An Explicit Approach to Elementary Number Theory

William Stein

Math 124 HARVARD UNIVERSITY **Fall 2001**

Contents

Chapter 1

Introduction

1.1 Who is Teaching this Course?

I am *William Stein*. Come see me during my office hours, which are Wednesdays and Fridays, 2:00–3:00.

Quick Bio: I received a Ph.D. from Berkeley just over a year ago, where I worked with Hendrik Lenstra, Ken Ribet, and Robert Coleman. After graduating, I visited math institutes in Europe, Australia, and Asia and was a postdoctoral fellow here at Harvard. Now I am a Benjamin Peirce Assistant Professor. Lucky for you, my research specialty is number theory, with a focus on computing with "elliptic curves and modular forms".

1.2 Evaluation

- In-class midterm on October 17 (20% of grade)
- Homework every Wednesday (40% of grade)
- Take-home final (40% of grade)

1.3 What is this Course About?

See the lecture plan. The main ideas include:

1.3.1 Factorization

Do you remember writing whole numbers as products of primes? For example,

$$12 = 2 \times 2 \times 3.$$

Can this sort of thing always be done? Is it really hard or really easy? For example, is factoring social security numbers "trivial" or hopeless? In fact, it's trivial; even my wristwatch can do it!! (Mine might be the only wristwatch in the world that can factor social security numbers, but that's another story.) What about bigger numbers?

These questions are important to your everyday life. If somebody out there secretly knows how to factor 200-digit numbers quickly, then that person could easily read you credit card number and expiration date when you send it to `amazon.com`.

1.3.2 Congruences and Public-key Cryptography

Two numbers a and b are *congruent modulo another number* n if $a = b + nk$ for some integer k. That a and b are congruent just means you can "get from a to b on the number line" by adding or subtracting lots of copies of n. For example, $14 \equiv 2 \pmod{12}$ since $14 = 2 + 12 \cdot 1$.

$$\mathbb{Z}/n\mathbb{Z} = \{ \text{ equivalence classes of numbers modulo } n \ \}.$$

Your web browser's "secret code language" uses arithmetic in $\mathbb{Z}/pq\mathbb{Z}$ to send messages in broad weblight to `amazon.com`. How can this possibly be safe!? You will find out exactly what is going on.

1.3.3 Computers

Computers make the study of properties of whole numbers vastly more interesting. A computer is to a number theorist, like a telescope is to an astronomer. It would be a shame to teach an astronomy class without touching a telescope; likewise, it would be shame to teach this class without telling you how to look at the integers "through the lens of a computer".

1.3.4 Sums of Two Squares

I will tell you how to decide whether or not your order number is a sum of two squares. For example, an odd prime number is a sum of two squares if and only if when divided by 4 it leaves a remainder of 1. For example, 7 is not a sum of two squares, but 29 is.

1.3.5 Elliptic Curves

My experience is that elliptic curves are extraordinarily fun to study. Every such curve is like a whole galaxy in itself, just like the rational numbers are. An elliptic curve over $\mathbb{Q}$ is a curve that can be put in the form

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

where the cubic has distinct roots and $a, b \in \mathbb{Q}$. The amazing thing is that the set of pairs

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} : y^2 = x^3 + ax + b\} \cup \{\infty\}$$

has a natural structure of "group". In particular, this means that given two points on E, there is a way to "add" the two solutions together to get another solution.

Many exciting problems in number theory can be translated into questions about elliptic curves. For example, Fermat's Last Theorem, which asserts that $x^n+y^n = z^n$ has no positive integer solutions when $n > 2$ was proved using elliptic curves. Giving a method to decide which numbers are the area of a right triangle with rational side lengths has *almost*, but not quite, been solved using elliptic curves.

The central question about elliptic curves is *The Birch and Swinnerton-Dyer Conjecture* which gives a simple conjectural criterion to decide whether or not $E(\mathbb{Q})$ is infinite (and more). Proving the BSD conjecture is one of the Clay Math Institute's million dollar prize problems. I'll tell you what this conjecture is.

Chapter 2

Prime Factorization

2.1 Prime Numbers

We call positive whole numbers the *natural numbers* and denote them by $\mathbb{N}$. Thus

$$\mathbb{N} = \{1, 2, 3, 4, \ldots\}.$$

We call all the whole numbers, both positive and negative, the *integers*, and write

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

They are denoted by $\mathbb{Z}$ because the German word for the integers is "**Z**ahlen" (and 19th century German number theorists rocked).

Definition 2.1.1. If $a, b \in \mathbb{Z}$ then "a divides b" if $ac = b$ for some $c \in \mathbb{Z}$.

To save time, we write

$$a \mid b.$$

For example, $2 \mid 6$ and $389 \mid 97734562907$. Also, everything divides 0.

Definition 2.1.2. A natural number $p > 1$ is a *prime* if 1 and p are the only divisors of p in $\mathbb{N}$. I.e., if $a \mid p$ implies $a = 1$ or $a = p$.

Primes:

$$2, 3, 5, 7, 11, \ldots, 389, \ldots, 2003, \ldots$$

Composites:

$$4, 6, 8, 9, 10, 12, \ldots, 666 = 2 \cdot 3^2 \cdot 37, \ldots, 2001 = 3 \cdot 23 \cdot 29, \ldots$$

Primes are "primal"—every natural number is built out of prime numbers.

Theorem 2.1.3 (The Fundamental Theorem of Arithmetic). *Every positive integer can be written as a product of primes, and this expression is unique (up to order).*

Warning: This theorem is harder to prove than I first thought it should be. Why?

First, we are lucky that there are any primes at all: if the natural numbers are replaced by the positive rational numbers then there are no primes; e.g., $2 = \frac{1}{2} \cdot 4$, so $\frac{1}{2} \mid 2$.

Second, we are fortunate to have *unique* factorization in $\mathbb{Z}$. In other "rings", such as $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$, unique factorization can fail. In $\mathbb{Z}[\sqrt{-5}]$, the number 6 factors in two different ways:

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

If you are worried about whether or not 2 and 3 are "prime", read this: If $2 = (a + b\sqrt{-5}) \cdot (c + d\sqrt{-5})$ with neither factor equal to ± 1, then taking norms implies that

$$4 = (a^2 + 5b^2) \cdot (c^2 + 5d^2),$$

with neither factor 1. Theorem 2.1.3 implies that $2 = a^2 + 5b^2$, which is impossible. Thus 2 is "prime" in the (nonstandard!) sense that it has no divisors besides ± 1 and ± 2. A similar argument shows that 3 has no divisors besides ± 1 and ± 3. On the other hand, as you will learn later, 2 should not be considered prime, because the *ideal* generated by 2 in $\mathbb{Z}[\sqrt{-5}]$ is not prime. We have $(1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}) = 6 \in (2)$, but neither $1 + \sqrt{-5}$ nor $1 - \sqrt{-5}$ is in (2). We also note that $(1 + \sqrt{-5})$ does not factor. If $(1 + \sqrt{-5}) = (a + b\sqrt{-5}) \cdot (c + d\sqrt{-5})$, then, upon taking norms,

$$2 \cdot 3 = (a^2 + 5b^2) \cdot (c^2 + 5d^2),$$

which is impossible.

2.2 Greatest Common Divisors

Let a and b be two integers. The *greatest common divisor* of a and b is the biggest number that divides both of them. We denote it by "$\gcd(a, b)$". Thus,

Definition 2.2.1.

$$\gcd(a, b) = \max\{d : d \mid a \text{ and } d \mid b\}.$$

Warning: We define $\gcd(0, 0) = 0$, instead of "infinity".

Here are a few gcd's:

$$\gcd(1, 2) = 1, \quad \gcd(0, a) = \gcd(a, 0) = a, \quad \gcd(3, 27) = 3, \quad \gcd(2261, 1275) = ?$$

Warning: In Davenport's book, he denotes our gcd by HCF and calls it the "highest common factor". I will use the notation gcd because it is much more common.

2.2.1 Euclid's Algorithm for Computing GCDs

Can we easily compute something like $\gcd(2261, 1275)$? Yep. Watch closely:

$$2261 = 1 \cdot 1275 + 986.$$

Notice that if a number d divides both 2261 and 1275, then it automatically divides 986, and of course d divides 1275. Also, if a number divides both 1275 and 986, then it has got to divide 2261 as well! So we have made progress:

$$\gcd(2261, 1275) = \gcd(1275, 986)$$

Let's try again:

$$1275 = 1 \cdot 986 + 289,$$

so $\gcd(1275, 986) = \gcd(986, 289)$. Just keep at it:

$$\begin{aligned} 986 &= 3 \cdot 289 + 119 \\ 289 &= 2 \cdot 119 + 51 \\ 119 &= 2 \cdot 51 + 17. \end{aligned}$$

Thus $\gcd(2261, 1275) = \cdots = \gcd(51, 17)$, which is 17 because $17 \mid 51$, so

$$\gcd(2261, 1275) = 17.$$

Cool. Aside from tedious arithmetic, that was quick and very mechanical.

The Algorithm: That was an illustration of **Euclid's algorithm**. You just "Divide and switch."

More formally, fix $a, b \in \mathbb{N}$ with $a > b$. Using "divide with quotient and remainder", write $a = bq + r$, with $0 \leq r < b$. Then, just as above,

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

Let $a_1 = b$, $b_1 = r$, and repeat until $r = 0$. Soon enough we have computed $\gcd(a, b)$.

Here's are two more examples:

Example 2.2.2. Set $a = 15$ and $b = 6$.

$$\begin{array}{rcll} 15 &=& 6 \cdot 2 + 3 & \gcd(15, 6) = \gcd(6, 3) \\ 6 &=& 3 \cdot 2 + 0 & \gcd(6, 3) = \gcd(3, 0) = 3 \end{array}$$

We can just as easily do an example that is "10 times as hard":

Example 2.2.3. Set $a = 150$ and $b = 60$.

$$\begin{array}{rcll} 150 &=& 60 \cdot 2 + 30 & \gcd(150, 60) = \gcd(60, 30) \\ 60 &=& 30 \cdot 2 + 0 & \gcd(60, 30) = \gcd(30, 0) = 30 \end{array}$$

With Euclid's algorithm in hand, we can prove that if a prime divides the product of two numbers, then it has got to divide one of them. This result is the *key* to proving that prime factorization is unique.

Theorem 2.2.4 (Euclid). *Let p be a prime and $a, b \in \mathbb{N}$. If $p \mid ab$ then $p \mid a$ or $p \mid b$.*

Proof. If $p \mid a$ we are done. If $p \nmid a$ then $\gcd(p, a) = 1$, since only 1 and p divide p. Stepping through the Euclidean algorithm from above, we see that $\gcd(pb, ab) = b$. At each step, we simply multiply the equation through by b. Since $p \mid pb$ and, by hypothesis, $p \mid ab$, it follows that $p \mid \gcd(pb, ab) = b$. $\square$

2.3 Numbers Do Factor

Let $n = 1275$, and recall from above that $17 \mid 1275$, so n is definitely composite, $n = 17 \cdot 75$. Next, 75 is $5 \cdot 15 = 5 \cdot 5 \cdot 3$. So, finally, $1275 = 3 \cdot 5 \cdot 5 \cdot 17$.

Now suppose n is any positive number. Then, just as above, n can be written as a product of primes:

- If n is prime, we are done.
- If n is composite, then $n = ab$ with $a, b < n$. By induction, a, b are products of primes, so n is also a product of primes.

What if we had done something differently when breaking 1275 apart as a product of primes? Could the primes that show up be different? Why not just try? We have $1275 = 5 \cdot 255$. Now $255 = 5 \cdot 51$ and $51 = 17 \cdot 3$, so everything turned out the same. Will it always?

Incidently, there's an open problem nearby:

Unsolved Question: Is there an algorithm which can factor any given integer n so quickly that its "running time" is bounded by a polynomial function of the number of decimal digits of n.

I think most people would guess "no", but nobody has yet proved that it can't be done (and told everyone...). If there were such an algorithm, then the cryptosystem that I use to send my girlfriend private emails would probably be easily broken.

2.3.1 A \$10,000 Challenge

If you factor the following 174-digit number, affectionality known as "RSA-576", then the RSA company will give you TEN THOUSAND DOLLARS!!!

> 188198812920607963838697239461650439807163563379417382700763356
> 422988859715234665485319060606504743045317388011303396716199692
> 321205734031879550656996221305168759307650257059

This number is called RSA-576, since it has 576 *binary* digits. See

http://www.rsasecurity.com/rsalabs/challenges/factoring/index.html

for more details.

2.4 The Fundamental Theorem of Arithmetic

We can now prove Theorem 2.1.3. The idea is simple. Suppose we have two factorization. Use Theorem 2.2.4 to cancel primes from each, one prime at a time. At the end of the game, we discover that the factorizations have to consist of exactly the same primes. The technical details, with all the p's and q's are given below:

Proof. We have

$$n = p_1 \cdot p_2 \cdots p_d,$$

with each p_i prime. Suppose that

$$n = q_1 \cdot q_2 \cdots q_m$$

is another expression of n as a product of primes. Since

$$p_1 \mid n = q_1 \cdot (q_2 \cdots q_m),$$

Euclid's theorem implies that $p_1 = q_1$ or $p_1 \mid q_2 \cdots q_m$. By induction, we see that $p_1 = q_i$ for some i.

Now cancel p_1 and q_i, and repeat the above argument. Eventually, we find that, up to order, the two factorizations are the same. $\square$

Chapter 3

Introduction to Computing and PARI

3.1 Introduction

> "The object of numerical computation is theoretical advance." – *Bryan Birch describing A. O. L. Atkin.*

Much progress in number theory has been driven by attempts to prove conjectures. It's reasonably easy to play around with integers, see a pattern, and make a conjecture. Frequently proving the conjecture is *extremely difficult.* In this direction, computers help us to

- find more conjectures
- disprove conjectures
- increase our confidence in a conjecture

They also frequently help to solve a specific problem. For example, the following problem would be hopelessly tedious by hand. Here's an example of such a problem:

> Find all integer $n < 100$ that are the area of a right triangle with integer side lengths.[1]

This problem can be solved by a combination of very deep theorems, a few big computer computations, and a little luck.

3.2 Some Assertions About Primes

A computer can quickly "convince" you that many assertions about prime numbers are true. Here are three.

- *The polynomial $x^2 + 1$ takes on infinitely many prime values.*
  Let
  $$f(n) = \{x : x < n : x \text{ and } x^2 + 1 \text{ is prime } \}.$$

[1] We will discuss the "The Congruent Number Problem" in more depth later in this course.

With a computer, we quickly find that

$$f(10^2) = 19, \quad f(10^3) = 112, \quad f(10^4) = 841, \quad f(10^5) = 6656.$$

Surely $f(n)$ is unbounded! The PARI code to compute $f(n)$ is very simple:

```
? f(n) = s=0; for(x=1,n,if(isprime(x^2+1),s++)); s
? f(100)
%1 = 19
? f(1000)
%2 = 112
? f(10000)
%3 = 841
? f(100000)
%4 = 6656
```

- *Every even integer $n > 2$ is a sum of two primes.*
With a computer we find that this seems true

n	p	q
4	2	2
6	3	3
8	3	5
10	3	7
12	5	7

... and much further. In practice, it's easy to write an even number as a sum of two primes. Why should there be any weird even numbers out there for which this can't be done? PARI code to find p and q:

```
? gb(n) = forprime(p=2,n,if(isprime(n-p),return([p,n-p])));
? gb(4)
%7 = [2, 2]
? gb(6)
%8 = [3, 3]
? gb(100)
%9 = [3, 97]
? gb(1000)
%10 = [3, 997]
? gb(570)          \\ takes no time at all!
%11 = [7, 563]
```

- *There are infinitely many primes p such that $p + 2$ is also prime.*
Let $t(n) = \#\{p : p \leq n \text{ and } p + 2 \text{ is prime }\}$. Using a computer we quickly find that

$$t(10^2) = 8, \quad t(10^3) = 35, \quad t(10^4) = 205, \quad t(10^5) = 1024.$$

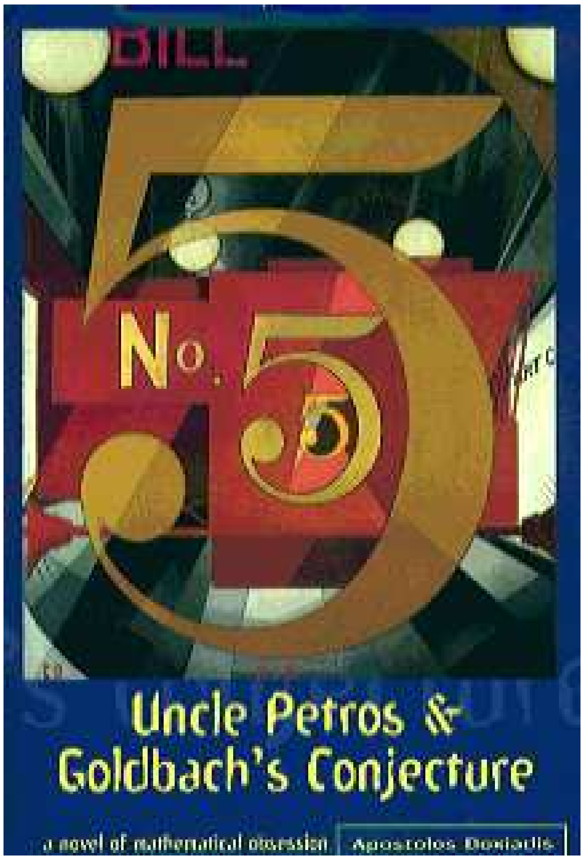
The PARI code to compute $t(n)$ is very simple:

```
? t(n) = s=0; forprime(p=2,n,if(isprime(p+2),s++)); s
? t(10^2)
%12 = 8
? t(10^3)
%13 = 35
? t(10^4)
%14 = 205
? t(10^5)
%15 = 1224
```

Surely $t(n)$ keeps getting bigger!!

As it turns out, these three assertions are *all* OLD famous extremely difficult unsolved problems! Anyone who proves one of them will be very famous.

Assertion 2 is called "The Goldbach Conjecture"; Goldbach reformulated it in a letter to Euler in 1742. It's featured in the following recent novel:



The publisher of that novel offers a MILLION dollar prize for the solution to the Goldbach conjecture:

`http://www.faber.co.uk/faber/million_dollar.asp?PGE=&ORD=faber&TAG=&CID=`

The Goldbach conjecture is true for all $n < 4 \cdot 10^{14}$, see

`http://www.informatik.uni-giessen.de/staff/richstein/ca/Goldbach.html`

Assertion 3 is the "Twin Primes Conjecture". According to

`http://perso.wanadoo.fr/yves.gallot/primes/chrrcds.html#twin`

on May 17, 2001, David Underbakke and Phil Carmody discovered a 32220 digits twin primes record with a set of different programs: $318032361 \cdot 2^{107001} \pm 1$. This is the current "world record".

With a computer, even if you can't solve one of these "Grand Challenge" problems, at least you can perhaps work very hard and prove it for more cases than anybody before you, especially since computers keep getting more powerful. This can be very fun, especially as you search for a more efficient algorithm to extend the computations.

3.3 Some Tools for Computing

Calculator: A TI-89 can deal with integers with 1000s of digits, factor, and do most basic number theory. I am not aware if anyone has programmed basic "elliptic curve" computations into this calculator, but it could be done.

Mathematica and Maple: Both are commercial, but they are very powerful, can draw pretty pictures, and there are elliptic curve packages available for each (`apecs` for Maple, and something by Silverman for Mathematica).

PARI: Free, open source, excellent for our course, simple, runs on Macs, MS Windows, Linux, etc.

MAGMA: Huge, non-free but nonprofit, what I usually use for my research. I can legally give you a Linux executable if you are registered for 124.

My Wristwatch: Perhaps the only wristwatch in the world that can factor your social security number? `:-)`

3.4 Getting Started with PARI

3.4.1 Documentation

The documentation for PARI is available at

`http://modular.fas.harvard.edu/docs/`

Some PARI documentation:

1. **Installation Guide:** Help for setting up PARI on a UNIX computer.
2. **Tutorial:** 42-page tutorial that starts with `2 + 2`.
3. **User's Guide:** 226-page reference manual; describes every function
4. **Reference Card:** hard to print, so I printed it for you (handout)

3.4.2 A Short Tour

```
$ gp
Appele avec : /usr/local/bin/gp -s 10000000 -p 500000 -emacs

                  GP/PARI CALCULATOR Version 2.1.1 (released)
                i686 running linux (ix86 kernel) 32-bit version
                (readline v4.2 enabled, extended help available)

                       Copyright (C) 2000 The PARI Group

PARI/GP is free software, covered by the GNU General Public License, and
comes WITHOUT ANY WARRANTY WHATSOEVER.

Type ? for help, \q to quit.
Type ?12 for how to get moral (and possibly technical) support.

   realprecision = 28 significant digits
   seriesprecision = 16 significant terms
   format = g0.28

parisize = 10000000, primelimit = 500000
? \\ this is a comment
? x = 571438063;
? print(x)
571438063
? x^2+17
%2 = 326541459845191986
? factor(x)
%3 =
[7 1]

[81634009 1]
? gcd(x,56)
%5 = 7
? x^20
%6 = 137842550376658549303577840675412507732229154958280209139358
845011397194393261309756046226816251290119446623115998366224179
760816483100648674388195744425584150472890085928660801
```

3.4.3 Help in PARI

```
? ?
Help topics:
  0: list of user-defined identifiers (variable, alias, function)
  1: Standard monadic or dyadic OPERATORS
  2: CONVERSIONS and similar elementary functions
```

```
  3: TRANSCENDENTAL functions
  4: NUMBER THEORETICAL functions
  5: Functions related to ELLIPTIC CURVES
  6: Functions related to general NUMBER FIELDS
  7: POLYNOMIALS and power series
  8: Vectors, matrices, LINEAR ALGEBRA and sets
  9: SUMS, products, integrals and similar functions
 10: GRAPHIC functions
 11: PROGRAMMING under GP
 12: The PARI community

Further help (list of relevant functions): ?n (1<=n<=11).
Also:
  ? functionname (short on-line help)
  ?\             (keyboard shortcuts)
  ?.             (member functions)
Extended help looks available:
  ??             (opens the full user's manual in a dvi previewer)
  ??  tutorial   (same with the GP tutorial)
  ??  refcard    (same with the GP reference card)

  ??  keyword    (long help text about "keyword" from the user's manual)
  ??? keyword    (a propos: list of related functions).
? ?4

addprimes     bestappr      bezout        bezoutres     bigomega
binomial      chinese       content       contfrac      contfracpnqn
core          coredisc      dirdiv        direuler      dirmul
divisors      eulerphi      factor        factorback    factorcantor
factorff      factorial     factorint     factormod     ffinit
fibonacci     gcd           hilbert       isfundamental isprime
ispseudoprime issquare      issquarefree  kronecker     lcm
moebius       nextprime     numdiv        omega         precprime
prime         primes        qfbclassno    qfbcompraw    qfbhclassno
qfbnucomp     qfbnupow      qfbpowraw     qfbprimeform  qfbred
quadclassunit quaddisc      quadgen       quadhilbert   quadpoly
quadray       quadregulator quadunit      removeprimes  sigma
sqrtint       znlog         znorder       znprimroot    znstar

? ?gcd
gcd(x,y,{flag=0}): greatest common divisor of x and y. flag is optional, and
can be 0: default, 1: use the modular gcd algorithm (x and y must be
polynomials), 2 use the subresultant algorithm (x and y must be polynomials).

? ??gcd
\\ if set up correctly, brings up the typeset section from the manual on gcd
```

We will discuss writing more complicated PARI programs on October 10.

Chapter 4

The Sequence of Prime Numbers

This lecture is about the following three questions:

1. Are there infinitely many primes? (yes)
2. Are there infinitely many primes of the form $ax + b$? (yes, if $\gcd(a, b) = 1$)
3. How many primes are there? (asymptotically $x/\log(x)$ primes less than x)

4.1 There are infinitely many primes

Theorem 4.1.1 (Euclid). *There are infinitely many primes.*

Note that this is not obvious. There are completely reasonable rings where it is false, such as

$$R = \left\{ \frac{a}{b} : a, b \in \mathbb{Z} \text{ and } \gcd(b, 30) = 1 \right\}$$

There are exactly three primes in R, and that's it.

Proof of theorem. Suppose not. Let $p_1 = 2, p_2 = 3, \ldots, p_n$ be all of the primes. Let

$$N = 2 \times 3 \times 5 \times \cdots \times p_n + 1$$

Then $N \neq 1$ so, as proved in Lecture 2,

$$N = q_1 \times q_2 \times \cdots \times q_m$$

with each q_i prime and $m \geq 1$. If $q_1 \in \{2, 3, 5, \ldots, p_n\}$, then $N = q_1 a + 1$, so $q_1 \nmid N$, a contradiction. Thus our assumption that $\{2, 3, 5, \ldots, p_n\}$ are all of the primes is false, which proves that there must be infinitely many primes. $\square$

If we were to try a similar proof in R, we run into trouble. We would let $N = 2 \cdot 3 \cdot 5 + 1 = 31$, which is a unit, hence not a nontrivial product of primes.

