answer:*

D	Fundamental Solution
2	$(3, 2)$
3	$(2, 1)$
5	$(9, 4)$
6	$(5, 2)$
7	$(8, 3)$
8	$(3, 1)$
10	$(19, 6)$

29.2. If (a, b) is a solution then from

$$(a + b\sqrt{5})^2 = (a^2 + 5b^2) + (2ab)\sqrt{5}$$

we see that $(a^2 + 5b^2, 2ab)$ is a solution. Starting with $(a, b) = (9, 4)$ and repeatedly applying this rule, we get:

$$(161, 72),\ (51841, 23184),\ (5374978561, 2403763488),$$

$$\text{and}\ \ (57780789062419261441, 25840354427429161536).$$

29.3. (a) If (a, b) is the fundamental solution then $(a, b) * (u, v)$ will be a solution to $x^2 - Dy^2 = m$. This is because

$$N((u + v\sqrt{D})(a + b\sqrt{D})) = N(u + v\sqrt{D})N(a + b\sqrt{D}) = m \cdot 1.$$

(b) By inspection, we find that $(x, y) = (4, 1)$ is a solution to $x^2 - 2y^2 = 14$. Since $(8, 3)$ is the fundamental solution to $x^2 - 2y^2 = 1$, it follows that $(8, 3)^n * (4, 1)$ will be solutions to $x^2 - 2y^2 = 14$ for $n = 1, 2, \ldots$.

29.4. If there is a solution, then reducing mod 3 we see that 14 must be a square mod 3. However,

$$\left(\frac{14}{3}\right) = \left(\frac{2}{3}\right)\left(\frac{7}{3}\right) = (-1)(1) = -1.$$

So 14 isn't a square mod 3.

29.5. Suppose $p \equiv 3 \mod 4$ divides D. Then reducing $x^2 - Dy^2 = -1 \mod p$ gives that -1 is a square mod p. This is a contradiction because $\left(\frac{-1}{p}\right) = -1$ as $p \equiv 3 \pmod 4$.

Lecture 30

Exercises

30.4. *Answer:* $[2,\overline{4}]$ and $[2,\overline{2,4}]$.

30.9. *Answer:* $1, 2, \frac{5}{3}, \frac{7}{4}, \frac{19}{11}$.

30.16. *No solution provided.*

Practice Problems

30.1. (a) From $\alpha = 1 + \frac{1}{\alpha}$ we get

$$\alpha^2 - \alpha - 1 = 0$$

and hence $\alpha = \dfrac{1+\sqrt{5}}{2}$ (since α must be the positive root of the above quadratic).

(b) Let's first determine $\gamma = [\overline{2,6}]$:

$$\gamma = 2 + \cfrac{1}{6 + \cfrac{1}{2 + \cfrac{1}{6 + \cfrac{1}{\ddots}}}}$$

$$= 2 + \cfrac{1}{6 + \cfrac{1}{\gamma}}.$$

This simplifies to give

$$3\gamma^2 - 6\gamma - 1 = 0$$

hence

$$\gamma = \frac{3 + 2\sqrt{3}}{3}.$$

Consequently,

$$\beta = 1 + \cfrac{1}{1 + \cfrac{1}{\gamma}} = \frac{9 + 4\sqrt{3}}{6 + 2\sqrt{3}} = \frac{5 + \sqrt{3}}{4}.$$

30.2. (a) Let $\alpha = \sqrt{n^2+1}$. Then $\lfloor \alpha \rfloor = n$ so

$$\alpha = n + (\sqrt{n^2+1} - n) = n + \frac{1}{\frac{1}{\sqrt{n^2+1}-n}}.$$

Now we must deal with

$$\frac{1}{\sqrt{n^2+1}-n} = \frac{1}{\sqrt{n^2+1}-n} \cdot \frac{\sqrt{n^2+1}-n}{\sqrt{n^2+1}-n} = \sqrt{n^2+1} + n.$$

The floor of this number is $2n$, so

$$\sqrt{n^2+1} + n = 2n + (\sqrt{n^2+1} - n) = \frac{1}{\frac{1}{\sqrt{n^2+1}-n}}.$$