Joke (Lenstra). "There are infinitely many composite numbers. *Proof:* Multiply together the first n primes and don't add 1."

According to

http://www.utm.edu/research/primes/largest.html

the largest known prime is

$$p = 2^{6972593} - 1,$$

which is a number having over two million[1] decimal digits. Euclid's theorem implies that there definitely *is* a bigger prime number. However, nobody has yet found it *and proved that they are right.* In fact, determining whether or not a number is prime is an extremely interesting problem. We will discuss this problem more later.

4.2 Primes of the form $ax + b$

Next we turn to primes of the form $ax + b$. We assume that $\gcd(a, b) = 1$, because otherwise there is no hope that $ax + b$ is prime *infinitely* often. For example, $3x + 6$ is only prime for one value of x.

Proposition 4.2.1. *There are infinitely many primes of the form* $4x - 1$.

Why might this be true? Let's list numbers of the form $4x - 1$ and underline the ones that are prime:

$$\underline{3}, \underline{7}, \underline{11}, 15, \underline{19}, \underline{23}, 27, \underline{31}, 35, 39, \underline{43}, \underline{47}, \ldots$$

It certainly looks plausible that underlined numbers will continue to appear. The following PARI program can be used to further convince you:

```
f(n, s=0) = for(x=1, n, if(isprime(4*x-1), s++); s
```

Proof. The proof is similar to the proof of Euclid's Theorem, but, for variety, I will explain it in a slightly different way.

Suppose $p_1, p_2, \ldots, p_n$ are primes of the form $4x - 1$. Consider the number

$$N = 4p_1 \times p_2 \times \cdots \times p_n - 1.$$

Then $p_i \nmid N$ for any i. Moreover, not every prime $p \mid N$ is of the form $4x + 1$; if they all were, then N would also be of the form $4x + 1$, which it is not. Thus there is a $p \mid N$ that is of the form $4x - 1$. Since $p \neq p_i$ for any i, we have found another prime of the form $4x - 1$. We can repeat this process indefinitely, so the set of primes of the form $4x - 1$ is infinite. $\square$

Example 4.2.2. Set $p_1 = 3$, $p_2 = 7$. Then

$$N = 4 \times 3 \times 7 - 1 = \underline{83}$$

is a prime of the form $4x - 1$. Next

$$N = 4 \times 3 \times 7 \times 83 - 1 = \underline{6971},$$

which is a again a prime of the form $4x - 1$. Again:

$$N = 4 \times 3 \times 7 \times 83 \times 6971 - 1 = 48601811 = 61 \times \underline{796751}.$$

[1] It has exactly 2098960 decimal digits.

This time 61 is a prime, but it is of the form $4x+1 = 4 \times 15 + 1$. However, 796751 is prime and $(796751 - (-1))/4 = 199188$. We are unstoppable

$$N = 4 \times 3 \times 7 \times 83 \times 6971 \times 796751 - 1 = \underline{5591} \times 6926049421.$$

This time the small prime, 5591, is of the form $4x - 1$ and the large one is of the form $4x + 1$. Etc!

Theorem 4.2.3 (Dirichlet). *Let a and b be integers with $\gcd(a, b) = 1$. Then there are infinitely many primes of the form $ax + b$.*

The proof is out of the scope of this course. You will probably see a proof if you take Math 129 from Cornut next semester.

4.3 How many primes are there?

There are infinitely many primes.

Can we say something more precise?

Let's consider a similar question:

Question 4.3.1. How many even integers are there?

Answer: *Half* of all integers.

Question 4.3.2. How many integers are there of the form $4x - 1$?

Answer: *One fourth* of all integers.

Question 4.3.3. How many perfect squares are there?

Answer: Zero percent of all numbers, in the sense that the limit of the proportion of perfect squares to all numbers converges to 0. More precisely,

$$\lim_{x \to \infty} \#\{n : n \leq x \text{ and } n \text{ is a perfect square }\}/x = 0,$$

since the numerator is roughly $\sqrt{x}$ and $\sqrt{x}/x \to 0$.

A better question is:

Question 4.3.4. How many numbers $\leq x$ are perfect squares, as a function of x?

Answer: Asymptotically, the answer is $\sqrt{x}$.

So a good question is:

Question 4.3.5. How many numbers $\leq x$ are prime?

Let

$$\pi(x) = \#\{\text{ primes } p \leq x\}.$$

For example,

$$\pi(6) = \#\{2, 3, 5\} = 3.$$

We can compute a few more values of $\pi(x)$ using PARI:

```
? pi(x, c=0) = forprime(p=2,x,c++); c;
? for(n=1,7,print(n*100,"\t",pi(n*100)))
100 25
200 46
300 62
400 78
500 95
600 109
700 125
```

Now draw a graph on the blackboard. It will look like a straight line...

Gauss spent some of his free time counting primes. By the end of his life, he had computed $\pi(x)$ for x up to 3 million.

$$\pi(3000000) = 216816.$$

(I don't know if Gauss got the right answer.) Gauss conjectured the following:

Theorem 4.3.6 (Hadamard, Vallée Poussin, 1896). *$\pi(x)$ is asymptotic to $x/\log(x)$, in the sense that*

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

I will not prove this theorem in this class. The theorem implies that $x/(\log(x) - a)$ can be used to approximate $\pi(x)$, for any a. In fact, $a = 1$ is the best choice.

```
? pi(x, c=0) = forprime(p=2,x,c++); c;
? for(n=1,10,print(n*1000,"\t",pi(n*1000),"\t",n*1000/(log(n*1000)-1)))
1000 168 169.2690290604408165186256278
2000 303 302.9888734545463878029800994
3000 430 428.1819317975237043747385740
4000 550 548.3922097278253264133400985
5000 669 665.1418784486502172369455815
6000 783 779.2698885854778626863677374
7000 900 891.3035657223339974352567759
8000 1007 1001.602962794770080754784281
9000 1117 1110.428422963188172310675011
10000 1229 1217.976301461550279200775705
```

Remark 4.3.7.

4.3.1 Counting Primes Today

People all over the world are counting primes, probably even as we speak. See, e.g.,

`http://www.utm.edu/research/primes/howmany.shtml`

`http://numbers.computation.free.fr/Constants/Primes/Pix/pixproject.html`

A huge computation:

$$\pi(10^{22}) = 201467286689315906290$$

(I don't know for sure if this is right...)

4.3.2 The Riemann Hypothesis

The function

$$\mathrm{Li}(x) = \int_2^x \frac{1}{\log(x)} dx.$$

is also a good approximation to $\pi(x)$.

The famous **Riemann Hypothesis** is equivalent to the assertion that

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

(This is another $1000000 prize problem.)

```
pi(10^22)      = 201467286689315906290
Li(10^22)      = 201467286691248261498.1505...    (using Maple)
Log(x)/(x-1)   = 201381995844659893517.7648...    (pari)
```

Chapter 5

Congruences

The point of this lecture:
Define the ring $\mathbb{Z}/n\mathbb{Z}$ of integers modulo n. Prove Fermat's little theorem, which asserts that if $\gcd(x,n) = 1$, then $x^{\varphi(n)} \equiv 1 \pmod{n}$.

5.1 Notation

Definition 5.1.1 (Congruence). Let $a, b \in \mathbb{Z}$ and $n \in \mathbb{N}$. Then

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

if $n \mid a - b$.

That is, there is $c \in \mathbb{Z}$ such that

$$nc = a - b.$$

One way I think about it: a is congruent to b modulo n, if we can get from b to a by adding multiples of n.

Congruence modulo n is an *equivalence relation*. Let

$$\mathbb{Z}/n\mathbb{Z} = \{ \text{ the set of equivalence classes } \}$$

The set $\mathbb{Z}/n\mathbb{Z}$ is a *ring*, the "ring of integers modulo n". It is the quotient of the ring $\mathbb{Z}$ by the ideal generated by n.

Example 5.1.2.

$$\mathbb{Z}/3\mathbb{Z} = \{\{\dots,-3,0,3,\dots\},\{\dots,-2,1,4,\dots\},\{\dots,-1,2,5,\dots\}\} = \{[0],[1],[2]\}$$

where we let $[a]$ denote the equivalence class of a.

5.2 Arithmetic Modulo N

Suppose a, $a', b\,b' \in \mathbb{Z}$ and

$$a \equiv a' \pmod{n}, \qquad b \equiv b' \pmod{n}.$$

Then

$$a + b \equiv a' + b' \pmod{n} \tag{5.1}$$
$$a \times b \equiv a' \times b' \pmod{n} \tag{5.2}$$

So it makes sense to define $+$ and $\times$ by $[a] + [b] = [a + b]$ and $[a] \times [b] = [a \times b]$.

5.2.1 Cancellation

Proposition 5.2.1. *If* $\gcd(c, n) = 1$ *and*

$$ac \equiv bc \pmod{n}$$

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

Proof. By definition

$$n \mid ac - bc = (a - b)c.$$

Since $\gcd(n, c) = 1$, it follows that $n \mid a - b$, so

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

as claimed. □

5.2.2 Rules for Divisibility

Proposition 5.2.2. *A number* $n \in \mathbb{Z}$ *is divisible by* 3 *if and only if the sum of the digits of* n *is divisible by* 3.

Proof. Write

$$n = a + 10b + 100c + \cdots .$$

Since $10 \equiv 1 \pmod{3}$,

$$n = a + 10b + 100c + \cdots \equiv a + b + c + \cdots \pmod{3},$$

from which the proposition follows. □

Similarly, you can find rules for divisibility by 5, 9 and 11. What about divisibility by 7?

5.3 Linear Congruences

Definition 5.3.1 (Complete Set of Residues). A *complete set of residues* modulo n is a subset $R \subset \mathbb{Z}$ of size n whose reductions modulo n are distinct. In other words, a complete set of residues is a choice of representive for each equivalence class in $\mathbb{Z}/n\mathbb{Z}$.

Some examples:

$$R = \{0, 1, 2, \ldots, n - 1\}$$

is a complete set of residues modulo n. When $n = 5$, a complete set of residues is

$$R = \{0, 1, -1, 2, -2\}.$$

Lemma 5.3.2. *If R is a complete set of residues modulo n and $a \in \mathbb{Z}$ with $\gcd(a, n) = 1$, then $aR = \{ax : x \in R\}$ is also a complete set of residues.*

Proof. If $ax \equiv ax' \pmod{n}$ with $x, x' \in R$, then Proposition 5.2.1 implies that $x \equiv x' \pmod{n}$. Because R is a complete set of residues, this implies that $x = x'$. Thus the elements of aR have distinct reductions modulo n. It follows, since $\#aR = n$, that aR is a complete set of residues modulo n. □

Definition 5.3.3 (Linear Congruence). A *linear congruence* is an equation of the form

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

Proposition 5.3.4. *If $\gcd(a, n) = 1$, then the equation*

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

must have a solution.

Proof. Let R be a complete set of residues modulo n (for example, $R = \{0, 1, \ldots, n-1\}$). Then by Lemma 5.3.2, aR is also a complete set of residues. Thus there is an element $ax \in aR$ such that $ax \equiv b \pmod{n}$, which proves the proposition. □

The point in the proof is that left multiplication by a defines a map $\mathbb{Z}/n\mathbb{Z} \hookrightarrow \mathbb{Z}/n\mathbb{Z}$, which must be surjective because $\mathbb{Z}/n\mathbb{Z}$ is finite.

Illustration:

$$2x \equiv 3 \pmod{7}$$

Set $R = \{0, 1, 2, 3, 4, 5, 6\}$. Then

$$2R = \{0, 2, 4, 6, 8 \equiv 1, 10 \equiv 3, 12 \equiv 5\},$$

so $2 \cdot 5 \equiv 3 \pmod{7}$.

Warning:
Note that the equation $ax \equiv b \pmod{n}$ might have a solution even if $\gcd(a, n) \neq 1$. To construct such examples, let a be any divisor of n, x any number, and set $b = ax$. For example, $2x \equiv 6 \pmod{8}$ has a solution!

5.4 Fermat's Little Theorem

Definition 5.4.1 (Order). Let $n \in \mathbb{N}$ and $x \in \mathbb{Z}$ with $\gcd(x, n) = 1$. The *order* of x modulo n is the smallest $m \in \mathbb{N}$ such that

$$x^m \equiv 1 \pmod{n}.$$

We must show that this definition makes sense. To do so, we verify that such an m exists. Consider $x, x^2, x^3, \ldots$ modulo n. There are only finitely many residue classes modulo n, so we must eventually find two integers i, j with $i < j$ such that

$$x^i \equiv x^j \pmod{n}.$$

Since $\gcd(x, n) = 1$, Proposition 5.2.1 implies that we can cancel x's and conclude that

$$x^{j-i} \equiv 1 \pmod{n}.$$

Definition 5.4.2 (Euler Phi function). Let

$$\varphi(n) = \#\{a \in \mathbb{N} : a \leq n \text{ and } \gcd(a, n) = 1\}.$$

For example,

$$\begin{aligned} \varphi(1) &= \#\{1\} = 1, \\ \varphi(5) &= \#\{1, 2, 3, 4\} = 4, \\ \varphi(12) &= \#\{1, 5, 7, 11\} = 4. \end{aligned}$$

If p is any prime number then

$$\varphi(p) = \#\{1, 2, \ldots, p-1\} = p - 1.$$

Theorem 5.4.3 (Fermat's Little Theorem). *If* $\gcd(x, n) = 1$, *then*

$$x^{\varphi(n)} \equiv 1 \pmod{n}.$$

Proof. Let

$$P = \{a : 1 \leq a \leq n \text{ and } \gcd(a, n) = 1\}.$$

In the same way that we proved Lemma 5.3.2, we see that the reductions modulo n of the elements of xP are exactly the same as the reductions of the elements of P. Thus

$$\prod_{a \in P} (xa) = \prod_{a \in P} a \pmod{n},$$

since the products are over exactly the same numbers modulo n. Now cancel the a's on both sides to get

$$x^{\#P} \equiv 1 \pmod{n},$$

as claimed. □

5.4.1 Group-theoretic Interpretation

The set of invertible elements of $\mathbb{Z}/n\mathbb{Z}$ is a group

$$(\mathbb{Z}/n\mathbb{Z})^* = \{[a] \in \mathbb{Z}/n\mathbb{Z} : \gcd(a, n) = 1\}.$$

This group has order $\varphi(n)$. Theorem 5.4.3 asserts that the order of an element of $(\mathbb{Z}/n\mathbb{Z})^\times$ divides the order $\varphi(n)$ of $(\mathbb{Z}/n\mathbb{Z})^\times$. This is a special case of the more general theorem that if G is a finite group and $g \in G$, then the order of g divide $\#G$.

5.5 What happened?

Take out a piece of paper and answer the following two questions:

1. What is a central idea that you learned in this lecture?
2. What part of this lecture did you find murky?

Chapter 6

Congruences, Part II

Key Ideas Today

- Wilson's theorem
- Chinese Remainder Theorem
- Multiplicativity of φ

6.1 Wilson's Theorem

Theorem 6.1.1 (John Wilson's theorem, from the 1770s). *An integer $p > 1$ is prime if and only if*

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

Example 6.1.2.

```
? p=3
%1 = 3
? (p-1)! % 3
%2 = 2
? p=17
%3 = 17
? (p-1)!
%4 = 20922789888000
? (p-1)! % p
%5 = 16
```

Proof. We first **assume that** p **is prime** and prove that $(p-1)! \equiv -1 \pmod{p}$. If $a \in \{1, 2, \ldots, p-1\}$ then the equation

$$ax \equiv 1 \pmod{p}$$

has a unique solution $a' \in \{1, 2, \ldots, p-1\}$. If $a = a'$, then $a^2 \equiv 1 \pmod{p}$, so $p \mid a^2 - 1 = (a-1)(a+1)$, so $p \mid (a-1)$ or $p \mid (a+1)$, so $a \in \{1, -1\}$. We can thus pair off the elements of $\{2, 3, \ldots, p-2\}$, each with its inverse. Thus

$$2 \cdot 3 \cdot \cdots \cdot (p-2) \equiv 1 \pmod{p}.$$

Multiplying both sides by $p-1$ proves that $(p-1)! \equiv -1 \pmod{p}$.

Next we **assume that** $(p-1)! \equiv -1 \pmod{p}$ and prove that p must be prime. Suppose not, so that p is a composite number ≥ 4. Let ℓ be a prime divisor of p. Then $\ell < p$, so $\ell \mid (p-1)!$. Also,

$$\ell \mid p \mid ((p-1)! - 1).$$

This is a contradiction, because a prime can't divide a number a and also divide $a-1$, since it would then have to divide $a-(a-1)=1$. $\square$

Example 6.1.3. When $p = 17$, we have

$$2 \cdot 3 \cdot \cdots \cdot 15 = (2 \cdot 9) \cdot (3 \cdot 6) \cdot (4 \cdot 13) \cdot (5 \cdot 7) \cdot (8 \cdot 15) \cdot (10 \cdot 12) \cdot (14 \cdot 11) \equiv 1 \pmod{17},$$

where we have paired up the numbers a, b for which $ab \equiv 1 \pmod{17}$.

Let's test Wilson's Theorem in PARI:

```
? wilson(n) = Mod((n-1)!,n) == Mod(-1,n)
? wilson(5)
%9 = 1
? wilson(10)
%10 = 0
? wilson(389)
%11 = 1
? wilson(2001)
%12 = 0
```

Warning: In practice, this is a horribly inefficient way to check whether or not a number is prime.

6.2 The Chinese Remainder Theorem

Sun Tsu Suan-Ching (4th century AD):

> There are certain things whose number is unknown. Repeatedly divided by 3, the remainder is 2; by 5 the remainder is 3; and by 7 the remainder is 2. What will be the number?

In modern notation, Sun is asking us to solve the following system of equations:

$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7} \end{aligned}$$

The Chinese Remainder Theorem asserts that a solution to Sun's question exists, and the proof gives a method to find a solution.

Theorem 6.2.1 (The Chinese Remainder Theorem). *Let $a, b \in \mathbb{Z}$ and $n, m \in \mathbb{N}$ such that $\gcd(n, m) = 1$. Then there exists $x \in \mathbb{Z}$ such that*

$$\begin{aligned} x &\equiv a \pmod{m} \\ x &\equiv b \pmod{n} \end{aligned}$$

Proof. The equation
$$tm \equiv b - a \pmod{n}$$
has a solution t since $\gcd(m, n) = 1$. Set $x = a + tm$. We next verify that x is a solution to the two equations. Then
$$x \equiv a + (b - a) \equiv b \pmod{n},$$
and
$$x = a + tm \equiv a \pmod{m}.$$
□

Now we can solve Sun's problem:
$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5} \\ x &\equiv 2 \pmod{7}. \end{aligned}$$
First, we use the theorem to find a solution to the pair of equations
$$\begin{aligned} x &\equiv 2 \pmod{3} \\ x &\equiv 3 \pmod{5}. \end{aligned}$$
Set $a = 2$, $b = 3$, $m = 3$, $n = 5$. Step 1 is to find a solution to $t \cdot 3 \equiv 3 - 2 \pmod{5}$. A solution is $t = 2$. Then $x = a + tm = 2 + 2 \cdot 3 = 8$. Since any x' with $x' \equiv x \pmod{15}$ is also a solution to those two equations, we can solve all three equations by finding a solution to the pair of equations
$$\begin{aligned} x &\equiv 8 \pmod{15} \\ x &\equiv 2 \pmod{7}. \end{aligned}$$
Again, we find a solution to $t \cdot 15 \equiv 2 - 8 \pmod{7}$. A solution is $t = 1$, so
$$x = a + tm = 8 + 15 = 23.$$
Note that there are other solutions. Any $x' \equiv x \pmod{3 \cdot 5 \cdot 7}$ is also a solution; e.g., $23 + 3 \cdot 5 \cdot 7 = 128$.

We can also solve Sun's problem in PARI:

```
? chinese(Mod(2,3),Mod(3,5))
%13 = Mod(8, 15)
? chinese(Mod(8,15),Mod(2,7))
%14 = Mod(23, 105)
```

6.3 Multiplicative Functions

Definition 6.3.1. A function $f : \mathbb{N} \to \mathbb{Z}$ is *multiplicative* if, whenever $m, n \in \mathbb{N}$ and $\gcd(m, n) = 1$, we have
$$f(mn) = f(m) \cdot f(n).$$

Recall that the *Euler φ-function* is

$$\varphi(n) = \#\{a : 1 \leq a \leq n \text{ and } \gcd(a,n) = 1\}.$$

Proposition 6.3.2. *φ is a multiplicative function.*

Proof. Suppose that $m, n \in \mathbb{N}$ and $\gcd(m,n) = 1$. Consider the map

$$\{c : 1 \leq c \leq mn \text{ and } \gcd(c, mn) = 1\} \xrightarrow{f}$$
$$\{a : 1 \leq a \leq m \text{ and } \gcd(a,m) = 1\} \times \{b : 1 \leq b \leq n \text{ and } \gcd(b,n) = 1\}\}$$

defined by

$$f(c) = (c \bmod m, \quad c \bmod n).$$

The map f is injective: If $f(c) = f(c')$, then $m \mid c - c'$ and $n \mid c - c'$, so, since $\gcd(n,m) = 1$, $nm \mid c - c'$, so $c = c'$.

The map f is surjective: Given a, b with $\gcd(a,m) = 1$, $\gcd(b,n) = 1$, the Chinese Remainder Theorem implies that there exists c with $c \equiv a \pmod{m}$ and $c \equiv b \pmod{n}$. We may assume that $1 \leq c \leq nm$, ans since $\gcd(a,m) = 1$ and $\gcd(b,n) = 1$, we must have $\gcd(c, nm) = 1$. Thus $f(c) = (a,b)$.

Because f is a bijection, the set on the left has the same size as the product set on the right. Thus

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

□

Example 6.3.3. The proposition makes it easier to compute $\varphi(n)$. For example,

$$\varphi(12) = \varphi(2^2) \cdot \varphi(3) = 2 \cdot 2 = 4.$$

Also, for $n \geq 1$, we have

$$\varphi(p^n) = p^n - \frac{p^n}{p},$$

since $\varphi(p^n)$ is the number of numbers less than p^n minus the number of those that are divisible by p. Thus, e.g.,

$$\varphi(389 \cdot 11^2) = 388 \cdot (11^2 - 11) = 388 \cdot 110 = 42680.$$

The φ function is also available in PARI:

```
? eulerphi(389*11^2)
%15 = 42680
```

Question 6.3.4. Is computing $\varphi(1000 \text{ digit number})$ really easy or really hard?

Chapter 7

Congruences, Part III

Key Ideas

1. How to solve $ax \equiv 1 \pmod{n}$ efficiently.
2. How to compute $a^m \pmod{n}$ efficiently.
3. A probabilistic primality test.

7.1 How to Solve $ax \equiv 1 \pmod{n}$

Let $a, n \in \mathbb{N}$ with $\gcd(a, n) = 1$. Then we know that $ax \equiv 1 \pmod{n}$ has a solution. How can we find x?

7.1.1 More About GCDs

Proposition 7.1.1. *Suppose $a, b \in \mathbb{Z}$ and $\gcd(a, b) = d$. Then there exists $x, y \in \mathbb{Z}$ such that*

$$ax + by = d.$$

I won't give a formal proof of this proposition, though there are many in the literature. Instead I will show you how to find x and y in practice, because that's what you will need to do in order to solve equations like $ax \equiv 1 \pmod{n}$.

Example 7.1.2. Let $a = 5$ and $b = 7$. The steps of the Euclidean gcd algorithm are:

$$\begin{aligned} \underline{7} &= 1 \cdot \underline{5} + \underline{2} & \quad &\text{so } \underline{2} = \underline{7} - \underline{5} \\ \underline{5} &= 2 \cdot \underline{2} + \underline{1} & &\text{so } \underline{1} = \underline{5} - 2 \cdot \underline{2} = 3 \cdot \underline{5} - 2 \cdot \underline{7} \end{aligned}$$

On the right, we have written each partial remainder as a linear combination of a and b. In the last step, we write $\gcd(a, b)$ as a linear combination of a and b, as desired.

That example wasn't too complicated, next we try a much longer example.

Example 7.1.3. Let $a = 130$ and $b = 61$. We have

$$\begin{array}{rlcl}
\underline{130} &= 2 \cdot \underline{61} + \underline{8} & \qquad\qquad & \text{so } \underline{8} = \underline{130} - 2 \cdot \underline{61} \\
\underline{61} &= 7 \cdot \underline{8} + \underline{5} & & \text{so } \underline{5} = -7 \cdot \underline{130} + 15 \cdot \underline{61} \\
\underline{8} &= 1 \cdot \underline{5} + \underline{3} & & \text{so } \underline{3} = 8 \cdot \underline{130} - 17 \cdot \underline{61} \\
\underline{5} &= 1 \cdot \underline{3} + \underline{2} & & \text{so } \underline{2} = -15 \cdot \underline{130} + 32 \cdot \underline{61} \\
\underline{3} &= 1 \cdot \underline{2} + \underline{1} & & \text{so } \underline{1} = 23 \cdot \underline{130} - 49 \cdot \underline{61}
\end{array}$$

Thus $x = 130$ and $y = -49$.

Remark 7.1.4. For our present purposes it will always be sufficient to find one solution to $ax + by = d$. In fact, there are always infinitely many solutions. If x, y is a solution to

$$ax + by = d,$$

then for any $\alpha \in \mathbb{Z}$,

$$a\left(x + \alpha \cdot \frac{b}{d}\right) + b\left(y - \alpha \cdot \frac{a}{d}\right) = d,$$

is also a solution, and all solutions are of the above form for some α.

It is also possible to compute x and y using PARI.

```
? ?bezout
bezout(x,y): gives a 3-dimensional row vector [u,v,d] such that
             d=gcd(x,y) and u*x+v*y=d.
? bezout(130,61)
%1 = [23, -49, 1]
```

7.1.2 To solve $ax \equiv 1 \pmod{n}$

Suppose $\gcd(a, n) = 1$. To solve

$$ax \equiv 1 \pmod{n},$$

find x and y such that $ax + ny = 1$. Then

$$ax \equiv ax + ny \equiv 1 \pmod{n}.$$

Example 7.1.5. Solve $17x \equiv 1 \pmod{61}$. First, we use the Euclidean algorithm to find x, y such that $17x + 61y = 1$:

$$\begin{array}{rlcl}
\underline{61} &= 3 \cdot \underline{17} + \underline{10} & \qquad\qquad & \text{so } \underline{10} = \underline{61} - 3 \cdot \underline{17} \\
\underline{17} &= 1 \cdot \underline{10} + \underline{7} & & \text{so } \underline{7} = -\underline{61} + 4 \cdot \underline{17} \\
\underline{10} &= 1 \cdot \underline{7} + \underline{3} & & \text{so } \underline{3} = 2 \cdot \underline{61} - 7 \cdot \underline{17} \\
\underline{3} &= 2 \cdot \underline{3} + \underline{1} & & \text{so } \underline{1} = -5 \cdot \underline{61} + 18 \cdot \underline{17}
\end{array}$$

Thus $x = 18$ is a solution to $17x \equiv 1 \pmod{61}$.

7.2 How to Compute $a^m \pmod{n}$ Efficiently

As we will see on Friday, a quick method to compute $a^m \pmod{n}$ is absolutely *essential* to public-key cryptography.

Naive Algorithm: Compute $a \cdot a \cdot \cdots \cdot a \pmod{n}$ by repeatedly multiplying by a and reducing modulo m. This is BAD because it takes $m-1$ multiplications.

Clever Algorithm: The following observation is the key idea which makes the clever algorithm work. Write $m = \sum_{i=1}^{r} \varepsilon_i 2^i$ with each $\varepsilon_i \in \{0,1\}$, i.e., write m in base 2 (binary). Then

$$a^m = \prod_{\varepsilon_i = 1} a^{2^i} \pmod{n}.$$

It is straightforward to write a number m in binary, as follows: If m is odd, then $\varepsilon_0 = 1$, otherwise $\varepsilon_0 = 0$. Replace m by floor$(\frac{m}{2})$. If the new m is odd then $\varepsilon_1 = 1$, otherwise $\varepsilon_1 = 0$. Keep repeating until $m = 0$.

Example 7.2.1.
Problem: *Compute the last* 2 *digits of* 6^{91}.

Solution: We compute $6^{91} \pmod{100}$.

i	m	ε_i	6^{2^i} mod 100
0	91	1	6
1	45	1	36
2	22	0	96
3	11	1	16
4	5	1	56
5	2	0	36
6	1	1	96

As a check, note that $91 = 1011011_2 = 2^6 + 2^4 + 2^3 + 2 + 2^0$. Finally, we have

$$6^{91} = 6^{2^6} \cdot 6^{2^4} \cdot 6^{2^3} \cdot 6^2 \cdot 6 \equiv 96 \cdot 56 \cdot 16 \cdot 36 \cdot 6 \equiv 56 \pmod{100}.$$

Summary of above table: The first column, labeled i, is just to keep track of i. The second column, labeled m, is got by dividing the entry above it by 2 and taking the integer part of the result. The third column, labeled ε_i, simply records whether or not the second column is odd. The forth column is computed by squaring, modulo 100, the entry above it.

Some examples in PARI to convince you that powering isn't too difficult:

```
? Mod(17,389)^5000
%13 = Mod(330, 389)
? Mod(2903,49084098)^498494
%14 = Mod(13189243, 49084098)
```

These both take no noticeable time.

7.3 A Probabilistic Primality Test

Recall,

Theorem 7.3.1. *A natural number p is prime if and only if for* every $a \not\equiv 0 \pmod{p}$,

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

Thus if $p \in \mathbb{N}$ and, e.g., $2^{p-1} \not\equiv 1 \pmod{p}$, then we have proved that p is *not* prime. If, however, $a^{p-1} \equiv 1 \pmod{p}$ for a couple of a, then it is "highly likely" that p is prime. I will not analyze this probability here, but we might later in this course.

Example 7.3.2. Let $p = 323$. Is p prime? Let's compute 2^{322} modulo 323. Making a table as above, we have

i	m	ε_i	$2^{2^i} \bmod 323$
0	322	0	2
1	161	1	4
2	80	0	16
3	40	0	256
4	20	0	290
5	10	0	120
6	5	1	188
7	2	0	137
8	1	1	35

Thus

$$2^{322} \equiv 4 \cdot 188 \cdot 35 \equiv 157 \pmod{323},$$

so 323 is not prime. In fact, $323 = 17 \cdot 19$.

It's possible to prove that a large number is composite, but yet be unable to (easily) find a factorization! For example if

$$n = 95468093486093450983409583409850934850938459083,$$

then $2^{n-1} \not\equiv 1 \pmod{n}$, so n is composite. This is something one could verify in a reasonable amount of time by hand. (Though finding a factorization by hand would be very difficult!)

7.3.1 Finding large numbers that are probably prime

```
? probprime(n, a=2) = Mod(a,n)^(n-1) == Mod(1,n)
? x = 0948609348698406983409580934859034509834095809348509834905809345
%36 = 948609348698406983409580934859034509834095809348509834905809345
? for(i=0,100,if(probprime(x+2*i,2),print(i)))
27
? p = x + 2*27
%37 = 948609348698406983409580934859034509834095809348509834905809399
? probprime(p,3)
%39 = 1
```

Chapter 8

Public-key Crypto I: Diffie-Hellman Key Exchange

Key Ideas

- Public-key cryptography
- The Diffie-Hellman key exchange

8.1 Public-key Cryptography

Nikita must communicate vital information to Michael, who is a thousand kilometers away. Their communications are being monitored by The Collective, which must not discover the message. If Nikita and Michael could somehow agree on a secret encoding key, they could encrypt their message. Fortunately, Nikita knows about an algorithm developed by Diffie and Hellman in 1976.

8.2 The Diffie-Hellman Key Exchange Protocol

Nikita and Michael agree on a prime number p and an integer g that has order $p-1$ modulo p. (So $g^{p-1} \equiv 1 \pmod{p}$, but $g^n \not\equiv 1 \pmod{p}$ for any positive $n < p-1$.) Nikita chooses a random number $n < p$, and Michael chooses a random number $m < p$. Nikita sends $g^n \pmod{p}$ to Michael, and Michael sends $g^m \pmod{p}$ to Nikita. Nikita can now compute the secret key:

$$s = g^{mn} = (g^m)^n \pmod{p}.$$

Likewise, Michael computes the secret key:

$$s = g^{mn} = (g^n)^m \pmod{p}.$$

Now Nikita uses the secret key s to send Michael an encrypted version of her critical message. Michael, who also knows s, is able to decode the message.

Meanwhile, hackers in The Collective see both $g^n \pmod p$ and $g^m \pmod p$, but they aren't able to use this information to deduce either m, n, or $g^{mn} \pmod p$ quickly enough to stop Michael from thwarting their plans. Yeah!

The Diffie-Hellman key exchange is the first public-key cryptosystem every published (1976). The system was discovered by GCHQ (British intelligence) a few years before Diffie and Hellman found it, but they couldn't tell anyone about their work; perhaps it was discovered by others before. That this system was discovered independently more than once shouldn't surprise you, given how simple it is!

8.2.1 Some Quotes

A review of Diffie and Hellman's groundbreaking article is amusing, because the reviewer, J.S. Joel, says "They propose a couple of techniques for implementing the system, but the reviewer was unconvinced."

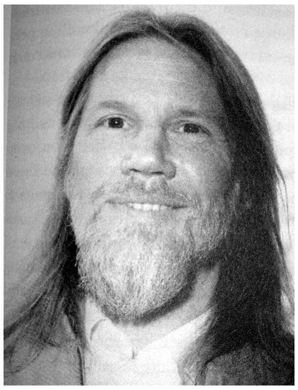
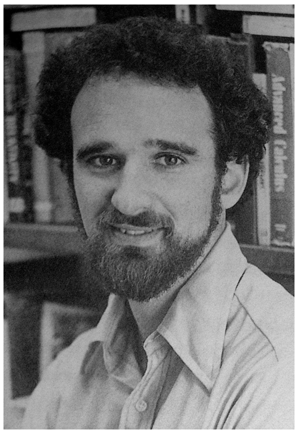
```
Diffie, Whitfield; Hellman, Martin E.
 New directions in cryptography.
 IEEE Trans. Information Theory IT-22 (1976), no. 6, 644--654.
```

> The authors discuss some of the recent results in communications theory that have arisen out of the need for security in the key distribution channels. They concentrate on the use of ciphers to restrict the extraction of information from a communication over an insecure [channel]. As is well known, the transmission and distribution is then likely to become a problem, in efficiency if not in security. The authors suggest various possible approaches to avoid these further problems that arise. The first they call a "public key distribution system", which has the feature that an unauthorized "eavesdropper" will find it computationally infeasible to decipher the message since the enciphering and deciphering are governed by distinct keys. They propose a couple of techniques for implementing the system, but the reviewer was unconvinced.

Somebody named Alan Westrope wrote in 1998 about political implications:

> The 1976 publication of "New Directions in Cryptography", by Whitfield Diffie and Martin Hellman, was epochal in cryptographic history. Many regard it as the beginning of public-key cryptography, analogous to a first shot in what has become an ongoing battle over privacy, civil liberties, and the meaning of sovereignty in cyberspace.

Here is what Diffie and Hellman look like, respectively:

8.3 Let's try it!

To make finding g easier, let's choose a prime p such that $(p-1)/2 = q$ is prime (so $p-1 = 2q$, with q prime). Since for any g with $\gcd(g,p) = 1$,

$$g^{2q} \equiv 1 \pmod{p},$$

the order of g is 1, 2, q, or $2q = p-1$, so the order of g is easy to compute.

For our first example, let $p = 23$. Then $g = 5$ has order $p - 1 = 22$. (I found $g = 5$ using the function `znprimroot` in PARI. You can also just compute the order of 2, 3, etc., until you find a number with order $p-1$.)

Nikita: Chooses secret $n = 12$; sends $g^{12} = 5^{12} \equiv \mathbf{18} \pmod{23}$.

Michael: Chooses secret $n = 5$; sends $g^5 = 5^5 \equiv \mathbf{20} \pmod{23}$.

Compute Shared Secret:
Nikita: $20^{12} \equiv \mathbf{3} \pmod{23}$
Michael: $18^5 \equiv \mathbf{3} \pmod{23}$.

8.4 The Discrete Logarithm Problem

Let a, b, n be positive real numbers. Recall that

$$\log_b(a) = n \text{ if and only if } a = b^n.$$

Thus the $\log_b$ function solves the following problem: Given a base b and a power a of b, find an exponent n such that

$$a = b^n.$$

That is, given b^n and b, find n.

Example 8.4.1. $a = 19683$, $b = 3$. A calculator quickly gives that

$$n = \log(19683)/\log(3) = 9.$$

The discrete log problem is the analogue of this problem modulo p:

Discrete Log Problem: Given $b \pmod{p}$ and $b^n \pmod{p}$, find n. Put another way, compute $\log_b(a)$, when $a, b \in \mathbb{Z}/p\mathbb{Z}$.

As far as we know, this problem is **VERY HARD** to solve quickly. Nobody has admitted publicly to having proved that the discrete log can't be solved quickly, but many very smart people have tried hard and not succeeded. It's easy to write a *slow* program to solve the discrete log problem. (There are better methods but we won't discuss them in this class.)

```
? dislog(x,g, s) = s=g; for(n=1,znorder(g),if(x==s, return(n), s=s*g)); 0;
? dislog(18,Mod(5,23))
%6 = 12
? dislog(20,Mod(5,23))
%7 = 5
```

So the example above was far too simple. Let's try a slightly larger prime:

```
? p=nextprime(9584)
%8 = 9587
? isprime((p-1)\2)
%9 = 1
? znorder(Mod(2,p))
%10 = 9586
? g=Mod(2,p)
%11 = Mod(2, 9587)
? a = g^389
%15 = Mod(7320, 9587)
? dislog(a,g)
%16 = 389
```

This is still very easy to "crack". Let's try an even bigger one.

```
? p = 9048610007
%1 = 9048610007
?  g = Mod(5,p)
%2 = Mod(5, 9048610007)
? a = g^948603
%3 = Mod(3668993056, 9048610007)
? dislog(a,g)              \\ this take a while
%4 = 948603
? znlog(a,g)        \\ builtin super-optimized version takes about 1/2 second
%31 = 948603
```

Computing the discrete log gets slow quickly, the larger we make the p. Doubling the number of digits of the modulus makes the discrete log much much harder.

8.4.1 The State of the Art

```
Discrete logarithms in GF(2^n)
From: Reynald LERCIER <lercier@club-internet.fr>
```

```
 To: NMBRTHRY@LISTSERV.NODAK.EDU
 Date: Tue, 25 Sep 2001 13:37:18 -0400

We are pleased to announce a new record for the discrete logarithm
problem. We were able to compute discrete logarithms in
GF(2^521). This was done in one month on a unique 525MHz
quadri-processors Digital Alpha Server 8400 computer. The approach
that we followed is a careful implementation of the general Function
Field Sieve as described from a theoretical point of view by Adleman
[Ad94].

As far as we know, the largest such computation previously done was
performed in GF(2^401) [GoMc92] using an algorithm due to Coppersmith
[Co84].

[...]
So, as a conclusion, time that we need for computing discrete
logarithms in GF(2^521) on a 525 MHz quadri-processor alpha server
8400 computer is approximatively 12 hours for each, once the sieving
step (21 days) and the linear algebra step (10 days) is performed.

Antoine JOUX    (DCSSI, Issy les Moulineaux, France, Antoine.Joux@ens.fr),
Reynald LERCIER (CELAR, Rennes, France, lercier@celar.fr).
```

8.5 Realistic Example

```
? p=nextprime(93450983094850938450983409583)
%17 = 93450983094850938450983409611
? isprime((p-1)\2)
%18 = 0
? nextgoodprime(p) = while(!isprime((p-1)\2), p=nextprime(p+1)); p
? nextgoodprime(p)
%19 = 93450983094850938450983409623
? g=2
%21 = 2
? znorder(Mod(g,p))
%22 = 93450983094850938450983409610
? ?random
random({N=2^31}): random integer between 0 and N-1.
? nikita = random(p)
%23 = 18319922375531859171613379181
? michael = random(p)
%24 = 82335836243866695680141440300
? nikita_say = Mod(g,p)^nikita
%26 = Mod(17037287637415625385373411504, 93450983094850938450983409611)
? michael_say=Mod(g,p)^michael
%27 = Mod(22014258943243699707729405 47, 93450983094850938450983409611)
```

```
? secret = nikita_say^michael
%28 = Mod(25591938014843312529239952955, 93450983094850938450983409611)
? secret = michael_say^nikita
%29 = Mod(25591938014843312529239952955, 93450983094850938450983409611)
```

Chapter 9

The RSA Public-Key Cryptosystem, I

Key Ideas:

- Creating an RSA public key
- Encrypting and decrypting messages
- Breaking RSA and factoring

Contents

9.1 How RSA works

9.1.1 One-way Functions

The *fundamental idea* behind RSA is to try to construct a "one-way function", i.e., an "encryption" function

$$E : X \to X$$

such that it is easy for Nikita, say, to compute E^{-1}, but very hard for anybody else to compute E^{-1}.

9.1.2 How Nikita Makes an RSA Public Key

Here is how Nikita makes a one-way function E:

1. Nikita picks two large primes p and q, and lets $n = pq$.

2. It is easy for Nikita to then compute

$$\varphi(n) = \varphi(p) \cdot \varphi(q) = (p-1) \cdot (q-1).$$

3. Nikita next chooses a "random" integer e with

$$1 < e < \varphi(n) \text{ and } \gcd(e, \varphi(n)) = 1.$$

4. Finally, Nikita uses the algorithm from Lecture 7 to find a solution d to the equation

$$ex \equiv 1 \pmod{\varphi(n)}.$$

The Encoding Function:
Nikita defines a function $E : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$

$$E(x) = x^e.$$

(Recall that $\mathbb{Z}/n\mathbb{Z} = \{0, 1, \ldots, n-1\}$ with addition and multiplication modulo n.) Then anybody can compute E fairly quickly using the repeated-squaring algorithm from Lecture 7.

Nikita's **public key** is the pair of integers (n, e), which is just enough information for people to easily compute E. Nikita knows a number d such that $ed \equiv 1 \pmod{\varphi(n)}$, so, as we will see below, she can quickly compute E^{-1}.

Now Michael or even The Collective can send Nikita a message whenever they want, even if Nikita is asleep. They look up how to compute E and compute E(their message).

9.1.3 Sending Nikita an Encrypted Message

Encode your message as a sequence of numbers modulo n (see Section 9.2):

$$m_1, \ldots, m_r \in \mathbb{Z}/n\mathbb{Z}.$$

Send

$$E(m_1), \ldots, E(m_r)$$

to Nikita. (Recall that $E(m) = m^e$.)

9.1.4 How Nikita Decrypts a Message

When Nikita receives an $E(m_i)$, she finds m_i as follows:

$$m_i = E^{-1}(E(m_i)) = E(m_i)^d = (m_i^e)^d = m_i.$$

The following proposition proves that the last equality holds.

Proposition 9.1.1. *Let n be a square-free integer and let $d, e \in \mathbb{N}$ such that $p - 1 \mid de - 1$ for each prime $p \mid n$. Then $a^{de} \equiv a \pmod{n}$ for all $a \in \mathbb{Z}$.*

Proof. Since $n \mid a^{de} - a$ if and only if $p \mid a^{de} - a$ for each prime divisor of p, it suffices to prove that $a^{de} \equiv a \pmod{p}$ for each prime divisor p of n. If $\gcd(a, p) \neq 0$, then $a \equiv 0 \pmod{p}$, so $a^{de} \equiv a \pmod{p}$. If $\gcd(a, p) = 1$, then Fermat's Little Theorem asserts that $a^{p-1} \equiv 1 \pmod{p}$. Since $p - 1 \mid de - 1$, we have $a^{de-1} \equiv 1 \pmod{p}$ as well. Multiplying both sides by a shows that $a^{de} \equiv a \pmod{p}$. $\square$

9.2 Encoding a Phrase in a Number

Think of a sequence of letters and spaces as a number in base 27. Let a single-space correspond to 0, the letter A to 1, B to 2, ..., Z to 26. Thus, e.g., "HARVARD" denotes a number written in base 27. The corresponding number written in decimal is 1808939906:

$$\text{HARVARD} \quad \leftrightarrow \quad 8+27\cdot 1+27^2\cdot 18+27^3\cdot 22+27^4\cdot 1+27^5\cdot 18+27^6\cdot 4 = 1808939906$$

To recover the digits of the number, repeatedly divide by 27:

$$\begin{array}{lcrcrl} 1808939906 & = & 66997774 \cdot 27 & + & 8 & \text{H} \\ 66997774 & = & 2481399 \cdot 27 & + & 1 & \text{A} \end{array}$$

and so on.

9.2.1 How Many Letters Can a Number "Hold"?

If $27^k < n$, then k letters can be encoded in a number $< n$. Put another way,

$$k < \log(n)/\log(27) = \log_{27}(n).$$

9.3 Examples

9.3.1 A Small Example

So the arithmetic is easy to follow, we use small primes p and q and encrypt the single letter "X".

1. Choose p and q: Let $p = 17$, $q = 19$, so $n = pq = 323$.
2. Compute $\varphi(n)$:

   $$\varphi(n) = \varphi(p \cdot q) = \varphi(p) \cdot \varphi(q) = (p-1)(q-1) = pq-p-q+1 = 323-17-19+1 = 288.$$

3. Randomly choose an $e \in \mathbb{Z}/323\mathbb{Z}$: We choose $e = 95$.
4. Solve

   $$95x \equiv 1 \pmod{288}.$$

   Using the GCD algorithm, we find that $d = 191$ solves the equation.

The public key is $(323, 95)$. So $E : \mathbb{Z}/323\mathbb{Z} \to \mathbb{Z}/323\mathbb{Z}$ is defined by

$$E(x) = x^{95}.$$

Next, we encrypt the letter "X". It is encoded as the number 24, since X is the 24th letter of the alphabet. We have

$$E(24) = 24^{95} = 294 \in \mathbb{Z}/323\mathbb{Z}.$$

To decrypt, we compute E^{-1}:

$$E^{-1}(294) = 294^{191} = 24 \in \mathbb{Z}/323\mathbb{Z}.$$

9.3.2 A Bigger Example in PARI

```
? p=nextprime(random(10^30))
%3 = 738873402423833494183027176953
? q=nextprime(random(10^25))
%4 = 3787776806865662882378273
? n=p*q
%5 = 2798687536910915970127263606347911460948554197853542169
? e=random(n)
%6 = 1483959194866204179348536010284716655442139024915720699
? phin=(p-1)*(q-1)
%7 = 2798687536910915970127262867470721260308194351943986944
? while(gcd(e,phin)!=1,e=e+1)
? e
%8 = 1483959194866204179348536010284716655442139024915720699
? d = lift(Mod(e,phin)^(-1));
%9 = 2113367928496305469541348387088632973457802358781610803
? (e*d)%phin
%10 = 1
? log(n)/log(27)
%11 = 38.03851667699197952338510248
```

We can encode single blocks of up to 38 letters. Let's encode "HARVARD":

```
? m=8+27*1+27^2*18+27^3*22+27^4*1+27^5*18+27^6*4
%12 = 1808939906
? E(x)=lift(Mod(x,n)^e)
? D(x)=lift(Mod(x,n)^d)
? secret_message = E(m)
%14 = 625425724974078486559370130768554070421628674916144724
? D(secret_message)
%15 = 1808939906
```

The following complete PARI program automates the whole process, though it is a little clumsy. Call this file `rsa.gp`. It uses { and } so that functions can be extended over more than one line.

```
/* rsa.gp ------------------------------------------------- */
{alphabet=[" ","A","B","C","D","E","F","G","H","I","J","K","L","M",
           "N","O","P","Q","R","S","T","U","V","W","X","Y","Z"];
}
{letter_to_number(l,
     n)=
     for(n=1,27,if(alphabet[n]==l,return(n-1)));
     error("invalid input.")
}
{number_to_message(n,
     s="")=
     while(n>0, s = concat(s,alphabet[n%27+1]); n = n \ 27);
     return(s)
}
{message_to_number(w,
     i,n=0)=
     for(i=1,length(w), n = n + 27^(i-1)*letter_to_number(w[i]));
     return(n);
}
{make_rsa_key(len,
     p,q,n,e,d)=
     p = nextprime(random(10^(len\2+1)));
     q = nextprime(random(10^(len\2+3)));
     n = p*q; phin = (p-1)*(q-1);
     e = random(phin);
     while(gcd(e,phin)!=1,e=e+1);
     d = lift(Mod(e,phin)^(-1));
     return([n,e,d]);
}
encrypt(message, n, e) = lift(Mod(message_to_number(message),n)^e);
decrypt(secret, n, d)  = number_to_message(lift(Mod(secret,n)^d));
/* rsa.gp ------------------------------------------------- */
```

Here is an example that uses the above little program.

```
? \r rsa
? setrand(1)   \\ default random number seed is 1!
? rsa=make_rsa_key(20)   \\ returns [n, e, d]
%2 = [89050154117716728145939, 33735260657253161660951,
      49244741969289756040079]
? n = rsa[1]; e = rsa[2]; d = rsa[3];
? public_key = [n,e]
%3 = [89050154117716728145939, 33735260657253161660951]
? msg = ["H", "A", "R", "V", "A", "R", "D"];   \\ clumsy!!!
? secret = encrypt(msg,n,e)
%36 = 75524965161901413275866
? decrypt(secret, n, d)
%37 = "HARVARD"
```

9.4 A Connection Between Breaking RSA and Factoring Integers

Nikita's public key is (n, e). If we compute the factorization of $n = pq$, then we can compute $\varphi(n)$ and hence deduce her secret decoder number d.
It is no easier to $\varphi(n)$ than to factor n:
Suppose $n = pq$. Given $\varphi(n)$, it is very easy to compute p and q. We have

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

so we know $pq = n$ and $p + q = n + 1 - \varphi(n)$. Thus we know the polynomial

$$x^2 - (p+q)x + pq = (x-p)(x-q)$$

whose roots are p and q.

There is also a more complicated "probabilistic algorithm" to find p and q given the secret decoding number d. I might describe it in the next lecture.