And now the pattern repeats! Thus,

$$\begin{aligned}\alpha &= n + \cfrac{1}{2n + \cfrac{1}{2n + \cfrac{1}{2n + \cfrac{1}{2n + \ddots}}}} \\ &= [n, \overline{2n}].\end{aligned}$$

(b) From part (a), we see that the continued fraction expansion of $\sqrt{D}$ is periodic with odd period $\ell = 1$, so the fundamental solution is $(x, y) = (p_{2\ell-1}, q_{2\ell-1}) = (p_1, q_1)$. The first convergent is

$$\frac{p_1}{q_1} = n + \frac{1}{2n} = \frac{2n^2+1}{2n}.$$

Thus, the fundamental solution is $(x, y) = (2n^2 + 1, 2n)$.

30.3. (a) We have $\sqrt{34} = [5; \overline{1, 4, 1, 10}]$. Using this, we can show that the fundamental solution is $(x, y) = (p_3, q_3) = (35, 6)$.

(b) If $x^2 - 34y^2 = -1$ has an integer solution then (by switching signs if necessary) it will have a positive integer solution hence it will have a minimal positive solution (p, q). Then, by the fact given in the problem and by part (a),

$$p^2 + 34q^2 = 35 \quad \text{and} \quad 2pq = 6.$$

The second equation implies that $(p, q) \in \{(3, 1), (1, 3)\}$. However, neither of these satisfy the first equation. We conclude that $x^2 - 34y^2 = -1$ cannot have any integer solutions.

Lecture 31

Exercises

31.2. Here is the full proof.

If (a, b, c) is a positive solution to $x^4 + y^4 = z^2$ then (a^2, b^2, c) is a Pythagorean triple. If d is a common divisor of a^2, b^2, c then we can just factor it out and obtain a smaller solution; so we may as well assume that (a, b, c^2) is a primitive triple. In particular, we may assume that a and c are odd and b is even.

Thus, by our classification of primitive Pythagorean triples, we have

$$a^2 = n^2 - m^2, \quad b^2 = 2nm \quad \text{and} \quad c = n^2 + m^2$$

where $n > m$ are coprime positive integers of different parity. Now, $a^2 + m^2 = n^2$ so (a, m, n) is a Pythagorean triple. It is also clearly primitive because if p divides a, m, n then d will divide $b = 2nm$ and $c = n^2 + m^2$ contradicting the fact that (a, b, c^2) is primitive. Since (a, m, n) is primitive and since a is odd, it follows that m is even and n is odd.

Appealing to our classification of Pythagorean triples once again, we have

$$a = u^2 - v^2, \quad m = 2uv \quad \text{and} \quad n = u^2 + v^2.$$

for some coprime $u, v \in \mathbb{Z}_{>0}$. Hence $b^2 = 2mn = 4uv(u^2+v^2)$. Observe that u, v, u^2+v^2 are pairwise coprime. Thus, since the product $uv(u^2 + v^2) = (b/2)^2$ is a square (note that $b/2 \in \mathbb{Z}$ because b is even), it follows that each of u, v and $u^2 + v^2$ is a square, say $u = (a')^2$, $v = (b')^2$ and $u^2 + v^2 = (c')^2$, where $a', b', c' \in \mathbb{Z}_{>0}$.

Note that $(a')^4 + (b')^4 = u^2 + v^2 = (c')^2$, so (a', b', c') is a positive integer solution to $x^4 + y^4 = z^2$. Finally, we have

$$c = n^2 + m^2 \geq n^2 = (u^2 + v^2)^2 = (c')^4 > c'.$$

This completes our descent.

31.3. If $n > 4$ is divisible by an odd prime p then $n = pk$ and hence

$$x^n + y^n = z^n \quad \iff \quad (x^k)^p + (y^k)^p = (z^k)^p.$$

So if there are no solutions to FLT with exponent p, then there are no solutions to FLT with exponent n.

On the other hand, if $n > 4$ is not divisible by any odd prime, then $n = 2^k$ must be divisible by 4. The same kind of argument used above applies here.

Practice Problems

31.1. *No solution provided. Margin too narrow.*