Chapter 10

Attacking RSA

Nikita's public key is (n, e). If we compute the factorization of $n = pq$, then we can compute $\varphi(n)$ and hence deduce her secret decoding number d. Thus attempting to factor n is a way to try to break an RSA public-key cryptosystem. In this lecture we consider several approaches to "cracking" RSA, and relate them to the difficulty of factoring n.

10.1 Factoring n Given $\varphi(n)$

If you know $\varphi(n)$ then it is easy to factor n:
Suppose $n = pq$. Given $\varphi(n)$, it is very easy to compute p and q. We have

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

so we know both $pq = n$ and $p + q = n + 1 - \varphi(n)$. Thus we know the polynomial

$$x^2 - (p+q)x + pq = (x-p)(x-q)$$

whose roots are p and q. These roots can be found using the quadratic formula.

Example 10.1.1.

```
? n=nextprime(random(10^10))*nextprime(random(10^10));
? phin=eulerphi(n);
? f = x^2 - (n+1-phin)*x + n
%6 = x^2 - 12422732288*x + 31615577110997599711
? polroots(f)
%7 = [3572144239, 8850588049]
? n
%8 = 31615577110997599711
? 3572144239*8850588049
%9 = 31615577110997599711
```

10.2 When p and q Are Close

Suppose that p and q are "close" to each other. Then it is easy to factor n using a factorization method of Fermat.

Suppose $n = pq$ with $p > q$, say. Then

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

Since p and q are "close",

$$s = \frac{p-q}{2}$$

is small,

$$t = \frac{p+q}{2}$$

is only slightly larger than $\sqrt{n}$, and $t^2 - n = s^2$ is a perfect square. So we just try

$$t = \mathrm{ceil}(\sqrt{n}), \quad t = \mathrm{ceil}(\sqrt{n}) + 1, \quad t = \mathrm{ceil}(\sqrt{n}) + 2, \ldots$$

until $t^2 - n$ is a perfect square s^2. Then

$$p = t + s, \qquad q = t - s.$$

Example 10.2.1. Suppose $n = 23360947609$. Then

$$\sqrt{n} = 152842.88\ldots.$$

If $t = 152843$, then $\sqrt{t^2 - n} = 187.18\ldots$.
If $t = 152844$, then $\sqrt{t^2 - n} = 583.71\ldots$.
If $t = 152845$, then $\sqrt{t^2 - n} = 804 \in \mathbb{Z}$.
Thus $s = 804$. We find that $p = t + s = 153649$ and $q = t - s = 152041$.

Here is a bigger example in PARI:

```
? q=nextprime(random(10^50))
%20 = 78177096444230804504075122792410749354743712880803
? p=nextprime(q+1)  \\ a nearby prime
%21 = 78177096444230804504075122792410749354743712880899
? n=p*q
%22 = 61116584084505646970856342018459768505099085809499868895257O4...
      ...2596503421573992791632896516937224818 97
? t=floor(sqrt(n))+1
  ***   precision loss in truncation
? \p150        \\ set precision of floating-point computations.
   realprecision = 154 significant digits (150 digits displayed)
? t=floor(sqrt(n))+1
%29 = 78177096444230804504075122792410749354743712880851
? sqrt(t^2-n)
%30 = 48.000000000000000000000000000000000000000000000000000....
? s=48
%31 = 48
? t + s     \\ p
%33 = 78177096444230804504075122792410749354743712880899
? t - s     \\ q
%35 = 78177096444230804504075122792410749354743712880803
```

10.3 Factoring n Given d

Suppose that we crack an RSA cryptosystem by finding a d such that

$$a^{ed} \equiv a \pmod{n}$$

for all a. Then we've found an m $(= ed - 1)$ such that $a^m \equiv 1 \pmod{n}$ for all a with $\gcd(a, n) = 1$. Knowing a does not lead to a factorization of n in as direct a manner as knowing $\varphi(n)$ does (see Section 10.1). However, there is a probabilistic procedure that, given an m such that $a^m \equiv 1 \pmod{n}$, will with high probability find a factorization of n.

Probabilistic procedure to factor n:

1. m is even since $(-1)^m \equiv 1 \pmod{n}$.
2. If $a^{m/2} \equiv 1 \pmod{n}$ for all a coprime to n, replace m by $m/2$. In practice, it is not possible to determine whether or not this condition holds, because it would require doing a computation for too many a. Instead, we try a few random a; if $a^{m/2} \equiv 1 \pmod{n}$ for the a we check, then we divide m by 2. (If there exists even a single a such that $a^{m/2} \not\equiv 1 \pmod{n}$, then at least half the a have this property.)

   Keep repeating this step until we find an a such that $a^{m/2} \not\equiv 1 \pmod{n}$.
3. There is a 50% chance that a randomly chosen a will have the property that

   $$a^{m/2} \equiv +1 \pmod{p}, \qquad a^{m/2} \equiv -1 \pmod{q}$$

   or

   $$a^{m/2} \equiv -1 \pmod{p}, \qquad a^{m/2} \equiv +1 \pmod{q}.$$

   If the first case occurs, then

   $$p \mid a^{m/2} - 1, \qquad \text{but } q \nmid a^{m/2} - 1,$$

   so

   $$\gcd(a^{m/2} - 1, pq) = p,$$

   and we have factored n. Just keep trying a's until one of the cases occurs.

```
? \r rsa   \\ load the file rsa.gp, available at Lecture 9 web page.
? rsa = make_rsa_key(10)
%34 = [32295194023343, 29468811804857, 11127763319273]
? n = rsa[1]; e = rsa[2]; d = rsa[3];
? m = e*d-1
%38 = 327921963064646896263108960
? for(a=2,20, if(Mod(a,n)^m!=1,print(a)))   \\ prints nothing...
? m = m/2
%39 = 163960981532323448131554480
? for(a=2,20, if(Mod(a,n)^m!=1,print(a)))
? m = m/2
%40 = 81980490766161724065777240
```

```
? for(a=2,20, if(Mod(a,n)^m!=1,print(a)))
? m = m/2
%41 = 40990245383080862032888620
? for(a=2,20, if(Mod(a,n)^m!=1,print(a)))
? m = m/2
%42 = 20495122691540431016444310
? for(a=2,20,if(Mod(a,n)^m!=1,print(a)))
2
5
6
... etc.
? gcd(2^m,n)
  ***   power overflow in pow_monome.
? x = lift(Mod(2,n)^m)-1
%43 = 4015382800098
? gcd(x,n)
%46 = 737531
? p = gcd(x,n)
%53 = 737531
? q = n/p
? p*q
%54 = 32295194023343
? n
%55 = 32295194023343
```

10.4 RSA Challenge n

The easiest challenge at

`http://www.rsasecurity.com/rsalabs/challenges/factoring/numbers.html`

is the 576-bit number

```
Name:         RSA-576
Prize:        $10000
Digits:       174
Digit Sum:    785
188198812920607963838697239461650439807163563379417382700763356422988859
715234665485319060606504743045317388011303396716199692321205734031879550
656996221305168759307650257059
```

Chapter 11

Primitive Roots

Key Idea: *There is an element of $(\mathbb{Z}/p\mathbb{Z})$ of order $p-1$.*

11.1 Polynomials over $\mathbb{Z}/p\mathbb{Z}$

Proposition 11.1.1. *Let $f \in (\mathbb{Z}/p\mathbb{Z})[x]$ be a nonzero polynomial over the ring $\mathbb{Z}/p\mathbb{Z}$. Then there are at most* $\deg(f)$ *elements $\alpha \in \mathbb{Z}/p\mathbb{Z}$ such that $f(\alpha) = 0$.*

Proof. We proceed by induction on $\deg(f)$. The cases $\deg(f) = 0, 1$ are clear. Write $f = a_n x^n + \cdots a_1 x + a_0$. If $f(\alpha) = 0$ then

$$\begin{aligned} f(x) &= f(x) - f(\alpha) \\ &= a_n(x^n - \alpha^n) + \cdots a_1(x - \alpha) + a_0(1 - 1) \\ &= (x - \alpha)(a_n(x^{n-1} + \cdots + \alpha^{n-1}) + \cdots a_1) \\ &= (x - \alpha)g(x), \end{aligned}$$

for some polynomial $g(x) \in (\mathbb{Z}/p\mathbb{Z})[x]$. Next suppose that $f(\beta) = 0$ with $\beta \neq \alpha$. Then $(\beta - \alpha)g(\beta) = 0$, so, since $\beta - \alpha \neq 0$ (hence $\gcd(\beta - \alpha, p) = 1$, we have $g(\beta) = 0$. By our inductive hypothesis, g has at most $n - 1$ roots, so there are at most $n - 1$ possibilities for β. It follows that f has at most n roots. $\square$

Proposition 11.1.2. *Let p be a prime number and let d be a divisor of $p-1$. Then $f(x) = x^d - 1 \in (\mathbb{Z}/p\mathbb{Z})[x]$ has exactly d solutions.*

Proof. Let e be such that $de = p - 1$. We have

$$\begin{aligned} x^{p-1} - 1 &= (x^d)^e - 1 \\ &= (x^d - 1)((x^d)^{e-1} + (x^d)^{e-2} + \cdots + 1) \\ &= (x^d - 1)g(x), \end{aligned}$$

where $\deg(g(x)) = p-1-d$. Recall that Fermat's little theorem implies that $x^{p-1}-1$ has exactly $p - 1$ roots in $\mathbb{Z}/p\mathbb{Z}$. By Proposition 11.1.1, $f(x)$ has *at most* $p - 1 - d$ roots and $x^d - 1$ has at most d roots, so $g(x)$ has exactly $p - 1$ roots and $x^d - 1$ has exactly d roots, as claimed. $\square$

WARNING: The analogue of this theorem is false for some $f \in (\mathbb{Z}/n\mathbb{Z})[x]$ with n composite. For example, if $n = n_1 \cdot n_2$ with $n_1, n_2 \neq 1$, then $f = nx$ has at least *two* distinct zeros, namely 0 and $n_2 \neq 0$.

11.2 The Structure of $(\mathbb{Z}/p\mathbb{Z})^* = \{1, 2, \ldots, p-1\}$

In this section, we prove that the group $(\mathbb{Z}/p\mathbb{Z})^*$ is cylic.

Definition 11.2.1. A *primitive root* modulo p is an element of $(\mathbb{Z}/p\mathbb{Z})^*$ of order $p-1$.

Question: For which primes p is there a primitive root? (Ans. Every prime.)

Lemma 11.2.2. *Suppose $a, b \in (\mathbb{Z}/n\mathbb{Z})^*$ have orders r and s, respectively, and that* $\gcd(r, s) = 1$. *Then ab has order rs.*

This is a general fact about commuting elements of a group.

Proof. Since $(ab)^{rs} = a^{rs}b^{rs} = 1$, the order of ab is a divisor $r_1 s_1$ of rs, where $r_1 \mid r$ and $s_1 \mid s$. Thus

$$a^{r_1 s_1} b^{r_1 s_1} = (ab)^{r_1 s_1} = 1.$$

Raise both sides to the power r_2, where $r_1 r_2 = r$. Then

$$a^{r_1 r_2 s_1} b^{r_1 r_2 s_1} = 1,$$

so, since $a^{r_1 r_2 s_1} = (a^{r_1 r_2})^{s_1} = 1$,

$$b^{r_1 r_2 s_1} = 1.$$

This implies that $s \mid r_1 r_2 s_1$, and, since $\gcd(s, r_1 r_2) = 1$, it follows that $s = s_1$. A similar argument shows that $r = r_1$, so the order of ab is rs. $\square$

Theorem 11.2.3. *For every prime p there is a primitive root mod p. In other words, the group $(\mathbb{Z}/p\mathbb{Z})^*$ is a cyclic group of order $p-1$.*

Proof. Write

$$p - 1 = q_1^{n_1} q_2^{n_2} \cdots q_r^{n_r}$$

as a product of distinct primes q_i.

By Proposition 11.1.2, the polynomial $x^{q_i^{n_i}} - 1$ has exactly $q_i^{n_i}$ roots, and the polynomial $x^{q_i^{n_i - 1}} - 1$ has exactly $q_i^{n_i-1}$ roots. Thus there is an $a_i \in \mathbb{Z}/p\mathbb{Z}$ such that $a_i^{q_i^{n_i}} = 1$ but $a_i^{q_i^{n_i-1}} \neq 1$. This a_i has order $q_i^{n_i}$. For each $i = 1, \ldots, r$, choose such an a_i. By repeated application of Lemma 11.2.2, we see that

$$a = a_1 a_2 \cdots a_r$$

has order $q_1^{n_1} \cdots q_r^{n_r} = p - 1$, so a is a primitive root. $\square$

Remark 11.2.4. There are $\varphi(p-1)$ primitive roots modulo p, since there are $q_i^{n_i} - q_i^{n_i-1}$ ways to choose a_i. To see this, we check that two distinct choices of sequence $a_1, \ldots, a_r$ define two different primitive roots. Suppose that

$$a_1 a_2 \cdots a_r = a_1' a_2' \cdots a_r',$$

with a_i, a_i' of order $q_i^{n_i}$, for $i = 1, \ldots, r$. Upon raising both sides of this equality to the power $s = q_2^{n_2} \cdots q_r^{n_r}$, we see that $a_1^s = a_1'^s$. Since $\gcd(s, q_1^{n_1}) = 1$, there exists t such that $st \equiv 1 \pmod{q_1^{n_1}}$. It follows that

$$a_1 = (a_1^s)^t = (a_1'^s)^t = a_1'.$$

Upon canceling a_1 from both sides, we see that $a_2 \cdots a_r = a_2' \cdots a_r'$; by repeating the above argument, we see that $a_i = a_i'$ for all i. Thus, different choices of the a_i must lead to different primitive roots; in other words, if the primitive roots are the same, then the a_i were the same.

For example, there are $\varphi(16) = 2^4 - 2^4 = 8$ primitive roots mod 17:

```
? for(n=1,16,if(znorder(Mod(n,17))==16,print1(n," ")))
3 5 6 7 10 11 12 14
```

Example 11.2.5. In this example, we illustrate the proof of Theorem 11.2.3 when $p = 13$. We have

$$p - 1 = 12 = 2^2 \cdot 3.$$

The polynomial $x^4 - 1$ has roots $\{1, 5, 8, 12\}$ and $x^2 - 1$ has roots $\{1, 12\}$, so we take $a_1 = 5$. The polynomial $x^3 - 1$ has roots $\{1, 3, 9\}$, so set $a_2 = 3$. Finally, $a = 5 \cdot 3 = 15 = 2$. Note that the successive powers of 2 are

$$2,\ 4,\ 8,\ 3,\ 6,\ 12,\ 11,\ 9,\ 5,\ 10,\ 7,\ 1,$$

so 2 really does have order 12.

Example 11.2.6. The result is false if, e.g., p is replaced by a big power of 2. The elements of $(\mathbb{Z}/8\mathbb{Z})^*$ all have order dividing 2, but $\varphi(8) = 4$.

Theorem 11.2.7. *Let p^n be a power of an odd prime. Then there is an element of $(\mathbb{Z}/p^n\mathbb{Z})^*$ of order $\varphi(p^n)$. Thus $(\mathbb{Z}/p^n\mathbb{Z})^*$ is cyclic.*

I will not prove Theorem 11.2.7 in class. I will probably put a problem on your next homework set that will guide you to a proof.

11.3 Artin's Conjecture

Conjecture 11.3.1 (Emil Artin). *If $a \in \mathbb{Z}$ is not -1 or a perfect square, then the number $N(x, a)$ of primes $p \leq x$ such that a is a primitive root modulo p is asymptotic to $C(a)\pi(x)$, where $C(a)$ is a constant that depends only on a. In particular, there are infinitely many primes p such that a is a primitive root modulo p.*

Nobody has proved this conjecture for even a single choice of a. There are partial results, e.g., that there are infinitely many p such that the order of a is divisible by the largest prime factor of $p - 1$. (See, e.g., Moree, Pieter, *A note on Artin's conjecture.*)

Chapter 12

Quadratic Reciprocity I

Key Ideas:

- *Euler's Criterion:* When is a a square modulo p?
- Quadratic reciprocity
- Lemma of Gauss

12.1 Euler's Criterion

Proposition 12.1.1 (Euler's Criterion). *Let p be an odd prime and a an integer not divisible by p. Then $x^2 \equiv a \pmod{p}$ has a solution if and only if*

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

Proof. By the theorem from Lecture 11, there is an integer g that has order $p-1$ modulo p. Every integer coprime to p is congruent to a power of g. First suppose that a is congruent to a perfect square modulo p, so

$$a \equiv (g^r)^2 \equiv g^{2r} \pmod{p}$$

for some r. Then

$$a^{(p-1)/2} \equiv g^{2r\cdot\frac{p-1}{2}} \equiv g^{r(p-1)} \equiv 1 \pmod{p}.$$

Conversely, suppose that $a^{(p-1)/2} \equiv 1 \pmod{p}$. We have $a \equiv g^r \pmod{p}$ for some integer r. Thus $g^{r(p-1)/2} \equiv 1 \pmod{p}$, so

$$p-1 \mid r(p-1)/2$$

which implies that r is even. Thus $a \equiv (g^{r/2})^2 \pmod{p}$, so a is congruent to a square modulo p. □

Corollary 12.1.2. *If $x^2 \equiv a \pmod{p}$ has no solutions if and only if $a^{(p-1)/2} \equiv -1 \pmod{p}$.*

Proof. This follows from Proposition 12.1.1 and that the polynomial x^2-1 has no roots besides $+1$ and -1. □

Example 12.1.3. Suppose $p = 11$. By squaring each element of $(\mathbb{Z}/11\mathbb{Z})^*$, we see exactly which numbers are squares modulo 11:

$$1^2 = 1,\ 2^2 = 4,\ 3^2 = 9,\ 4^2 = 5,\ 5^2 = 3,\ 6^2 = 3,\ 7^2 = 5,\ 8^2 = 9,\ 9^2 = 4,\ 10^2 = 1.$$

Thus the squares are $\{1, 3, 4, 5, 9\}$. Next, we compute $a^{(p-1)/2} = a^5$ for each $a \in (\mathbb{Z}/11\mathbb{Z})^*$.

$$1^5 = 1,\ 2^5 = -1,\ 3^5 = 1,\ 4^5 = 1,\ 5^5 = 1,\ 6^5 = -1,\ 7^5 = -1,\ 8^5 = -1,\ 9^5 = 1,\ 10^5 = -1.$$

The a with $a^5 = 1$ are $\{1, 3, 4, 5, 9\}$, which is exactly the same as the set of squares, just as Proposition 12.1.1 predicts.

Example 12.1.4. Determine whether or not 3 is a square modulo $p = 726377359$.
Answer: We compute $3^{(p-1)/2}$ modulo p using PARI:

```
? Mod(3,p)^((p-1)/2)
%5 = Mod(726377358, 726377359)   \\ class of -1 modulo 726377359.
```

Thus 3 is not a square modulo p. This computation wasn't too difficult, but it would have been very tedious to carry about by hand. The law of quadratic reciprocity, which we will state in the next section, is a vastly more powerful way to answer such questions. For example, you could easily answer the above question by hand using quadratic reciprocity.

Remark 12.1.5. Proposition 12.1.1 can be reformulated in more group-theoretic language as follows. The map

$$(\mathbb{Z}/p\mathbb{Z})^* \to \{\pm 1\}$$

that sends a to $a^{(p-1)/2} \pmod{p}$ is a homomorphism of groups, whose kernel is the subgroup of squares of elements of $(\mathbb{Z}/p\mathbb{Z})^*$.

Definition 12.1.6. An element $a \in \mathbb{Z}$ with $p \nmid a$ is called a *quadratic residue* modulo p if a is a square modulo p.

12.2 The Quadratic Reciprocity Law

Let p be an odd prime and let a be an integer with $p \nmid a$. Set

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

Proposition 12.1.1 implies that

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

Also, notice that

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

because $\left(\frac{\cdot}{p}\right)$ is a homomorphism (see Remark 12.1.5).

The symbol $\left(\frac{a}{p}\right)$ only depends on the residue class of a modulo p. Thus tabulating the value of $\left(\frac{a}{5}\right)$ for hundreds of a would be silly. *Would it be equally silly to make a table of* $\left(\frac{5}{p}\right)$ *for hundreds of primes* p*?* Let's begin making such a table and see whether or not there is an obvious pattern. (To compute $\left(\frac{a}{p}\right)$ in PARI, use the command `kronecker(a,b)`.)

p	$\left(\frac{5}{p}\right)$	$p \bmod 5$
7	-1	2
11	1	1
13	-1	3
17	-1	2
19	1	4
23	-1	3
29	1	4
31	1	1
37	-1	2
41	1	1
43	-1	3
47	-1	2

The evidence suggests that $\left(\frac{5}{p}\right)$ depends only on the congruence class of p; more precisely, $\left(\frac{5}{p}\right) = 1$ if and only if $p \equiv 1, 4 \pmod 5$, i.e., p is a square modulo 5. However, when I think directly about the equation

$$5^{(p-1)/2} \pmod p,$$

I see no way that knowing that $p \equiv 1, 4 \pmod 5$ helps us to evaluate that strange expression! And yet, the numerical evidence is so *compelling*! Argh!

Based on such computations, various mathematicians found a conjectural explanation for this mystery in the 18th century. Finally, on April 8, 1796, at your age (age 19), Gauss proved their conjecture.

Theorem 12.2.1 (The Law of Quadratic Reciprocity). *Suppose that p and q are odd primes. Then*

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

We will prove this theorem in the next lecture.

In the case considered above, this theorem implies that

$$\left(\frac{5}{p}\right) = (-1)^{2\cdot\frac{p-1}{2}}\left(\frac{p}{5}\right) = \left(\frac{p}{5}\right) = \begin{cases} +1 & \text{if } p \equiv 1,4 \pmod 5 \\ -1 & \text{if } p \equiv 2,3 \pmod 5. \end{cases}$$

Thus the quadratic reciprocity law "explains" why knowing p modulo 5 helps in computing $5^{\frac{p-1}{2}} \pmod p$.

Here is a list of almost 200 proofs of Theorem 13.3.1:

`http://www.rzuser.uni-heidelberg.de/~hb3/rchrono.html`

12.3 A Lemma of Gauss

The proof we will give of Theorem 13.3.1 was first discovered by Gauss, though not when he was 19. This proof is given in many elementary number theory texts (including Davenport). It depends on the following lemma of Gauss:

Lemma 12.3.1. *Let p be an odd prime and let a be an integer $\not\equiv 0 \pmod{p}$. Form the numbers*

$$a,\ 2a,\ 3a,\ \ldots,\ \frac{p-1}{2}a$$

and reduce them modulo p to lie in the interval $(-\frac{p}{2},\ \frac{p}{2})$. Let ν be the number of negative numbers in the resulting set. Then

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

Proof. In defining ν, we expressed each number in

$$S = \left\{a, 2a, \ldots, \frac{p-1}{2}a\right\}$$

as congruent to a number in the set

$$\left\{1, -1, 2, -2, \ldots, \frac{p-1}{2}, -\frac{p-1}{2}\right\}.$$

No number $1, 2, \ldots \frac{p-1}{2}$ appears more than once, with either choice of sign, because if it did then either two elements of S are congruent modulo p or 0 is the sum of two elements of S, and both events are impossible. Thus the resulting set must be of the form

$$T = \left\{\varepsilon_1 \cdot 1, \varepsilon_2 \cdot 2, \ldots, \varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right\},$$

where each ε_i is either $+1$ or -1. Multiplying together the elements of S and of T, we see that

$$(1a) \cdot (2a) \cdot (3a) \cdot \cdots \cdot \left(\frac{p-1}{2}a\right) \equiv (\varepsilon_1 \cdot 1) \cdot (\varepsilon_2 \cdot 2) \cdots \left(\varepsilon_{(p-1)/2} \cdot \frac{p-1}{2}\right) \pmod{p},$$

so

$$a^{(p-1)/2} \equiv \varepsilon_1 \cdot \varepsilon_2 \cdot \cdots \cdot \varepsilon_{(p-1)/2} \pmod{p}.$$

The lemma then follows from Proposition 12.1.1. $\square$

Chapter 13

Quadratic Reciprocity II

In-class midterm this Wednesday, October 17!

Monday's lecture will be a review lecture; Grigor's review session is on Monday at 4pm; I will have an extra office hour in SC 515, Tuesday, 2:35–3:30.

13.1 Recall Gauss's Lemma

We proved the following lemma in the previous lecture.

Lemma 13.1.1. *Let p be an odd prime and a an integer with $p \nmid a$. Form the numbers a, $2a$, $3a$, $\ldots$, $\frac{p-1}{2}a$ and reduce them modulo p to lie in the interval $(-\frac{p}{2},\ \frac{p}{2})$. Let ν be the number of negative numbers in the resulting set. Then $\left(\frac{a}{p}\right) = (-1)^\nu$.*

13.2 Euler's Conjecture

Lemma 13.2.1. *Let $a, b \in \mathbb{Q}$. Then for any $n \in \mathbb{Z}$,*

$$\#\left((a,b)\cap\mathbb{Z}\right) \equiv \#\left((a,b+2n)\cap\mathbb{Z}\right) \equiv \#\left((a+2n,b)\cap\mathbb{Z}\right) \pmod{2}.$$

Proof. If $n > 0$, then

$$(a, b+2n) = (a,b) \cup [b, b+2n),$$

where the union is disjoint. Let $[x]$ denote the least integer $\geq x$. There are $2n$ integers,

$$[b], [b]+1, \ldots, [b]+2n-1,$$

in the interval $[b, b+2n)$, so the assertion of the lemma is true in this case. We also have

$$(a, b-2n) = (a,b)\backslash[b-2n, b)$$

and $[b-2n, b)$ also contains exactly $2n$ integers, so the lemma is also true when n is negative. The statement about $\#\left((a+2n,b)\cap\mathbb{Z}\right)$ is proved in a similar manner. $\square$

The following proposition was first conjectured by Euler, based on extensive numerical evidence. Once we've proved this proposition, it will be easy to deduce the quadratic reciprocity law.

Proposition 13.2.2 (Euler's Conjecture). *Let p be an odd prime and $a \in \mathbb{N}$ a natural number with $p \nmid a$.*

1. *The symbol $\left(\frac{a}{p}\right)$ depends only on p modulo $4a$.*

2. *If q is a prime with $q \equiv -p \pmod{4a}$, then $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$.*

Proof. To apply Gauss's lemma, we have to compute the parity of the intersection of

$$S = \left\{a, 2a, 3a, \ldots \frac{p-1}{2}a\right\}$$

and

$$I = \left(\frac{1}{2}p, p\right) \cup \left(\frac{3}{2}p, 2p\right) \cup \cdots \cup \left(\left(b - \frac{1}{2}\right)p, bp\right),$$

where $b = \frac{1}{2}a$ or $\frac{1}{2}(a-1)$, whichever is an integer. (Why? We have to check that every element of S that reduces to something in the interval $(-\frac{p}{2}, 0)$ lies in I. This is clear if $b = \frac{1}{2}a < \frac{p-1}{2}a$. If $b = \frac{1}{2}(a-1)$, then $bp + \frac{p}{2} > \frac{p-1}{2}a$, so $((b-\frac{1}{2})p, bp)$ is the last interval that could contain an element of of S that reduces to $(-\frac{p}{2}, 0)$.) Also note that the integer endpoints of I are not in S, since those endpoints are divisible by p, but no element of S is divisible by p.

Dividing I through by a, we see that

$$\#(S \cap I) = \#\left(\mathbb{Z} \cap \frac{1}{a}I\right),$$

where

$$\frac{1}{a}I = \left(\left(\frac{p}{2a}, \frac{p}{a}\right) \cup \left(\frac{3p}{2a}, \frac{2p}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)p}{2a}, \frac{bp}{a}\right)\right).$$

Write $p = 4ac + r$, and let

$$J = \left(\left(\frac{r}{2a}, \frac{r}{a}\right) \cup \left(\frac{3r}{2a}, \frac{2r}{a}\right) \cup \cdots \cup \left(\frac{(2b-1)r}{2a}, \frac{br}{a}\right)\right).$$

The only difference between I and J is that the endpoints of intervals are changed by addition of an even integer. By Lemma 13.2.1,

$$\nu = \#\left(\mathbb{Z} \cap \frac{1}{a}I\right) \equiv \#(\mathbb{Z} \cap J) \pmod{2}.$$

Thus $\left(\frac{a}{p}\right) = (-1)^\nu$ depends only on r, i.e., only on p modulo $4a$. WOW!

If $q \equiv -p \pmod{4a}$, then the only change in the above computation is that r is replaced by $4a - r$. This changes $\frac{1}{a}I$ into

$$K = \left(\left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right) \cup \left(6 - \frac{3r}{2a}, 8 - \frac{2r}{a}\right) \cup \cdots \cup \left(4b - 2 - \frac{(2b-1)r}{2a}, 4b - \frac{br}{a}\right)\right).$$

Thus K is the same as $-\frac{1}{a}I$, except even integers have been added to the endpoints. By Lemma 13.2.1,

$$\#(K \cap \mathbb{Z}) \equiv \#\left(\left(\frac{1}{a}I\right) \cap \mathbb{Z}\right) \pmod{2},$$

so $\left(\frac{a}{p}\right) = \left(\frac{a}{q}\right)$, which completes the proof. ∎

The following more careful analysis in the special case when $a = 2$ helps illustrate the proof of the above lemma, and is frequently useful in computations.

Proposition 13.2.3. *Let p be an odd prime. Then*

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

Proof. When $a = 2$, the set $S = \{a, 2a, \ldots, 2 \cdot \frac{p-1}{2}\}$ is

$$\{2, 4, 6, \ldots, p - 1\}.$$

We must count the parity of the number of elements of S that lie in the interval $I = (\frac{p}{2}, p)$. Writing $p = 8c + r$, we have

$$\begin{aligned}\#\,(I \cap S) &= \#\left(\frac{1}{2}I \cap \mathbb{Z}\right) = \#\left(\left(\frac{p}{4}, \frac{p}{2}\right) \cap \mathbb{Z}\right) \\ &= \#\left(\left(2c + \frac{r}{4}, 4c + \frac{r}{2}\right) \cap \mathbb{Z}\right) \equiv \#\left(\left(\frac{r}{4}, \frac{r}{2}\right) \cap \mathbb{Z}\right) \pmod 2,\end{aligned}$$

where the last equality comes from Lemma 13.2.1. The possibilities for r are $1, 3, 5, 7$. When $r = 1$, the cardinality is 0, when $r = 3, 5$ it is 1, and when $r = 7$ it is 2. $\square$

13.3 The Quadratic Reciprocity Law

With the lemma in hand, it is straightforward to deduce the quadratic reciprocity law.

Theorem 13.3.1 (Gauss). *Suppose that p and q are distinct odd primes. Then*

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

Proof. First suppose that $p \equiv q \pmod 4$. By swapping p and q if necessary, we may assume that $p > q$, and write $p - q = 4a$. Since $p = 4a + q$,

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

and

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

Proposition 17.2.4 implies that $\left(\frac{a}{q}\right) = \left(\frac{a}{p}\right)$, since $p \equiv q \pmod{4a}$. Thus

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

where the last equality is because $\frac{p-1}{2}$ is even if and only if $\frac{q-1}{2}$ is even.

Next suppose that $p \not\equiv q \pmod 4$, so $p \equiv -q \pmod 4$. Write $p + q = 4a$. We have

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

Since $p \equiv -q \pmod{4a}$, Proposition 17.2.4 implies that $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$. Since $(-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}} = 1$, the proof is complete. $\square$

13.3.1 Examples

Example 13.3.2. Is 6 a square modulo 389? We have

$$\left(\frac{6}{389}\right) = \left(\frac{2\cdot 3}{389}\right) = \left(\frac{2}{389}\right)\cdot\left(\frac{3}{389}\right) = (-1)\cdot(-1) = 1.$$

Here, we found that $\left(\frac{2}{389}\right) = -1$ using Proposition 13.2.3 and that $389 \equiv 3 \pmod 8$. We found $\left(\frac{3}{389}\right)$ as follows:

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

Thus 6 is a square modulo 389.

Annoyingly, though we know that 6 is a square modulo 389, we still don't know an x such that $x^2 \equiv 6 \pmod{389}$!

```
? for(a=1,388,if(Mod(a,389)^2==6,print1(a, " ")))
28 361
```

Example 13.3.3. Is 3 a square modulo $p = 726377359$? We proved that the answer is "no" in the previous lecture by computing $3^{p-1} \pmod p$. It's easier to prove that the answer is no using Theorem 13.3.1:

$$\left(\frac{3}{726377359}\right) = (-1)^{1\cdot\frac{726377358}{2}}\cdot\left(\frac{726377359}{3}\right) = -\left(\frac{1}{3}\right) = -1.$$

13.4 Some Homework Hints

Spend time studying for the midterm in addition to doing the homework. To point you in the right direction on the homework problems, here are some hints.

(i) Use the quadratic reciprocity law, just like in the above examples.

(ii) Use the quadratic reciprocity law.

(iii) Relate the statement for $n = 3$ to the statement for $n > 3$.

(iv) Write down an element of $(\mathbb{Z}/p^2\mathbb{Z})^*$ that looks like it might have order p, and prove that it does. Recall that if a, b have orders n, m, with $\gcd(n, m) = 1$, then ab has order nm.

(v)

(vi)

(vii) Replace $\sum \left(\frac{a}{p}\right)$ by $\sum \left(\frac{ab}{p}\right)$ and use that $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\cdot\left(\frac{b}{p}\right)$.

(viii) Write a little program.

Chapter 14

The Midterm Exam

Today I will briefly describe some key ideas that we've covered in this course up until now. Make sure you understand these, so you can do well on the midterm, which is Wednesday, October 17, and is worth 20% of your grade.

14.1 Some Basic Definitions

Greatest common divisor:

$$\gcd(a, b) = \max\{d : d \mid a \text{ and } d \mid b\}$$

Congruence: $a \equiv b \pmod{n}$ means that $n \mid a - b$.

Example 14.1.1. We have $7 \equiv -19 \pmod{13}$ since $13 \mid 7 - (-19) = 26$.

If a is an integer such that $\gcd(a, n) = 1$, then the order of a modulo n is

$$\min\left\{i \in \mathbb{N} : a^i \equiv 1 \pmod{n}\right\}.$$

For example, the order of 2 modulo 15 is 4.

Some Rings and Groups: We let $\mathbb{Z}/n\mathbb{Z}$ denote the ring of equivalence classes of integers modulo n. We also frequently consider the group

$$(\mathbb{Z}/n\mathbb{Z})^* = \{a : 1 \leq a \leq n \text{ and } \gcd(a, n) = 1\}.$$

The order of a modulo n is then the order of the image of a in the multiplicative group $(\mathbb{Z}/n\mathbb{Z})^*$.

14.2 Equations Modulo n

14.2.1 Linear Equations

The equation $ax \equiv b \pmod{n}$ must have a solution if $\gcd(a, n) = 1$. *Warning:* It might still have a solution even if $\gcd(a, n) \neq 1$.

Example 14.2.1. The equation $3x \equiv 2 \pmod{5}$ has the solution $x = 4$.
The equation $3x \equiv 9 \pmod{18}$ has a solution $x = 3$ even though $\gcd(3, 18) = 3 \neq 1$.

14.2.2 Quadratic Equations

Suppose a is an integer that is not divisible by p. The solvability or nonsolvability of the quadratic equation $x^2 \equiv a \pmod{p}$ is addressed by quadratic reciprocity. (So far we have not discussed how to find a solution, only whether or not one exists.) The quadratic residue symbol is

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } x^2 \equiv a \pmod{p} \text{ has a solution} \\ -1 & \text{otherwise.} \end{cases}$$

We have

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

The **Quadratic Reciprocity Law**, which was proved by Gauss, asserts that if p and q are distinct odd primes then

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

This is the deepest result that we've proved in the course so far. On the midterm, you will *not* be held responsible for understanding the proof I gave last Friday. However, you should know the statement of the quadratic reciprocity law and have some practice applying it.

14.3 Systems of Equations

Suppose that n and m are coprime integers. Then the **Chinese Remainder Theorem** (CRT) asserts that the system of equations

$$\begin{aligned} x &\equiv a \pmod{m}, \\ x &\equiv b \pmod{n} \end{aligned}$$

has solutions. (There is exactly one nonnegative solution $x < nm$.)

Example 14.3.1. Because of CRT, I know that there is an x such that

$$\begin{aligned} x &\equiv 1 \pmod{37}, \\ x &\equiv 17 \pmod{23} \end{aligned}$$

even though I am too lazy to find x right now.

14.4 The Euler φ Function

Define a function $\varphi : \mathbb{N} \to \mathbb{N}$ by

$$\varphi(n) = \#\{a : 1 \le a \le n \text{ and } \gcd(a,n) = 1\} = \#(\mathbb{Z}/n\mathbb{Z})^*.$$

Using the Chinese Remainder Theorem we proved that φ is a *multiplicative function*, i.e., if $m, n \in \mathbb{N}$ and $\gcd(m,n) = 1$, then

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

Also, if p is a prime then $\varphi(p^n) = p^n - \dfrac{p^n}{p} = p^n - p^{n-1}$.

Example 14.4.1.

$$\varphi(2^3 \cdot 5^2) = \varphi(2^3) \cdot \varphi(5^2) = (2^3 - 2^2) \cdot (5^2 - 5) = 4 \cdot 20 = 80.$$

14.5 Public-key Cryptography

14.5.1 The Diffie-Hellman Key Exchange

1. Nikita chooses a prime p and a number g that is a primitive root modulo p. She tells Michael both p and g.

2. Nikita secretely chooses a random number n and sends Michael $g^n \pmod p$.

3. Michael secretely chooses a random number m and sends Nikita $g^m \pmod p$.

4. The *secret key* is $s = g^{nm} \pmod p$. Both Michael and Nikita can easily compute s, but The Collective can't because of the difficulty of the "discrete logarithm problem".

14.5.2 The RSA Cryptosystem

1. Nikita creates her public key as follows:

   (a) She chooses two distinct large primes p and q, then computes both $n = pq$ and $\varphi(n) = (p-1)(q-1)$.

   (b) She picks a random natural number $e < \varphi(n)$ such that $\gcd(e, \varphi(n)) = 1$.

   (c) She computes a number d such that $ed \equiv 1 \pmod{\varphi(n)}$.

   (d) Her public key is (n, e). (And her private decoding key is d.)

2. To send Nikita a message, Michael encodes it (or a piece of it) as a number $m \pmod n$. He then sends $m^e \pmod n$ to Nikita.

3. Nikita recovers m from $m^e \pmod n$ by using that

$$m \equiv (m^e)^d \pmod n.$$

14.6 Important Algorithms

14.6.1 Euclid's Algorithm

Given integers a and b, a slight extension of Euclid's gcd algorithm enables us to find integers x and y such that

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

Example 14.6.1. $a = 12$, $b = 101$.

$$\begin{aligned} \underline{101} &= 8 \cdot \underline{12} + \underline{5} \qquad & \underline{5} &= \underline{101} - 8 \cdot \underline{12} \\ \underline{12} &= 2 \cdot \underline{5} + \underline{2} \qquad & \underline{2} &= -2 \cdot \underline{101} + 17 \cdot \underline{12} \\ \underline{5} &= 2 \cdot \underline{2} + \underline{1} \qquad & \underline{1} &= \underline{5} - 2 \cdot \underline{2} = 5 \cdot \underline{101} - 42 \cdot \underline{12}. \end{aligned}$$

Thus $x = -42$, $y = 5$ works, and $\gcd(a, b) = 1$.

We can use the result of this computation to solve

$$12x \equiv 1 \pmod{101}.$$

Indeed, $1 = (-42) \cdot 12 + 5 \cdot 101$, so $x = -42$ is a solution.

14.6.2 Powering Algorithm

There is a clever trick that makes computing a^n easier. Write n in binary, that is is write $n = \sum_{i=0}^{r} \varepsilon_i 2^i$ with $\varepsilon_i \in \{0, 1\}$. Then

$$a^n = \prod_{i \text{ with } \varepsilon_i \neq 0} a^{2^i}.$$

14.6.3 PARI

The midterm will **NOT** test knowledge of PARI.

14.7 The Midterm Exam

The students had 52 minutes to do all of these problems with no external aids such as a calculator or notes.

1. (5 points) Prove that a positive number n is divisible by 11 if and only if the alternating sum of the digits of n is divisible by 11.

2. Let $\varphi : \mathbb{N} \to \mathbb{N}$ be the Euler φ function.

   (i) (3 points) Find all natural numbers n such that $\varphi(n) = 1$.

   (ii) (2 points) Do there exist natural numbers m and n such that $\varphi(mn) \neq \varphi(m) \cdot \varphi(n)$?

3. (6 points) Nikita and Michael decide to agree on a secret encryption key using the Diffie-Hellman key exchange protocol. You observe the following:

   - Nikita chooses $p = 13$ for the modulus and $g = 2$ as generator.
   - Nikita sends 6 to Michael.
   - Michael sends 11 to Nikita.

   What is the secret key?

4. Consider the RSA public-key cryptosystem defined by $(n, e) = (77, 7)$.

   (i) (3 points) Encrypt the number 4 using this cryptosystem.

   (ii) (3 points) Find an integer d such that $ed \equiv 1 \pmod{\varphi(n)}$.

5. (5 points) How many natural numbers $x < 2^{13}$ satisfy the equation

   $$x^2 \equiv 5 \pmod{2^{13} - 1}?$$

   (You may assume that $2^{13} - 1$ is prime.)

6. (10 points) Which of the following systems of equations have at least one solution? Briefly justify your answers.

(a) $\begin{aligned} x &\equiv 1 \pmod 3 \\ x &\equiv -1 \pmod 9 \end{aligned}$

(b) $2x \equiv 1 \pmod{12345678910111213 14151}$

(c) $\begin{aligned} x^2 &\equiv 5 \pmod{29} \\ x^2 &\equiv 3 \pmod{47} \end{aligned}$

(d) $\begin{aligned} x &\equiv 3 \pmod{29} \\ x &\equiv 5 \pmod{47} \end{aligned}$

(e) $\begin{aligned} x^2 &\equiv 3 \pmod{29} \\ x^2 &\equiv 5 \pmod{47}. \end{aligned}$

7. (5 points) Find the natural number $x < 97$ such that $x \equiv 4^{48} \pmod{97}$. (You may assume that 97 is prime.)

14.8 Abbreviated Solutions

1. We use that $10 \equiv -1 \pmod{11}$ and facts about modular arithmetic. Write $n = \sum_{i=0}^{r} d_i 10^i$. Then $n \equiv \sum_{i=0}^{r} (-1)^i d_i \pmod{11}$, so $n \equiv 0 \pmod{11}$ if and only if the alternating sum of the digits $\sum(-1)^i d_i$ is congruent to 0 modulo 11.

2. For the first part, the answer is $\mathbf{n = 1, 2}$. On a previous homework we proved that $n = 1, 2$ are the only n such that $\varphi(n)$ is odd. For the second part, the fact that φ is multiplicative means that if $\gcd(m, n) = 1$ then $\varphi(mn) = \varphi(m) \cdot \varphi(n)$. When $\gcd(m, n) \neq 1$ this implication can fail; for example,
$$2 = \varphi(2 \cdot 2) \neq \varphi(2) \cdot \varphi(2) = 1.$$

3. Since $2^n \equiv 6 \pmod{13}$, a table of powers of 2 modulo 13 quickly reveals that n must be 5 (we solve the discrete log problem easily in this case since 13 is so small). Likewise, since $g^m \equiv 11 \pmod{13}$ we see that $m = 7$. The secret key is $\mathbf{s = 7}$ since $g^{nm} = 2^{35} \equiv 2^{11} \equiv 7 \pmod{13}$. (Some people who attempted this problem incorrectly thought the secret key should be $g^n \cdot g^m = g^{n+m}$.)

4. (a) We must compute $4^7 \pmod{77}$. Working modulo 77, we have that
$$4^7 = 64 \cdot 64 \cdot 4 = 13^2 \cdot 4 = 169 \cdot 4 = 15 \cdot 4 = 60,$$
so 4 encrypts as **60**.
   (b) First, $\varphi(n) = \varphi(77) = \varphi(7) \cdot \varphi(11) = 6 \cdot 10 = 60$. (Some people incorrectly thought that $\varphi(n) = 77$ for some reason.) We then use the extended Euclidean algorithm to find an integer e such that $7e \equiv 1 \pmod{60}$. We find that $2 \cdot 60 - 17 \cdot 7 = 1$, so $\mathbf{e = -17}$ is a solution.

5. First we use the law of quadratic reciprocity to decide whether or not there is a solution. We have
$$\left(\frac{5}{2^{13}-1}\right) = (-1)^{2 \cdot (2^{13}-2)/2} \left(\frac{2^{13}-1}{5}\right) = \left(\frac{1}{5}\right) = 1,$$
so the equation $x^2 \equiv 5 \pmod{2^{13}-1}$ has at least one solution a. Since the polynomial $x^2 - 5$ has degree two and $2^{13} - 1$ is prime, there are at most 2 solutions. Since $-a$ is also a solution and $a \neq 0$, there are **exactly two solutions**.

6. (a) has **no solutions** because $x \equiv -1 \pmod 9$ implies that $x \equiv -1 \pmod 3$. (b) has **a solution** because $\gcd(2, 1234567891011121314151) = 1$. (c) has **a solution** because $\left(\frac{5}{29}\right) = \left(\frac{3}{47}\right) = 1$ so there are a and b such that $a^2 \equiv 5 \pmod{29}$ and $b^2 \equiv 3 \pmod{47}$; the Chinese Remainder Theorem then implies that there is an x such that $x \equiv a \pmod{29}$ and $x \equiv b \pmod{47}$. (d) has **a solution** by the Chinese Remainder Theorem, since $\gcd(29, 47) = 1$. (e) has **no solution** since $\left(\frac{3}{29}\right) = -1$, so the first of the two equations doesn't even have a solution.

7. Since 97 is prime, Fermat's Little Theorem implies that $4^{48} = 2^{96} \equiv \mathbf{1} \pmod{97}$.

Chapter 15

Programming in PARI, II

15.1 Beyond One Liners

In today's relaxing but decidely non-mathematical lecture, you will learn a few new PARI programming commands. Feel free to try out variations of the examples below (especially because there is no homework due this coming Wednesday). Also, given that you know PARI fairly well by now, ask me questions during today's lecture!

15.1.1 Reading Files

The `\r` command allows you to read in a file.

Example 15.1.1. Create a file `pm.gp` that contains the following lines

```
{powermod(a, p, n) =
    return (lift(Mod(a,p)^n));}
```

Now use `\r` to load this little program into PARI:

```
> ?powermod
  ***   powermod: unknown identifier.
> \rpm               \\ \rpm.gp would do the same thing
? ?powermod
powermod(a, p, n) = return(lift(Mod(a,p)^n));
? powermod(2,101,7)
%1 = 27
```

If we change `pm.gp`, just type `\r` to reload it (omitting the file name reloads the last file loaded). For example, suppose we change `return (lift(Mod(a,p)^n))` in `pm.gp` to `return (lift(Mod(a,p)^n)-p)`. Then

```
? \r
? powermod(2,101,7)
%2 = -74
```

15.1.2 Arguments

PARI functions can have several arguments. For example,

```
{add(a, b, c)=
   return (a + b + c);}
? add(1,2,3)
%3 = 6
```

If you leave off arguments, they are set equal to 0.

```
? add(1,2)
%4 = 3
```

If you want the left-off arguments to default to something else, include that information in the declaration of the function:

```
{add(a, b=-1, c=2)=
   return (a + b + c);}
? add(1,2)
%6 = 5
? add(1)
%7 = 2
? add(1,2,3)
%8 = 6
```

15.1.3 Local Variables Done Right

Amidst the haste of a previous lecture, I mentioned that an unused argument can be used as a poor man's local variable. The following example illustrates the right way to declare local variables in PARI.

Example 15.1.2. The function `verybad` below sums the integers $1, 2, \dots n$ whilst wreaking havoc on the variable `i`.

```
{verybad(n)=
   i=0;
   for(j=1,n, i=i+j);
   return(i);}
? verybad(3)
%9 = 6
? i=4;
? verybad(3);
? i
%13 = 6                     \\ ouch!! what have you done to my eye!
```

The function `poormans` is better, but it uses a cheap hack to simulate a local variable.

```
{poormans(n,  i=0)=
   for(j=1,n, i=i+j);
   return(i);}
? i=4;
```

```
? poormans(3)
%16 = 6
? i
%17 = 4                \\ good
```

The following function is the best, because i is local and it's clearly declared as such.

```
{best(n)=
   local(i);
   i=0;  for(j=1,n, i=i+j);
   return(i);}
? i=4;
? best(3)
%18 = 6
? i
%19 = 4
```

15.1.4 Making Your Program Listen

The `input` command reads a PARI expression from the keyboard. The expression is evaluated and the result returned to your program. This behavior is at first disconcerting if, like me, you naively expect `input` to return a string. Here are some examples to illustrate the `input` command:

```
? ?input
input(): read an expression from the input file or standard input.
? s = input();
1+1
? s                    \\ s is not the string "1+1", as you might expect
%24 = 2
? s=input()
hi there
%25 = hithere
? type(s)              \\ PARI views s as a polynomial in the variable hithere
%26 = "t_POL"
? s=input()
"hi there"
%27 = "hi there"
? type(s)              \\ now it's a string
%28 = "t_STR"
```

15.1.5 Writing to Files

Use the `write` command:

```
? ?write
write(filename,a): write the string expression a to filename.
? write("testfile", "Hello Kitty!")
```

The write command above appended the line "Hello Kitty!" to the last line of testfile. This is useful if, e.g., you want to save key bits of work during a session or in a function. There is also a **logging facility** in PARI, which records most of what you type and PARI outputs to the file pari.log.

```
? \l
   log = 1 (on)
? 2+2
%29 = 4
? \l
   log = 0 (off)
   [logfile was "pari.log"]
```

15.2 Coming Attractions

The rest of this course is about continued fractions, quadratic forms, and elliptic curves. The following illustrates some relevant PARI commands which will help us to explore these mathematical objects.

```
? ?contfrac
contfrac(x,{b},{lmax}): continued fraction expansion of x ...
? contfrac(7/9)
%30 = [0, 1, 3, 2]
? contfrac(sqrt(2))
%31 = [1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, ...]
? ?qfbclassno
qfbclassno(x,{flag=0}): class number of discriminant x using Shanks's
method by default. If (optional) flag is set to 1, use Euler products.
? qfbclassno(-15,1)  \\ ALWAYS use flag=1, since ``the authors were too
%32 = 2              \\ lazy to implement Shanks' method completely...''
? E=ellinit([0,1,1,-2,0]);
? P=[0,0];
? elladd(E,P,P)
%36 = [3, 5]
? elladd(E,P,[3,5])
%37 = [-11/9, 28/27]
? a=-11/9;b=28/27;          \\ this is an ``amazing'' point on the curve.
? b^2+b == a^3+a^2-2*a
%38 = 1
```

Chapter 16

Continued Fractions, I

16.1 Introduction

A *continued fraction* is an expression of the form

$$a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \cdots}}}$$

which may or may not go on indefinitely. We denote[1] the value of this continued fraction by

$$[a_0, a_1, a_2, \ldots].$$

The a_n are called the *partial quotients* of the continued fraction (we will see why at the end of this lecture). Thus, e.g.,

$$[1, 2] = 1 + \frac{1}{2} = \frac{3}{2},$$

and

$$\frac{172}{51} = [3, 2, 1, 2, 6] = 3 + \cfrac{1}{2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{6}}}}.$$

Continued fractions have many applications, from the abstract to the concrete. They give good rational approximations to irrational numbers, and the have been used to understand why you can't tune a piano perfectly.[2] Continued fractions also suggest a sense in which e appears to be "less transcendental" than π.

There are many places to read about continued fractions, including Chapter X of Hardy and Wright's *Intro. to the Theory of Numbers*, §13.3 of Burton's *Elementary Number Theory*, Chapter IV of Davenport, and Khintchine's *Continued Fractions*. The notes you're reading right now draw primarily on Hardy and Wright, since their exposition is very clear and to the point. I found Davenport's chapter IV uneccessarily tedious; I felt marched through a thick jungle to see a beautiful river.

[1] *Warning:* This notation clashes with the notation used in Davenport. Our notation is standard.

[2] See `http://www.research.att.com/~njas/sequences/DUNNE/TEMPERAMENT.HTML`

16.2 Finite Continued Fractions

Definition 16.2.1. A *finite continued fraction* is an expression

$$a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_m}}},$$

where each a_n is a rational number and $a_n > 0$ for all $n \geq 1$. If the a_n are integers, we say that the continued fraction is *integral*.

To get a feeling for continued fractions, observe that

$$\begin{aligned}
[a_0] &= a_0, \\
[a_0, a_1] &= a_0 + \frac{1}{a_1} = \frac{a_0 a_1 + 1}{a_1}, \\
[a_0, a_1, a_2] &= a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{a_0 a_1 a_2 + a_0 + a_2}{a_1 a_2 + 1}.
\end{aligned}$$

Also,

$$\begin{aligned}
[a_0, a_1, \ldots, a_{m-1}, a_m] &= [a_0, a_1, \ldots, a_{m-2}, a_{m-1} + \frac{1}{a_m}] \\
&= a_0 + \frac{1}{[a_1, \ldots, a_m]} \\
&= [a_0, [a_1, \ldots, a_m]].
\end{aligned}$$

16.2.1 Partial Convergents

Fix a continued fraction $[a_0, \ldots, a_m]$.

Definition 16.2.2. For $0 \leq n \leq m$, the nth *convergent* of the continued fraction $[a_0, \ldots, a_m]$ is $[a_0, \ldots, a_n]$.

For each $n \geq -1$, define real numbers p_n and q_n as follows:

$$\begin{aligned}
p_{-1} &= 1, & p_0 &= a_0, & p_1 &= a_1 p_0 + p_{-1} = a_1 a_0 + 1, & p_n &= a_n p_{n-1} + p_{n-2}, \\
q_{-1} &= 0, & q_0 &= 1, & q_1 &= a_1 q_0 + q_{-1} = a_1, & q_n &= a_n q_{n-1} + q_{n-2}.
\end{aligned}$$

Exercise 16.2.3. [3] Compute p_n and q_n for the continued fractions $[-3, 1, 1, 1, 1, 3]$ and $[0, 2, 4, 1, 8, 2]$. Observe that the propositions below hold.

Proposition 16.2.4. $[a_0, \ldots, a_n] = \dfrac{p_n}{q_n}$

[3] Try to do this exercise, which is not part of the regular homework, before the next lecture.

Proof. We use induction. We already verified the assertion when $n = 0, 1$. Suppose the proposition is true for all continued fractions of length $n-1$. Then

$$\begin{aligned}
[a_0, \ldots, a_n] &= [a_0, \ldots, a_{n-2}, a_{n-1} + \frac{1}{a_n}] \\
&= \frac{\left(a_{n-1} + \frac{1}{a_n}\right) p_{n-2} + p_{n-3}}{\left(a_{n-1} + \frac{1}{a_n}\right) q_{n-2} + q_{n-3}} \\
&= \frac{(a_{n-1}a_n + 1)p_{n-2} + a_n p_{n-3}}{(a_{n-1}a_n + 1)q_{n-2} + a_n q_{n-3}} \\
&= \frac{a_n(a_{n-1}p_{n-2} + p_{n-3}) + p_{n-2}}{a_n(a_{n-1}q_{n-2} + q_{n-3}) + q_{n-2}} \\
&= \frac{a_n p_{n-1} + p_{n-2}}{a_n q_{n-1} + q_{n-2}} = \frac{p_n}{q_n}.
\end{aligned}$$

□

Proposition 16.2.5. *For $n \le m$,*

1. *the determinant of* $\begin{pmatrix} p_n & p_{n-1} \\ q_n & q_{n-1} \end{pmatrix}$ *is $(-1)^{n-1}$; equivalently,*

$$\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}};$$

2. *the determinant of* $\begin{pmatrix} p_n & p_{n-2} \\ q_n & q_{n-2} \end{pmatrix}$ *is $(-1)^n a_n$; equivalently,*

$$\frac{p_n}{q_n} - \frac{p_{n-2}}{q_{n-2}} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}}.$$

Proof. For the first statement, we proceed by induction. The case $n = 0$ holds because the determinant of $\begin{pmatrix} a_0 & 1 \\ 1 & 0 \end{pmatrix}$ is $-1 = (-1)^{-1}$. Suppose the statement is true for $n-1$. Then

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

This completes the proof of the first statement. For the second statement,

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

□

Corollary 16.2.6. *The fraction $\dfrac{p_n}{q_n}$ is in lowest terms.*

Proof. If $p \mid p_n$ and $p \mid q_n$ then $p \mid (-1)^{n-1}$. □

16.2.2 How the Convergents Converge

Let $[a_0, \ldots, a_m]$ be a continued fraction and for $n \leq m$ let

$$c_n = [a_0, \ldots, a_n] = \frac{p_n}{q_n}$$

denote the nth convergent.

Proposition 16.2.7. *The even convergents c_{2n} increase strictly with n, and the odd convergents c_{2n+1} decrease strictly with n. Moreover, the odd convergents c_{2n+1} are greater than all of the even convergents.*

Proof. For $n \geq 1$ the a_n are positive, so the q_n are all positive. By Proposition 16.2.5, for $n \geq 2$,

$$c_n - c_{n-2} = (-1)^n \cdot \frac{a_n}{q_n q_{n-2}},$$

which proves the first claim.

Next, Proposition 16.2.5 implies that for $n \geq 1$,

$$c_n - c_{n-1} = (-1)^{n-1} \cdot \frac{1}{q_n q_{n-1}}$$

has the sign of $(-1)^{n-1}$, so that $c_{2n+1} > c_{2n}$. Thus if there exists r, n such that $c_{2n+1} < c_{2r}$, then $r \neq n$. If $r < n$, then $c_{2n+1} < c_{2r} < c_{2n}$, a contradiction. If $r > n$, then $c_{2r+1} < c_{2n+1} < c_{2r}$, also a contradiction. $\square$

16.3 Every Rational Number is Represented

Proposition 16.3.1. *Every rational number is represented by a continued fraction.*

Proof. Let a/b, where $b > 0$, be any rational number. Euclid's algorithm gives:

$$\begin{aligned}
a &= b \cdot a_0 + r_1, & 0 < r_1 < b\\
b &= r_1 \cdot a_1 + r_2, & 0 < r_2 < r_1\\
&\cdots\\
r_{n-2} &= r_{n-1} \cdot a_{n-1} + r_n, & 0 < r_n < r_{n-1}\\
r_{n-1} &= r_n \cdot a_n + 0.
\end{aligned}$$

Note that $a_i > 0$ for $i > 0$. Rewrite the equations as follows:

$$\begin{aligned}
a/b &= a_0 + r_1/b = a_0 + 1/(b/r_1),\\
b/r_1 &= a_1 + r_2/r_1 = a_1 + 1/(r_1/r_2),\\
r_1/r_2 &= a_2 + r_3/r_2 = a_2 + 1/(r_2/r_3),\\
&\cdots\\
r_{n-1}/r_n &= a_n.
\end{aligned}$$

It follows that

$$\frac{a}{b} = [a_0, a_1, \ldots a_n].$$

$\square$

list of good links:
http://mathforum.org/electronic.newsletter/mf.intnews2.44.html

Chapter 17

Continued Fractions II: Infinite Continued Fractions

17.1 The Continued Fraction Algorithm

Let $x \in \mathbb{R}$ and write
$$x = a_0 + t_0$$
with $a_0 \in \mathbb{Z}$ and $0 \leq t_0 < 1$. If $t_0 \neq 0$, write
$$\frac{1}{t_0} = a_1 + t_1$$
with $a_1 \in \mathbb{N}$ and $0 \leq t_1 < 1$. Thus $t_0 = \frac{1}{a_1+t_1} = [0, a_1 + t_1]$, which is a (nonintegral) continued fraction expansion of t_0. Continue in this manner so long as $t_n \neq 0$ writing
$$\frac{1}{t_n} = a_{n+1} + t_{n+1}$$
with $a_{n+1} \in \mathbb{N}$ and $0 \leq t_{n+1} < 1$. This process, which associates to a real number x the sequence of integers $a_0, a_1, a_2, \ldots$, is called the *continued fraction algorithm.*

Example 17.1.1. Let $x = \frac{8}{3}$. Then $x = 2 + \frac{2}{3}$, so $a_0 = 2$ and $t_0 = \frac{2}{3}$. Then $\frac{1}{t_0} = \frac{3}{2} = 1 + \frac{1}{2}$, so $a_1 = 1$ and $t_1 = \frac{1}{2}$. Then $\frac{1}{t_1} = 2$, so $a_2 = 2$, $t_2 = 0$, and the sequence terminates. Notice that
$$\frac{8}{3} = [2, 1, 2],$$
so the continued fraction algorithm produces the continued fraction of $\frac{8}{3}$.

Proposition 17.1.2. *For every n such that a_n is defined, we have*
$$x = [a_0, a_1, \ldots, a_n + t_n],$$
and if $t_n \neq 0$ then $x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}]$.

Proof. Use induction. The statements are both true when $n = 0$. If the second statement is true for $n - 1$, then
$$x = [a_0, a_1, \ldots, a_{n-1}, \frac{1}{t_{n-1}}] = [a_0, a_1, \ldots, a_{n-1}, a_n + t_n] = [a_0, a_1, \ldots, a_{n-1}, a_n, \frac{1}{t_n}].$$

Similarly, the first statement is true for n if it is true for $n-1$.

$\square$

Example 17.1.3. Let $x = \frac{1+\sqrt{5}}{2}$. Then

$$x = 1 + \frac{-1+\sqrt{5}}{2},$$

so $a_0 = 1$ and $t_0 = \frac{-1+\sqrt{5}}{2}$. We have

$$\frac{1}{t_0} = \frac{2}{-1+\sqrt{5}} = \frac{-2-2\sqrt{5}}{-4} = \frac{1+\sqrt{5}}{2}$$

so again $a_1 = 1$ and $t_1 = \frac{-1+\sqrt{5}}{2}$. Likewise, $a_n = 1$ for all n. Does the following crazy-looking equality makes sense??

$$\frac{1+\sqrt{5}}{2} = 1 + \frac{1}{1 + \frac{1}{1+\frac{1}{1+\frac{1}{1+\frac{1}{1+\cdots}}}}}$$

Example 17.1.4. Next suppose $x = e$. Then

$$a_0, a_1, a_2, \ldots = 2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, \ldots$$

```
? contfrac(exp(1))
%1 = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1,
       12, 1, 1, 14, 1, 1, 16, 1, 1, 18, 1, 1, 20, 2]
? \\ to get more terms, increase the real precision:
? \p60
? contfrac(exp(1),[])
%12 = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1,
   12, 1, 1, 14, 1, 1, 16, 1, 1, 18, 1, 1, 20, 1, 1, 22, 1,
   1, 24, 1, 1, 26, 1, 1, 28, 1, 1, 30, 1, 1, 32, 1, 1, 34,
   1, 1, 36, 1, 1, 38, 1, 1, 40, 1, 1, 42, 2]
```

The following program uses a proposition we proved yesterday to compute the partial convergents of a continued fraction:

```
{convergents(v)=
    local(pp,qq,p,q,tp,tq,answer);
    pp=1; qq=0; p=v[1]; q=1;     \\ pp is p_{n-1} and p is p_n.
    answer = vector(length(v)); \\ put answer in this vector
    answer[1] = p/q;
    for(n=2,length(v),
       tp=p; tq=q; p=v[n]*p+pp; q=v[n]*q+qq; pp=tp; qq=tq;
       answer[n] = p/q;
    );
    return(answer);
}
```

Let's try this with π:

```
? contfrac(Pi)
%26 = [3, 7, 15, 1, 292, 1, 1, ...]
? convergents([3,7,15])
%27 = [3, 22/7, 333/106]
? convergents([3,7,15,1,292])
%28 = [3, 22/7, 333/106, 355/113, 103993/33102]
? %[5]*1.0
%29 = 3.1415926530119026040...
? % - Pi
%30 = -0.000000000577890634...
```

17.2 Infinite Continued Fractions

Theorem 17.2.1. *Let $a_0, a_1, a_2, \ldots$ be a sequence of integers such that $a_n > 0$ for all $n \geq 1$, and for each $n \geq 0$, set $c_n = [a_0, a_1, \ldots a_n]$. Then $\lim_{n\to\infty} c_n$ exists.*

Proof. For any $m \geq n$, the number c_n is a partial convergent of $[a_0, \ldots, a_m]$. Recall from the previous lecture that the even convergents c_{2n} form a strictly *increasing* sequence and the odd convergents c_{2n+1} form a strictly *decreasing* sequence. Moreover, the even convergents are all $\leq c_1$ and the odd convergents are all $\geq c_0$. Hence $\alpha_0 = \lim_{n\to\infty} c_{2n}$ and $\alpha_1 = \lim_{n\to\infty} c_{2n+1}$ both exist and $\alpha_0 \leq \alpha_1$. Finally, by a proposition from last time

$$|c_{2n} - c_{2n-1}| = \frac{1}{q_{2n} \cdot q_{2n-1}} \leq \frac{1}{2n(2n-1)} \to 0,$$

so $\alpha_0 = \alpha_1$. ∎

We define

$$[a_0, a_1, \ldots] = \lim_{n\to\infty} c_n.$$

Example 17.2.2. We use PARI to illustrate the convergence of the theorem for $x = \pi$.

```
? a = contfrac(Pi)
%38 = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, ...]
? c = convergents(a)
%39 = [3, 22/7, 333/106, 355/113, 103993/33102, 104348/33215, ...]
? \p9                          \\ so we can see.
   realprecision = 9 significant digits
? [c[1]*1.0, c[3]*1.0, c[5]*1.0, c[7]*1.0]  \\ odd ones converge up to pi
%43 = [3.00000000, 3.14150943, 3.14159265, 3.14159265]
? [c[2]*1.0,c[4]*1.0,c[6]*1.0,c[8]*1.0]     \\ even ones swoop down on pi.
%44 = [3.14285714, 3.14159291, 3.14159265, 3.14159265]
```

Theorem 17.2.3. *Let $x \in \mathbb{R}$ be a real number. Then*

$$x = [a_0, a_1, a_2, \ldots],$$

where $a_0, a_1, a_2, \ldots$ is the sequence produced by the continued fraction algorithm.

Proof. If the sequence is finite then some $t_n = 0$ and the result follows by Proposition 17.1.2. Suppose the sequence is infinite. By Proposition 17.1.2,

$$x = [a_0, a_1, \ldots, a_n, \frac{1}{t_n}].$$

By a proposition from the last lecture[1],

$$x = \frac{\frac{1}{t_n} p_n + p_{n-1}}{\frac{1}{t_n} q_n + q_{n-1}}.$$

Thus if $c_n = [a_0, a_1, \ldots, a_n]$, then

$$\begin{aligned}
x - c_n &= x - \frac{p_n}{q_n} \\
&= \frac{\frac{1}{t_n} p_n q_n + p_{n-1} q_n - \frac{1}{t_n} p_n q_n - p_n q_{n-1}}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)}. \\
&= \frac{p_{n-1} q_n - p_n q_{n-1}}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)} \\
&= \frac{(-1)^n}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)}.
\end{aligned}$$

Thus

$$\begin{aligned}
|x - c_n| &= \frac{1}{q_n \left(\frac{1}{t_n} q_n + q_{n-1}\right)} \\
&< \frac{1}{q_n (a_{n+1} q_n + q_{n-1})} \\
&= \frac{1}{q_n \cdot q_{n+1}} \leq \frac{1}{n(n+1)} \to 0.
\end{aligned}$$

(In the inequality we use that a_{n+1} is the integer part of $\frac{1}{t_n}$, and is hence $\leq \frac{1}{t_n}$.) $\square$

Proposition 17.2.4. *If x is a rational number then the sequence $a_0, a_1, a_2, \ldots$ terminates (at n say) and*

$$x = [a_0, a_1, a_2, \ldots, a_n].$$

Proof. Let $[b_0, b_1, \ldots, b_m]$ be the continued fraction representation of x that we obtain using the Euclidean algorithm. Then

$$x = b_0 + 1/[b_1, \ldots, b_m].$$

If $[b_1, \ldots, b_m] = 1$ then $m = 1$ and $b_1 = 1$, which would never happen using the Euclidean algorithm since x is expressed in lowest terms. Thus $[b_1, \ldots, b_m] > 1$, so in the continued fraction algorithm we choose $a_0 = b_0$ and $t_0 = 1/[b_1, \ldots, b_m]$. Repeating this argument enough times proves the claim. $\square$

[1] Which we apply in a case when the partial quotients of the continued fraction are not integers!

Chapter 18

Continued Fractions III: Quadratic Irrationals

In this lecture we prove that the continued fraction expansion of a number is periodic if and only if the number is a quadratic irrational.

18.1 Quadratic Irrationals

Definition 18.1.1. An element $\alpha \in \mathbb{R}$ is a *quadratic irrational* if it is irrational and satisfies a quadratic polynomial.

Thus, e.g., $(1+\sqrt{5})/2$ is a quadratic irrational. Recall that

$$\frac{1+\sqrt{5}}{2} = [1, 1, 1, \ldots].$$

The continued fraction of $\sqrt{2}$ is $[1, 2, 2, 2, 2, 2, \ldots]$, and the continued fraction of $\sqrt{389}$ is

$$[19, 1, 2, 1, 1, 1, 1, 2, 1, 38, 1, 2, 1, 1, 1, 1, 2, 1, 38, \ldots].$$

Does the $[1, 2, 1, 1, 1, 1, 2, 1, 38]$ pattern repeat over and over again??

18.2 Periodic Continued Fractions

Definition 18.2.1. A *periodic continued fraction* is a continued fraction $[a_0, a_1, \ldots, a_n, \ldots]$ such that

$$a_n = a_{n+h}$$

for a fixed positive integer h and all sufficiently large n. We call h the *period* of the continued fraction.

Example 18.2.2. Consider the periodic continued fraction $[1, 2, 1, 2, \ldots] = [\overline{1, 2}]$. What does it converge to?

$$[\overline{1, 2}] = 1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cdots}}}},$$

so if $\alpha = [\overline{1,2}]$ then

$$\alpha = 1 + \frac{1}{2+\alpha}.$$

Thus $2\alpha + \alpha^2 = 2 + \alpha + 1$, so

$$\alpha^2 + \alpha - 3 = 0 \quad \text{and} \quad \alpha = \frac{-1+\sqrt{7}}{2}.$$

Theorem 18.2.3. *An infinite integral continued fraction is periodic if and only if it represents a quadratic irrational.*

Proof. ($\Longrightarrow$) First suppose that

$$[a_0, a_1, \ldots, a_n, \overline{a_{n+1}, \ldots, a_{n+h}}]$$

is a periodic continued fraction. Set $\alpha = [a_{n+1}, a_{n+2}, \ldots]$. Then

$$\alpha = [a_{n+1}, \ldots, a_{n+h}, \alpha],$$

so

$$\alpha = \frac{\alpha p_{n+h} + p_{n+h-1}}{\alpha q_{n+h} + q_{n+h-1}}.$$

(We use that α is the last partial convergent.) Thus α satisfies a quadratic equation. Since the a_i are all integers, the number

$$\begin{aligned}[a_0, a_1, \ldots] &= [a_0, a_1, \ldots, a_n, \alpha] \\ &= a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \cdots + \alpha}}\end{aligned}$$

can be expressed as a polynomial in α with rational coefficients, so $[a_0, a_1, \ldots]$ also satisfies a quadratic polynomial. Finally, $\alpha \notin \mathbb{Q}$ because periodic continued fractions have infinitely many terms.

($\Longleftarrow$) This direction was first proved by Lagrange. The proof is much more exciting! Suppose $\alpha \in \mathbb{R}$ satisfies a quadratic equation

$$a\alpha^2 + b\alpha + c = 0$$

with $a, b, c \in \mathbb{Z}$. Let $[a_0, a_1, \ldots]$ be the expansion of α. For each n, let

$$r_n = [a_n, a_{n+1}, \ldots],$$

so that

$$\alpha = [a_0, a_1, \ldots, a_{n-1}, r_n].$$

We have

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

Substituting this expression for α into the quadratic equation for α, we see that

$$A_n r_n^2 + B_n r_n + C_n = 0,$$

where

$$\begin{aligned}
A_n &= ap_{n-1}^2 + bp_{n-1}q_{n-1} + cq_{n-1}^2,\\
B_n &= 2ap_{n-1}p_{n-2} + b(p_{n-1}q_{n-2} + p_{n-2}q_{n-1}) + 2cq_{n-1}q_{n-2},\\
C_n &= ap_{n-2}^2 + bp_{n-2}q_{n-2} + cp_{n-2}^2.
\end{aligned}$$

Note that $A_n, B_n, C_n \in \mathbb{Z}$, that $C_n = A_{n-1}$, and that

$$B^2 - 4A_nC_n = (b^2 - 4ac)(p_{n-1}q_{n-2} - q_{n-1}p_{n-2})^2 = b^2 - 4ac.$$

Recall from the proof of Theorem 2.3 of the previous lecture that

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

Thus

$$|\alpha q_{n-1} - p_{n-1}| < \frac{1}{q_n} < \frac{1}{q_{n+1}},$$

so

$$p_{n-1} = \alpha q_{n-1} + \frac{\delta}{q_{n-1}} \qquad \text{with } |\delta| < 1.$$

Hence

$$\begin{aligned}
A_n &= a\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right)^2 + b\left(\alpha q_{n-1} + \frac{\delta}{q_{n-1}}\right)q_{n-1} + cq_{n-1}^2\\
&= (a\alpha^2 + b\alpha + c)q_{n-1}^2 + 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta\\
&= 2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta.
\end{aligned}$$

Thus

$$|A_n| = \left|2a\alpha\delta + a\frac{\delta^2}{q_{n-1}^2} + b\delta\right| < 2|a\alpha| + |a| + |b|.$$

Thus there are only finitely many possibilities for the integer A_n. Also,

$$|C_n| = |A_{n-1}| \quad \text{and} \quad |B_n| = \sqrt{b^2 - 4(ac - A_nC_n)},$$

so there are only finitely many triples (A_n, B_n, C_n), and hence only finitely many possibilities for r_n as n varies. Thus for some $h > 0$,

$$r_n = r_{n+h}.$$

This shows that the continued fraction for α is periodic. □

18.3 What About Higher Degree?

Definition 18.3.1. An *algebraic number* is a root of a polynomial $f \in \mathbb{Q}[x]$.

Open Problem: [1] What is the continued fraction expansion of the algebraic number $\sqrt[3]{2}$?

[1]As far as I know this is still an open problem.

```
? contfrac(2^(1/3))
%5 = [1, 3, 1, 5, 1, 1, 4, 1, 1, 8, 1, 14, 1, 10, 2, 1, 4, 12, 2, 3,
 2, 1, 3, 4, 1, 1, 2, 14, 3, 12, 1, 15, 3, 1, 4, 534, 1, 1, 5, 1, 1,
 121, 1, 2, 2, 4, 10, 3, 2, 2, 41, 1, 1, 1, 3, 7, 2, 2, 9, 4, 1, 3, 7,
 6, 1, 1, 2, 2, 9, 3, 1, 1, 69, 4, 4, 5, 12, 1, 1, 5, 15, 1, 4, 1, 1,
 1, 1, 1, 89, 1, 22, 186, 5, 2, 4, 3, 3, 1, \ldots]
```

I sure don't see a pattern, and that 534 strips me of any confidence that I ever will. One could at least try to analyze the first few terms of the continued fraction statistically (see Lang and Trotter, 1972).

Khintchine (1963), page 59:

> No properties of the representing continued fractions, analogous to those which have just been proved, are known for algebraic numbers of higher degree. [...] It is of interest to point out that up till the present time no continued fraction development of an algebraic number of higher degree than the second is known. It is not even known if such a development has bounded elements. Generally speaking the problems associated with the continued fraction expansion of algebraic numbers of degree higher than the second are extremely difficult and virtually unstudied.

Richard Guy *Unsolved Problems in Number Theory* (1994), page 260:

> Is there an algebraic number of degree greater than two whose simple continued fraction has unbounded partial quotients? Does *every* such number have unbounded partial quotients?

Chapter 19

Continued Fractions IV: Applications

In this lecture we will learn about two applications of continued fractions. The first is a solution to the computational problem of recognizing a rational number using a computer. The second application is to the following ancient question: Given a positive nonsquare integer d, find *integers* x and y such that $x^2 - dy^2 = 1$.

19.1 Recognizing Rational Numbers

Suppose that you can compute approximations to a rational number using a computer, and desparately want to know what the rational number is. As Henri Cohen explains in his book *A Course in Computational Algebraic Number Theory*, continued fraction are very helpful.

> Consider the following apparently simple problem. Let $x \in \mathbb{R}$ be given by an approximation (for example a decimal or binary one). Decide if x is a rational number or not. Of course, this question as posed does not really make sense, since an approximation is usually itself a rational number. In practice however the question does make a lot of sense in many different contexts, and we can make it algorithmically more precise. For example, assume that one has an algorithm which allows us to compute x to as many decimal places as one likes (this is usually the case). Then, if one claims that x is (approximately) equal to a rational number p/q, this means that p/q should still be extremely close to x whatever the number of decimals asked for, p and q being fixed. This is still not completely rigorous, but it comes quite close to actual practice, so we will be content with this notion.
>
> Now how does one find p and q if x is indeed a rational number? The standard (and algorithmically excellent) answer is to compute the continued fraction expansion $[a_0, a_1, \ldots]$ of x. The number x is rational if and only if its continued fraction expansion is finite, i.e., if and only if one of the a_i is *infinite*. Since x is only given with the finite precision, x we be considered rational if x has a *very* large partial quotient a_i in its continued fraction expansion.

The following example illustrates Cohen's remarks:

Example 19.1.1.
```
? x
%13 = 9495/3847
? x*1.0
%14 = 2.4681570054587990642058747075643358461138549519105
? contfrac(x)
%15 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 2]
? contfrac(2.468157005458799064)
%16 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 328210621945, 2, 1, 1, 1, 1, 7]
? contfracpnqn([2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1])
%17 =
[9495 5852]
[3847 2371]
? contfrac(2.4681570054587990642058747075643)
%18 = [2, 2, 7, 2, 1, 5, 1, 1, 1, 1, 1, 1, 1885009518355562936415046, 1, 4]
? \p300
? x*1.0  \\ notice that no repeat is immediately evident in the digits of x
%19 = 2.468157005458799064205874707564335846113854951910579672472056147647517..
? \\ in fact, the length of the period of the decimal expansion
  \\ of 1/3847 is 3846 (the order of 10 modulo 3847)!!
```

19.2 Pell's Equation

In February of 1657, Pierre Fermat issued the following challenge:

> Given a positive integer d, find a positive integer y such that $dy^2 + 1$ is a perfect square.

In other words, find a solution to $x^2 - dy^2 = 1$ with $y \in \mathbb{N}$.

Note Fermat's emphasis on *integer* solutions. It is easy to find rational solutions to the equation $x^2 - dy^2 = 1$. Simply divide the relation

$$(r^2 + d)^2 - d(2r)^2 = (r^2 - d)^2$$

by $(r^2 - d)^2$ to arrive at

$$x = \frac{r^2 + d}{r^2 - d}, \qquad y = \frac{2r}{r^2 - d}.$$

Fermat said: "Solutions in fractions, which can be given at once from the merest elements of arithmetic, do not satisfy me."

The equation $x^2 - dy^2 = 1$ is called **Pell's equation**. This is because Euler (in about 1759) accidently called it "Pell's equation" and the name stuck, though Pell (1611–1685) had nothing to do with it.

If d is a perfect square, $d = n^2$, then

$$(x + ny)(x - ny) = x^2 - dy^2 = 1$$

which implies that $x + ny = x - ny = 1$, so

$$x = \frac{x + ny + x - ny}{2} = \frac{1 + 1}{2} = 1.$$

We will thus always assume that d is not a perfect square. You can read about Pell's equation in Section 0.6 of Kato-Kurokawa-Saito and on pages 107–111 of Davenport. Pell's equation is best understood in terms of units in real quadratic fields.

19.3 Units in Real Quadratic Fields

Let d be a nonsquare positive integer, and set

$$\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}$$
$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Z}\}.$$

Then $\mathbb{Q}(\sqrt{d})$ is a *real quadratic field* and $\mathbb{Z}[\sqrt{d}]$ is a ring. There is a homomorphism called norm:

$$N : \mathbb{Q}(\sqrt{d})^* \to \mathbb{Q}^*, \qquad N\left(a + b\sqrt{d}\right) = \left(a + b\sqrt{d}\right)\left(a - b\sqrt{d}\right) = a^2 - b^2 d.$$

Definition 19.3.1. An element $x \in R$ is a *unit* if there exists $y \in R$ such that $xy = 1$.

Proposition 19.3.2. *The units of $\mathbb{Z}[\sqrt{d}]$ are exactly the elements of norm ± 1 in $\mathbb{Z}[\sqrt{d}]$.*

Proof. Suppose $u \in \mathbb{Z}[\sqrt{d}]$ is a unit. Then

$$1 = N(1) = N(uu^{-1}) = N(u) \cdot N(u^{-1}).$$

Since $N(u), N(u^{-1}) \in \mathbb{Z}$, we have $N(u) = N(u^{-1}) = \pm 1$ □

Thus Fermat's challenge amounts to determing the group U^+ of units in $\mathbb{Z}[\sqrt{d}]$ of the form $a + b\sqrt{d}$ with $a, b \geq 0$.

Theorem 19.3.3. *The group U^+ is an infinite cyclic group. It is generated by $p_m + q_m\sqrt{d}$, where $\frac{p_m}{q_m}$ is one of the partial convergents of the continued fraction expansion of $\sqrt{d}$. (In fact, if m is the period of the continued fraction of $\sqrt{d}$ then $n = m - 1$ when m is even and $2n - 1$ when m is odd.)*

The theorem implies that *Pell's equation always has a solution*! Warning: the smallest solution is typically shockingly large. For example, the value of x in the smallest solution to $x^2 - 1000099y^2 = 1$ has **1118 digits**.

The following example illustrates how to use Theorem 19.3.3 to solve Pell's equation when $d = 61$, where the simplest solution is already quite large.

Example 19.3.4. Suppose $d = 61$. Then

$$\sqrt{d} = [7, \overline{1, 4, 3, 1, 2, 2, 1, 3, 4, 1, 14}],$$

which has odd period $n = 11$. Thus the group U^+ is generated by

$$x = p_{21} = 1766319049$$
$$y = q_{21} = 226153980.$$

That is, we have

$$U^+ = \langle u \rangle = \langle 1766319049 + 226153980\sqrt{61} \rangle,$$

and $x = 1766319049$, $y = 226153980$ gives a solution to $x^2 - dy^2 = 1$. All the other solutions arise from u^n for some n. For example,

$$u^2 = 6239765965720528801 + 798920165762330040\sqrt{61}$$

leads to another solution.

Remark 19.3.5. To help with your homework, note that if the equation

$$x^2 - dy^2 = n$$

has at least one (nonzero) solution $(x_0, y_0) \in \mathbb{Z} \times \mathbb{Z}$, then it must have infinitely many solutions. This is because if $x_0^2 - dy_0^2 = n$ and u is a generator of the cyclic group U^+, then for any integer i,

$$N(u^i(x_0 + y_0\sqrt{d})) = N(u^i) \cdot N(x_0 + y_0\sqrt{d}) = 1 \cdot n = n,$$

so

$$x_1 + y_1\sqrt{d} = u^i(x_0 + y_0\sqrt{d})$$

provides another solution to $x^2 + dy^2 = n$.

19.4 Some Proofs

The rest of this lecture is devoted to proving most of Theorem 19.3.3. We will prove that partial convergents to continued fractions contribute infinitely many solutions to Pell's equation. We will not prove that every solution to Pell's equation is a partial convergent, though this is true.[1]

Fix a positive nonsquare integer d.

Definition 19.4.1. A quadratic irrational $\alpha = a + b\sqrt{d}$ is *reduced* if $\alpha > 1$ and if the conjugate of α, denoted by α', satisfies $-1 < \alpha' < 0$.

For example, the number $\alpha = 1 + \sqrt{2}$ is reduced.

Definition 19.4.2. A continued fraction is *purely periodic* if it is of the form $[\overline{a_0, a_1, \ldots, a_n}]$.

The continued fraction $[\overline{2}]$ of $1 + \sqrt{2}$ is purely periodic.

Lemma 19.4.3. *If α is a reduced quadratic irrational, then the continued fraction expansion of α is purely periodic. (The converse is also true, and is easy to prove.)*

Proof. The proof can be found on pages 102–103 of Davenport's book. □

Lemma 19.4.4. *The continued fraction expansion of $\sqrt{d}$ is of the form*

$$[a_0, \overline{a_1, \ldots, a_{n-1}, 2a_0}].$$

[1]There is a complete proof in Section 13.5 of Burton's *Elementary Number Theory*. It just involves more of the same sort of computations that we've been doing with continued fractions.

Proof. Let a_0 be the floor of $\sqrt{d}$. Then $\alpha = \sqrt{d} + a_0$ is reduced because $\alpha > 1$ and $\alpha' = -\sqrt{d} + a_0$ satisfies $-1 < \alpha' < 0$. Let $[a_0, a_1, a_2, \ldots]$ be the continued fraction expansion of $\sqrt{d}$. Then the continued fraction expansion of $\sqrt{d} + a_0$ is $[2a_0, a_1, a_2, \ldots]$. By Lemma 19.4.3, the continued fraction expansion of $\sqrt{d} + a_0$ is purely periodic, so

$$[2a_0, a_1, a_2, \ldots] = [\overline{2a_0, a_1, a_2, \ldots, a_{n-1}}],$$

where n is the period. It follows that $a_n = 2a_0$, as claimed. $\square$

The following proposition shows that there are infinitely many solutions to Pell's equation that arise from continued fractions.

Proposition 19.4.5. *Let p_k/q_k be the partial convergents of the continued fraction expansion of $\sqrt{d}$, and let n be the period of the expansion of $\sqrt{d}$. Then*

$$p_{kn-1}^2 - dq_{kn-1}^2 = (-1)^{kn}$$

for $k = 1, 2, 3, \ldots$.

Proof. [2] By Lemma 19.4.4, for $k \geq 1$, the continued fraction of $\sqrt{d}$ can be written in the form

$$\sqrt{d} = [a_0, a_1, a_2, \ldots, a_{kn-1}, r_{kn}]$$

where

$$r_{kn} = [2a_0, \overline{a_1, a_2, \ldots, a_n}] = a_0 + \sqrt{d}.$$

Because $\sqrt{d}$ is the last partial convergent of the continued fraction above, we have

$$\sqrt{d} = \frac{r_{kn}p_{kn-1} + p_{kn-2}}{r_{kn}q_{kn-1} + q_{kn-2}}.$$

Upon substituting $r_{kn} = a_0 + \sqrt{d}$ and simplifying, this reduces to

$$\sqrt{d}(a_0 a_{kn-1} + q_{kn-2} - p_{kn-1}) = a_0 p_{kn-1} + p_{kn-2} - dq_{kn-1}.$$

Because the right-hand side is rational and $\sqrt{d}$ is irrational,

$$a_0 a_{kn-1} + q_{kn-2} = p_{kn-1}, \quad \text{and} \quad a_0 p_{kn-1} + p_{kn-2} = dq_{kn-1}.$$

Multiplying the first of these equations by p_{kn-1} and the second by $-q_{kn-1}$, and then adding them, gives

$$p_{kn-1}^2 - dq_{kn-1}^2 = p_{kn-1}q_{kn-2} - q_{kn-1}p_{kn-2}.$$

But

$$p_{kn-1}q_{kn-2} - q_{kn-1}p_{kn-2} = (-1)^{kn-2} = (-1)^{kn},$$

which proves the proposition. $\square$

[2]This proof is from Section 13.5 of Burton's *Elementary Number Theory.*

Chapter 20

Binary Quadratic Forms I: Sums of Two Squares

Today we study the question of which integers are the sum of two squares.

20.1 Sums of Two Squares

During the next four lectures, we will study binary quadratic forms. A simple example of a binary quadratic form that will occupy us today is

$$x^2 + y^2.$$

A typical question that one asks about a quadratic form is which integers does it represent. "Are there integers x and y so that $x^2 + y^2 = 389$? So that $x^2 + y^2 = 2001$?"

20.1.1 Which Numbers are the Sum of Two Squares?

The main goal of today's lecture is to prove the following theorem.

Theorem 20.1.1. *A number n is a sum of two squares if and only if all prime factors of n of the form $4m + 3$ have even exponent in the prime factorization of n.*

Before tackling a proof, we consider a few examples.

Example 20.1.2.

- $5 = 1^2 + 2^2$.
- 7 is not a sum of two squares.
- 2001 is divisible by 3 because $2 + 1$ is, but not by 9 since $2 + 1$ is not, so 2001 is *not* a sum of two squares.
- $2 \cdot 3^4 \cdot 5 \cdot 7^2 \cdot 13$ is a sum of two squares.
- 389 is a sum of two squares, since $389 \equiv 1 \pmod 4$ and 389 is prime.
- $21 = 3 \cdot 7$ is *not* a sum of two squares even though $21 \equiv 1 \pmod 4$.

In preparation for the proof of Theorem 20.1.1, we recall a result that emerged when we analyzed how partial convergents of a continued fraction converge.

Lemma 20.1.3. *If $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then there is a fraction $\frac{a}{b}$ in lowest terms such that $0 < b \leq n$ and*

$$\left| x - \frac{a}{b} \right| \leq \frac{1}{b(n+1)}.$$

Proof. Let $[a_0, a_1, \ldots]$ be the continued fraction expansion of x. As we saw in the proof of Theorem 2.3 in Lecture 18, for each m

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}}.$$

Since q_{m+1} is always at least 1 bigger than q_m and $q_0 = 1$, either there exists an m such that $q_m \leq n < q_{m+1}$, or the continued fraction expansion of x is finite and n is larger than the denominator of the rational number x. In the first case,

$$\left| x - \frac{p_m}{q_m} \right| < \frac{1}{q_m \cdot q_{m+1}} \leq \frac{1}{q_m \cdot (n+1)},$$

so $\frac{a}{b} = \frac{p_m}{q_m}$ satisfies the conclusion of the lemma. In the second case, just let $\frac{a}{b} = x$. $\square$

Definition 20.1.4. A representation $n = x^2 + y^2$ is *primitive* if $\gcd(x, y) = 1$.

Lemma 20.1.5. *If n is divisible by a prime p of the form $4m + 3$, then n has no primitive representations.*

Proof. If n has a primitive representation, $n = x^2 + y^2$, then

$$p \mid x^2 + y^2 \quad \text{and} \quad \gcd(x, y) = 1,$$

so $p \nmid x$ and $p \nmid y$. Thus $x^2 + y^2 \equiv 0 \pmod{p}$ so, since $\mathbb{Z}/p\mathbb{Z}$ is a field we can divide by y^2 and see that

$$(x/y)^2 \equiv -1 \pmod{p}.$$

Thus the quadratic residue symbol $\left(\frac{-1}{p}\right)$ equals $+1$. However,

$$\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}} = (-1)^{\frac{4m+3-1}{2}} = (-1)^{2m+1} = -1.$$

$\square$

Proof of Theorem 20.1.1. ($\Longrightarrow$) Suppose that p is of the form $4m + 3$, that $p^r \parallel n$ (exactly divides) with r odd, and that $n = x^2 + y^2$. Letting $d = \gcd(x, y)$, we have

$$x = dx', \quad y = dy', \quad n = d^2 n'$$

with $\gcd(x', y') = 1$ and

$$(x')^2 + (y')^2 = n'.$$

Because r is odd, $p \mid n'$, so Lemma 20.1.5 implies that $\gcd(x', y') > 1$, a contradiction.

($\Longleftarrow$) Write $n = n_1^2 n_2$ where n_2 has no prime factors of the form $4m+3$. It suffices to show that n_2 is a sum of two squares. Also note that

$$(x_1^2 + y_1^2)(x_2^2 + y_2^2) = (x_1x_2 + y_1y_2)^2 + (x_1y_2 - x_2y_1)^2,$$

so a product of two numbers that are sums of two squares is also a sum of two squares.[1] Also, the prime 2 is a sum of two squares. It thus suffices to show that if p is a prime of the form $4m+1$, then p is a sum of two squares.

Since

$$(-1)^{\frac{p-1}{2}} = (-1)^{\frac{4m+1-1}{2}} = +1,$$

-1 is a square modulo p; i.e., there exists r such that $r^2 \equiv -1 \pmod{p}$. Taking $n = \lfloor \sqrt{p} \rfloor$ in Lemma 20.1.3 we see that there are integers a, b such that $0 < b < \sqrt{p}$ and

$$\left| -\frac{r}{p} - \frac{a}{b} \right| \leq \frac{1}{b(n+1)} < \frac{1}{b\sqrt{p}}.$$

If we write

$$c = rb + pa$$

then

$$|c| < \frac{pb}{b\sqrt{p}} = \frac{p}{\sqrt{p}} = \sqrt{p}$$

and

$$0 < b^2 + c^2 < 2p.$$

But $c \equiv rb \pmod{p}$, so

$$b^2 + c^2 \equiv b^2 + r^2b^2 \equiv b^2(1 + r^2) \equiv 0 \pmod{p}.$$

Thus $b^2 + c^2 = p$. $\square$

20.1.2 Computing x and y

Suppose p is a prime of the form $4m+1$. There is a construction of Legendre of x and y that is explained on pages 120–121 of Davenport. I'm unconvinced that it is any more efficient than the following naive algorithm: compute $\sqrt{p - x^2}$ for $x = 1, 2, \ldots$ until it's an integer. This takes at most $\sqrt{p}$ steps. Here's a simple PARI program which implements this algorithm.

```
{sumoftwosquares(n) =
   local(y);
   for(x=1,floor(sqrt(n)),
      y=sqrt(n-x^2);
      if(y-floor(y)==0, return([x,floor(y)]))
   );
   error(n," is not a sum of two squares.")
}
```

[1] This algebraic identity is secretely the assertion that the norm map $N : \mathbb{Q}(i)^* \to \mathbb{Q}^*$ sending $x + iy$ to $(x+iy)(x-iy) = x^2 + y^2$ is a homomorphism.

20.2 Sums of More Squares

Every natural number is a sum of **four** squares. See pages 124–126 of Davenport for a proof.

A natural number is a sum of **three** squares if and only if it is not a power of 4 times a number that is congruent to 7 modulo 8. For example, 7 is not a sum of three squares. This is more difficult to prove.

Chapter 21

Binary Quadratic Forms II: Basic Notions

21.1 Introduction

A *binary quadratic form* is a homogeneous polynomial

$$ax^2 + bxy + cy^2 \in \mathbb{Z}[x, y].$$

(There is a theory of quadratic forms in n-variables, but we will not study it in this course.) Chapter VI of Davenport's book is clear and well written. Read it.

The Classic Problem: Given a binary quadratic form $f(x, y) = ax^2 + bxy + cy^2$, what is the set of integers $\{f(x, y) : x, y \in \mathbb{Z}\}$?

That is, for which integers n are there integers x and y such that

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

We gave a clean answer to this question in the last lecture in the case when $f(x, y) = x^2 + y^2$. The set of sums of two squares is the set of integers n such that any prime divisor p of n of the form $4m + 3$ exactly divides n to an even power (along with 0). In your homework (Problem 5), you will give a simple answer to the question of which numbers are of the form $x^2 + 2y^2$. Is there a simple answer in general?

21.2 Equivalence

Definition 21.2.1. The modular group $\mathrm{SL}_2(\mathbb{Z})$ is the group of all 2×2 integer matrices with determinant $+1$.

If $g = \left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and $f(x, y) = ax^2 + bxy + cy^2$ is a quadratic form, let

$$f|_g(x, y) = f(px + qy, rx + sy) = f\left(\begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right),$$

where for simplicity we will sometimes write $f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right)$ for $f(x, y)$.

Proposition 21.2.2. *The above formula defines a right action of the group* $\mathrm{SL}_2(\mathbb{Z})$ *on the set of binary quadratic forms, in the sense that*

$$f|_{gh} = (f|_g)|_h.$$

Proof.

$$f|_{gh}(x,y) = f\left(gh\begin{bmatrix}x\\y\end{bmatrix}\right) = f|_g\left(h\left(\begin{bmatrix}x\\y\end{bmatrix}\right)\right) = (f|_g)|_h(x,y).$$

□

Proposition 21.2.3. *Let* $g \in \mathrm{SL}_2(\mathbb{Z})$ *and let* $f(x,y)$ *be a binary quadratic form. The set of integers represented by* $f(x,y)$ *is exactly the same as the set of integers represented by* $f|_g(x,y)$.

Proof. If $f(x_0,y_0) = n$ then since $g^{-1} \in \mathrm{SL}_2(\mathbb{Z})$, we have $g^{-1}\begin{bmatrix}x_0\\y_0\end{bmatrix} \in \mathbb{Z}^2$, so

$$f|_g\left(g^{-1}\begin{bmatrix}x_0\\y_0\end{bmatrix}\right) = f(x_0,y_0) = n.$$

Thus every integer represented by f is also represented by $f|_g$. Conversely, if $f|_g(x_0,y_0) = n$, then $f\left(g\begin{bmatrix}x_0\\y_0\end{bmatrix}\right) = n$, so f represents n. □

Define an equivalence relation $\sim$ on the set of all binary quadratic forms by declaring that f is equivalent to f' if there exists $g \in \mathrm{SL}_2(\mathbb{Z})$ such that $f|_g = f'$.

For simplicity, we will sometimes denote the quadratic form $ax^2 + bxy + cy^2$ by (a,b,c). Then, for example, since $g = \left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we see that $(a,b,c) \sim (c,-b,a)$, since if $f(x,y) = ax^2+bxy+cy^2$, then $f(-y,x) = ay^2 - bxy + cx^2$.

Example 21.2.4. Consider the binary quadratic form

$$f(x,y) = 458x^2 + 214xy + 25y^2.$$

Solving the representation problem for f might, at first glance, look hopeless. We find $f(x,y)$ for a few values of x and y:

$$\begin{aligned}
f(-1,-1) &= 17\cdot 41\\
f(-1,0) &= 2\cdot 229\\
f(0,-1) &= 5^2\\
f(1,1) &= 269\\
f(-1,2) &= 2\cdot 5\cdot 13\\
f(-1,3) &= 41
\end{aligned}$$

Each number is a sum of two squares! Letting $g = \left(\begin{smallmatrix}4 & -3\\ -17 & 13\end{smallmatrix}\right)$, we have

$$f|_g = 458(4x-3y)^2 + 214(4x-3y)(-17x+13y) + 25(-17x+13y)^2 = \cdots = x^2+y^2!!$$

By Proposition 21.2.3, f represents an integer n if and only if n is a sum of two squares.

21.3 Discriminants

Definition 21.3.1. The *discriminant* of $f(x, y) = ax^2 + bxy + cy^2$ is $b^2 - 4ac$.

Example 21.3.2. $\mathrm{disc}(x^2 + y^2) = -4$ and

$$\mathrm{disc}(458, 214, 25) = 214^2 - 4 \cdot 25 \cdot 458 = -4.$$

That the discriminants are the same is a good hint that $(1, 0, 1)$ and $(458, 214, 25)$ are closely related. Inspecting discriminants is more effective than simply computing $f(x, y)$ for many values of x and y and staring at the result.

Proposition 21.3.3. *If $f \sim f'$, then* $\mathrm{disc}(f) = \mathrm{disc}(f')$.

Proof. By tedious but elementary algebra (see page 133 of Davenport's book), one sees that if $g \in \mathrm{SL}_2(\mathbb{Z})$, then

$$\mathrm{disc}(f|_g) = \mathrm{disc}(f) \cdot (\det(g))^2 = \mathrm{disc}(f).$$

Since $f' = f|_g$ for some $g \in \mathrm{SL}_2(\mathbb{Z})$, the proposition follows. $\square$

WARNING: The converse of the proposition is false! Forms with the same discriminant need not be equivalent. For example, the forms $(1, 0, 6)$ and $(2, 0, 3)$ have discriminant -24, but are not equivalent. To see this, observe that $(1, 0, 6)$ represents 1, but $2x^2 + 3y^2$ does not represent 1.

Proposition 21.3.4. *The set of all discriminants of forms is exactly the set of integers d such that $d \equiv 0$ or $1 \pmod 4$.*

Proof. First note that $b^2 - 4ac$ is a square modulo 4, so it must equal 0 or 1 modulo 4. Next suppose d is an integer such that $d \equiv 0$ or $1 \pmod 4$. If we set

$$c = \begin{cases} -d/4, & \text{if } d \equiv 0 \pmod 4 \\ -(d-1)/4 & \text{if } d \equiv 1 \pmod 4, \end{cases}$$

then $\mathrm{disc}(1, 0, c) = d$ in the first case and $\mathrm{disc}(1, 1, c) = d$ in the second. $\square$

Definition 21.3.5. The form $(1, 0, -d/4)$ or $(1, 1, -(d-1)/4)$ of discriminant d that appears in the proof of the previous proposition is called the *principal form* of discriminant d.

d	principal form	
-4	(1,0,1)	$x^2 + y^2$
5	(1,1,-1)	$x^2 + xy - y^2$
-7	(1,1,2)	$x^2 + xy + 2y^2$
8	(1,0,-2)	$x^2 - 2y^2$
-23	(1,1,6)	$x^2 + xy + 6y^2$
389	(1,1,-97)	$x^2 + xy - 97y^2$

21.4 Definite and Indefinite Forms

Definition 21.4.1. A quadratic form with negative discriminant is called *definite*. A form with positive discriminant is called *indefinite*.

Let (a, b, c) be a quadratic form. Multiply by $4a$ and complete the square:

$$\begin{aligned} 4a(ax^2 + bxy + cy^2) &= 4a^2x^2 + 4abxy + 4acy^2 \\ &= (2ax + by)^2 + (4ac - b^2)y^2 \end{aligned}$$

If $\mathrm{disc}(a, b, c) < 0$ then $4ac - b^2 = -\mathrm{disc}(a, b, c) > 0$, so $ax^2 + bxy + cy^2$ takes only positive or only negative values, depending on the sign of a. In this sense, (a, b, c) is very definite about its choice of sign. If $\mathrm{disc}(a, b, c) > 0$, then $(2ax+by)^2+(4ac-b^2)y^2$ takes both positive and negative values, so (a, b, c) does also.

We will consider only definite forms in the next two lectures.

21.5 Real Life

The following text is from the documentation for binary quadratic forms in the MAGMA computer algebra system. A quick scan of the buzzwords emphasized (by me) below conveys an idea of where binary quadratic forms appear in mathematics.

> A binary quadratic form is an integral form $ax^2 + bxy + cy^2$ which is represented in MAGMA by a tuple (a, b, c). Binary quadratic forms play an central role in the *ideal theory of quadratic fields*, the classical theory of *complex multiplication*, and the theory of *modular forms*. Algorithms for binary quadratic forms provide efficient means of computing in the *ideal class group of orders* in a *quadratic field*. By using the explicit relation of definite quadratic forms with lattices with nontrivial endomorphism ring in the complex plane, one can apply *modular and elliptic functions* to forms, and exploit the analytic theory of complex multiplication.
>
> The structures of quadratic forms of a given discriminant D correspond to ordered bases of ideals in an order in a *quadratic number field*, defined up to scaling by the rationals. A form is primitive if the coefficients a, b, and c are coprime. For negative discriminants the primitive reduced forms in this structure are in bijection with the *class group of projective or invertible ideals*. For positive discriminants, the reduced orbits of forms are used for this purpose. Magma holds efficient algorithms for composition, enumeration of reduced forms, *class group computations*, and *discrete logarithms*. A significant novel feature is the treatment of nonfundamental discriminants, corresponding to nonmaximal orders, and the collections of homomorphisms between different class groups coming from the inclusions of these orders.
>
> The functionality for binary quadratic forms is rounded out with various functions for applying modular and elliptic functions to forms, and for *class polynomials* associated to *class groups* of definite forms.

Chapter 22

Binary Quadratic Forms III: Reduction Theory

Recall that a binary quadratic form is a function $f(x, y) = ax^2 + bxy + cy^2$. Our motivating problem is to decide which numbers are "represented" by f; i.e., for which integers n do there exist integers x, y such that $ax^2 + bxy + cy^2 = n$? If $g \in \mathrm{SL}_2(\mathbb{Z})$ then $f(x, y)$ and $f|_g(x, y) = f\left(g\begin{bmatrix}x\\y\end{bmatrix}\right)$ represent exactly the same set of integers. Also, $\mathrm{disc}(f) = \mathrm{disc}(f|_g)$, where $\mathrm{disc}(f) = b^2 - 4ac$, and f is called *positive definite* if $\mathrm{disc}(f) < 0$ and $a > 0$.

In today's lecture, we will learn about reduction theory, which allows us to decide whether or not two positive definite binary quadratic forms are equivalent under the action of $\mathrm{SL}_2(\mathbb{Z})$.

If, in the future, you would like to pursue the theory of binary quadratic forms in either a more algebraic or algorithmic direction, I highly recommend that you look at Chapter 5 of Henri Cohen's book *A Course in Computational Algebraic Number Theory* (GTM 138).

22.1 Reduced Forms

Definition 22.1.1 (Reduced). A positive definite quadratic form (a, b, c) is *reduced* if $|b| \leq a \leq c$ and if, in addition, when one of the two inequalities is an equality (i.e., either $|b| = a$ or $a = c$), then $b \geq 0$.

There is a geometric interpretation of reduced, which we will not use this later. Let $D = \mathrm{disc}(a, b, c) = b^2 - 4ac$ and set $\tau = \frac{-b+\sqrt{D}}{2a}$, so τ is the root of $ax^2 + bx + c$ with positive imaginary part. The right action of $\mathrm{SL}_2(\mathbb{Z})$ on positive definite binary quadratic forms corresponds to the left action of $\mathrm{SL}_2(\mathbb{Z})$ by linear fractional transformations on the complex upper half plane $\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. The standard fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathfrak{h}$ is

$$\mathcal{F} = \left\{\tau \in \mathfrak{h} : \mathrm{Re}(\tau) \in \left[-\frac{1}{2}, \frac{1}{2}\right), |\tau| > 1 \text{ or } |\tau| = 1 \text{ and } \mathrm{Re}(\tau) \leq 0\right\}.$$

Then (a, b, c) is reduced if and only if the corresponding complex number τ lies in $\mathcal{F}$. For example, if (a, b, c) is reduced then $\mathrm{Re}(\tau) = -b/2a \in [-1/2, 1/2)$ since $|b| \leq a$

and if $|b| = a$ then $b \geq 0$. Also

$$|\tau| = \sqrt{\frac{b^2 + 4ac - b^2}{4a^2}} = \sqrt{\frac{c}{a}} \geq 1$$

and if $|\tau| = 1$ then $b \geq 0$ so $\mathrm{Re}(\tau) \leq 0$.

The following theorem (which is not proved in Davenport) highlights the importance of reduced forms.

Theorem 22.1.2. *There is exactly one reduced form in each equivalence class of positive definite binary quadratic forms.*

Proof. We have to prove two things. First, that every class contains at least one reduced form, and second that this reduced form is the only one in the class.

We first prove that there is a reduced form in every class. Let $\mathcal{C}$ be an equivalence class of positive definite quadratic forms of discriminant D. Let (a, b, c) be an element of $\mathcal{C}$ such that a is minimal (amongst elements of $\mathcal{C}$). Note that for any such form we have $c \geq a$, since (a, b, c) is equivalent to $(c, -b, a)$ (use the matrix $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$). Applying the element $\left(\begin{smallmatrix} 1 & k \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ to (a, b, c) for a suitably chosen integer k (precisely, $k = \lfloor (a-b)/2a \rfloor$) results in a form (a', b', c') with $a' = a$ and $b' \in (-a', a']$. Since $a' = a$ is minimal, we have just as above that $a' \leq c'$, hence (a', b', c') is "just about" reduced. The only possible remaining problem would occur if $a' = c'$ and $b' < 0$. In that case, changing (a', b', c') to $(c'', b'', a'') = (c', -b', a')$ results in an equivalent form with $b'' > 0$, so that (c'', b'', a'') is reduced.

Next suppose (a, b, c) is a reduced form. We will now establish that (a, b, c) is the only reduced form in its equivalence class. First, we check that a is minimal amongst all forms equivalent to (a, b, c). Indeed, every other a' has the form $a' = ap^2 + bpr + cr^2$ with p, r coprime integers (see this by hitting (a, b, c) by $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$). The identities

$$ap^2 + bpr + cr^2 = ap^2\left(1 + \frac{b}{a}\frac{r}{p}\right) + cr^2 = ap^2 + cr^2\left(1 + \frac{b}{c}\frac{p}{r}\right)$$

then imply our claim since $|b| \leq a \leq c$ (use the first identity if $r/p < 1$ and the second otherwise). Thus any other reduced form (a', b', c') equivalent to (a, b, c) has $a' = a$. But the same identity implies that the only forms equivalent to (a, b, c) with $a' = a$ are obtained by applying a transformation of the form $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ (corresponding to $p = 1$, $r = 0$). Thus $b' = b + 2ak$ for some k. Since $a = a'$ we have $b, b' \in (-a, a]$, so $k = 0$. Finally

$$c' = \frac{(b')^2 - D}{4a'} = \frac{b^2 - D}{4a} = c,$$

so $(a', b', c') = (a, b, c)$. □

22.2 Finding an Equivalent Reduced Form

Here is how to find the reduced form equivalent to a given positive definite form (a, b, c). This algorithm is useful for solving problems 8 and 9 on the homework assignment. Consider the following two operations, which can be used to diminish one of a and $|b|$, without altering the other:

1. If $c < a$, replace (a,b,c) by the equivalent form $(c,-b,a)$.

2. If $|b| > a$, replace (a,b,c) by the equivalent form (a,b',c') where $b' = b+2ka$ and k is chosen so that $b' \in (-a,a]$ (more precisely, $k = \lfloor \frac{a-b}{2a} \rfloor$), and c' is found from the fact that $(b')^2 - 4ac' = D = \mathrm{disc}(a,b,c)$, so $c' = \frac{(b')^2-D}{4a}$.

Starting with (a,b,c), if you iterate the appropriate operation, eventually you will find the reduced form that is equivalent to (a,b,c).

Example 22.2.1. Let $f = 458x^2 + 214xy + 25y^2$.

Equivalent form	What I did	Matrix
$(458,214,25)$		
$(25,-214,458)$	(1)	$\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$
$(25,-14,2)$	(2) with $k=4$	$\left(\begin{smallmatrix} 1 & 4 \\ 0 & 1 \end{smallmatrix}\right)$
$(2,14,25)$	(1)	$\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$
$(2,2,1)$	(2) with $k=-3$	$\left(\begin{smallmatrix} 1 & -3 \\ 0 & 1 \end{smallmatrix}\right)$
$(1,-2,2)$	(1)	$\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$
$(1,0,1)$	(2) with $k=1$	$\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$

Let

$$\begin{aligned} g &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 4 \\ -13 & -17 \end{pmatrix}. \end{aligned}$$

Then

$$f|_g = x^2 + y^2!$$

22.3 Some PARI Code

The following PARI code checks whether or not a form is reduced, and computes the reduced form equivalent to a given form. You can download it from my web page if you don't want to type it in.

```
\\ true if and only if (a,b,c) is reduced.
{isreduced(a,b,c) =
   if(b^2-4*a*c>=0 || a<0,
      error("reduce: (a,b,c) must be positive definite."));
   if(!(abs(b)<=a && a<=c), return(0));
   if(abs(b)==a || a==c, return(b>=0));
   return(1);
}

\\ reduces, printing out each step.  returns the reduced form
\\ and a matrix that transforms the input form to the reduced form.
{reduce(a,b,c,s) =
   local(D, k, t, g);
   D=b^2-4*a*c;
```

```
   if(D>=0 || a<0, error("reduce: (a,b,c) must be positive definite."));
   g=[1,0;0,1];
   while(!isreduced(a,b,c),       \\ ! means ‘‘not’’
      if(c<a,
         b = -b; t = a; a = c; c = t;
         g = g*[0,-1;1,0];
         print([a,b,c], "  \t(1)"),  \\ backslash t means ‘‘tab’’
      \\ else
         if (abs(b)>a || -b==a,
            k = floor((a-b)/(2*a));
            b = b+2*k*a;
            c = (b^2-D)/(4*a);
            g = g*[1,k;0,1];
            print([a,b,c], "  \t(2) with k=",k)
         )
      )
   );
   return([a,b,c,g])
}

/* Here is an example:
? \r quadform
? reduce(458,214,25)
[25, -214, 458] (1)
[25, -14, 2]    (2) with k=4
[2, 14, 25]     (1)
[2, 2, 1]       (2) with k=-3
[1, -2, 2]      (1)
[1, 0, 1]       (2) with k=1
%22 = [1, 0, 1, [3, 4; -13, -17]]
*/
```

Chapter 23

Binary Quadratic Forms IV: The Class Group

23.1 Can You Hear the Shape of a Lattice?

After Lecture 23, Emanuele Viola asked me whether or not the following is true: *"If f_1 and f_2 are binary quadratic forms that represent exactly the same integers, is $f_1 \sim f_2$?"* The answer is no. For example, $f_1 = (2, 1, 3) = 2x^2 + xy + 3y^2$ and $f_2 = (2, -1, 3) = 2x^2 - xy + 3y^2$ are inequivalent reduced positive definite binary quadratic forms that represent exactly the same integers. Note that $\mathrm{disc}(f_1) = \mathrm{disc}(f_2) = -23$. There appears to be a sense in which all counterexamples resemble the one just given.

Questions like these are central to John H. Conway's book *The sensual (quadratic) form*, which I've never seen because the Cabot library copy is checked out and the Birkhoff copy has gone missing. The following is taken from the MathSciNet review (I changed the text slightly so that it makes sense):

> Chapter 2 begins by posing Mark Kac's question of "hearing the shape of a drum", and the author relates the higher-dimensional analogue of this idea on tori—quotients of $\mathbf{R}^n$ by a lattice—to the question of what properties of a positive definite integral quadratic form are determined by the numbers the form represents. A property of such a form is called "audible" if the property is determined by these numbers, or equivalently, by the theta function of the quadratic form. As examples, he shows that the determinant of the form and the theta function of the dual form are audible. He also provides counterexamples to the higher-dimensional Kac question, the first of which were found by J. Milnor...

23.2 Class Numbers

Proposition 23.2.1. *Let $D < 0$ be a discriminant. There are only finitely many equivalence classes of positive definite binary quadratic forms of discriminant D.*

Proof. Since there is exactly one reduced binary quadratic form in each equivalence class, it suffices to show that there are only finitely many reduced forms of discriminant D. Recall that if a form (a, b, c) is reduced, then $|b| \leq a \leq c$. If (a, b, c) has discriminant D then $b^2-4ac = D$. Since $b^2 \leq a^2 \leq ac$, we have $D = b^2-4ac \leq -3ac$, so

$$3ac \leq -D.$$

There are only finitely many positive integers a, c that satisfy this inequality. □

Definition 23.2.2. A binary quadratic form (a, b, c) is *primitive* if $\gcd(a, b, c) = 1$.

Definition 23.2.3. The *class number* h_D of discriminant $D < 0$ is the number of equivalence classes of primitive positive definite binary quadratic forms of discriminant D.

I computed the following table of class number h_D for $-D \leq 839$ using the built-in PARI function `qfbclassno(D,1)`. Notice that there are just a few 1s at the beginning and then no more.

$-D$	h_D	$-D$	h_D	$-D$	h_D	$-D$	h_D	$-D$	h_D	$-D$	h_D
3	**1**	123	2	243	3	363	4	483	4	603	4
7	**1**	127	5	247	6	367	9	487	7	607	13
11	**1**	131	5	251	7	371	8	491	9	611	10
15	2	135	6	255	12	375	10	495	16	615	20
19	**1**	139	3	259	4	379	3	499	3	619	5
23	3	143	10	263	13	383	17	503	21	623	22
27	**1**	147	2	267	2	387	4	507	4	627	4
31	3	151	7	271	11	391	14	511	14	631	13
35	2	155	4	275	4	395	8	515	6	635	10
39	4	159	10	279	12	399	16	519	18	639	14
43	**1**	163	**1**	283	3	403	2	523	5	643	3
47	5	167	11	287	14	407	16	527	18	647	23
51	2	171	4	291	4	411	6	531	6	651	8
55	4	175	6	295	8	415	10	535	14	655	12
59	3	179	5	299	8	419	9	539	8	659	11
63	4	183	8	303	10	423	10	543	12	663	16
67	**1**	187	2	307	3	427	2	547	3	667	4
71	7	191	13	311	19	431	21	551	26	671	30
75	2	195	4	315	4	435	4	555	4	675	6
79	5	199	9	319	10	439	15	559	16	679	18
83	3	203	4	323	4	443	5	563	9	683	5
87	6	207	6	327	12	447	14	567	12	687	12
91	2	211	3	331	3	451	6	571	5	691	5
95	8	215	14	335	18	455	20	575	18	695	24
99	2	219	4	339	6	459	6	579	8	699	10
103	5	223	7	343	7	463	7	583	8	703	14
107	3	227	5	347	5	467	7	587	7	707	6
111	8	231	12	351	12	471	16	591	22	711	20
115	2	235	2	355	4	475	4	595	4	715	4
119	10	239	15	359	19	479	25	599	25	719	31

$-D$	h_D
723	4
727	13
731	12
735	16
739	5
743	21
747	6
751	15
755	12
759	24
763	4
767	22
771	6
775	12
779	10
783	18
787	5
791	32
795	4
799	16
803	10
807	14
811	7
815	30
819	8
823	9
827	7
831	28
835	6
839	33

We can compute these numbers using Proposition 23.2.1. The following PARI program enumerates the primitive reduced forms of discriminant D.

```
{isreduced(a,b,c) =
   if(b^2-4*a*c>=0 || a<0,
      error("reduce: (a,b,c) must be positive definite."));
   if(!(abs(b)<=a && a<=c), return(0));
   if(abs(b)==a || a==c, return(b>=0));
   return(1);
}
{reduce(f) =
   local(D, k, t, a,b,c);
   a=f[1]; b=f[2]; c=f[3]; D=b^2-4*a*c;
   if(D>=0 || a<0, error("reduce: (a,b,c) must be positive definite."));
   while(!isreduced(a,b,c),       \\ ! means ‘‘not’’
      if(c<a,
         b = -b; t = a; a = c; c = t,
```

```
      \\ else
         if (abs(b)>a || -b==a,
            k = floor((a-b)/(2*a));
            b = b+2*k*a;
            c = (b^2-D)/(4*a);
         )
      )
   );
   return([a,b,c])
}
{reducedforms(D)=
   local(bound, forms, b, r);
   if (D > 0 || D%4 == 2 || D%4==3, error("Invalid discriminant"));
   bound = floor(-D/3);
   forms = [];
   for(a = 1, bound,
      for(c = 1, bound,
         if(3*a*c<=-D && issquare(4*a*c+D),
            b = floor(sqrt(4*a*c+D));
            r = reduce([a,b,c]);
            print1([a,b,c], " ----> ", r);
            if (gcd(r[1],gcd(r[2],r[3])) == 1,
               forms = setunion(forms,[r]); print(""),
               \\ else
               print ("  \t(not primitive)")
            )
         )
      )
   );
   return(eval(forms));    \\ eval gets rid of the annoying quotes.
}
```

For example, when $D = -419$ the program finds exactly 9 reduced forms:

```
? D = -419
%21 = -419
? qfbclassno(D,1)
%22 = 9
? reducedforms(D)
[1, 1, 105] ----> [1, 1, 105]
[1, 3, 107] ----> [1, 1, 105]
[1, 5, 111] ----> [1, 1, 105]
[1, 7, 117] ----> [1, 1, 105]
[1, 9, 125] ----> [1, 1, 105]
[1, 11, 135] ----> [1, 1, 105]
[3, 1, 35] ----> [3, 1, 35]
[3, 5, 37] ----> [3, -1, 35]
[3, 7, 39] ----> [3, 1, 35]
[3, 11, 45] ----> [3, -1, 35]
```

```
[5, 1, 21] ----> [5, 1, 21]
[5, 9, 25] ----> [5, -1, 21]
[5, 11, 27] ----> [5, 1, 21]
[7, 1, 15] ----> [7, 1, 15]
[9, 7, 13] ----> [9, 7, 13]
[9, 11, 15] ----> [9, -7, 13]
[13, 7, 9] ----> [9, -7, 13]
[15, 1, 7] ----> [7, -1, 15]
[15, 11, 9] ----> [9, 7, 13]
[21, 1, 5] ----> [5, -1, 21]
[25, 9, 5] ----> [5, 1, 21]
[27, 11, 5] ----> [5, -1, 21]
[35, 1, 3] ----> [3, -1, 35]
[37, 5, 3] ----> [3, 1, 35]
[39, 7, 3] ----> [3, -1, 35]
[45, 11, 3] ----> [3, 1, 35]
[105, 1, 1] ----> [1, 1, 105]
[107, 3, 1] ----> [1, 1, 105]
[111, 5, 1] ----> [1, 1, 105]
[117, 7, 1] ----> [1, 1, 105]
[125, 9, 1] ----> [1, 1, 105]
[135, 11, 1] ----> [1, 1, 105]
%23 = [[1, 1, 105], [3, -1, 35], [3, 1, 35], [5, -1, 21], [5, 1, 21],
       [7, -1, 15], [7, 1, 15], [9, -7, 13], [9, 7, 13]]
? length(%23)
%24 = 9
```

Theorem 23.2.4 (Heegner, Stark-Baker, Goldfeld-Gross-Zagier). *Suppose D is a negative discriminant that is either square free or 4 times a square-free number. Then*

- $h_D = 1$ *only for* $D = -3, -4, -7, -8, -11, -19, -43, -67, -163$.
- $h_D = 2$ *only for* $D = -15, -20, -24, -35, -40, -51, -52, -88, -91, -115, -123, -148, -187, -232, -235, -267, -403, -427$.
- $h_D = 3$ *only for* $D = -23, -31, -59, -83, -107, -139, -211, -283, -307, -331, -379, -499, -547, -643, -883, -907$.
- $h_D = 4$ *only for* $D = -39, -55, -56, -68, \ldots, -1555$.

To quote Henri Cohen: "The first two statements concerning class numbers 1 and 2 are very difficult theorems proved in 1952 by Heegner and in 1968–1970 by Stark and Baker. The general problem of determing all imaginary quadratic fields with a given class number has been solved in principle by Goldfeld-Gross-Zagier, but to my knowledge the explicit computations have been carried to the end only for class numbers 3 and 4 (in addition to the already known class numbers 1 and 2).

23.3 The Class Group

There are *much* more sophisticated ways to compute h_D than simply listing the reduced binary quadratic forms of discriminant D, which is an $O(|D|)$ algorithm. For example, there is an algorithm that can compute h_D for D having 50 digits in a reasonable amount of time. These more sophisticated algorithms use the fact that the set of primitive positive definite binary quadratic forms of given discriminant is a finite abelian group.

Definition 23.3.1. Let $f_1 = (a_1, b_1, c_1)$ and $f_2 = (a_2, b_2, c_2)$ be two quadratic forms of the same discriminant D. Set $s = (b_1 + b_2)/2$, $n = (b_1 - b_2)/2$ and let u, v, w and d be such that

$$ua_1 + va_2 + ws = d = \gcd(a_1, a_2, s)$$

(obtained by two applications of Euclid's algorithm), and let $d_0 = \gcd(d, c_1, c_2, n)$. Define the composite of the equivalence classes of the two forms f_1 and f_2 to be the equivalence class of the form

$$(a_3, b_3, c_3) = \left(d_0 \frac{a_1 a_2}{d^2}, b_2 + \frac{2a_2}{d}(v(s - b_2) - wc_2), \frac{b_3^2 - D}{4a_3}\right).$$

This mysterious-looking group law is induced by "multiplication of ideals" in the "ring of integers" of the quadratic imaginary number field $\mathbb{Q}(\sqrt{D})$. The following PARI program computes this group operation:

```
{composition(f1, f2)=
   local(a1,b1,c1,a2,b2,c2,D,s,n,bz0,bz1,u,v,w);
   a1=f1[1]; b1=f1[2]; c1=f1[3];
   a2=f2[1]; b2=f2[2]; c2=f2[3];
   D = b1^2 - 4*a1*c1;
   if(b2^2 - 4*a2*c2 != D, error("Forms must have the same discriminant."));
   s = (b1+b2)/2;
   n = (b1-b2)/2;
   bz0 = bezout(a1,a2);
   bz1 = bezout(bz0[3],s);
   u = bz1[1]*bz0[1];
   v = bz1[1]*bz0[2];
   w = bz1[2];
   d = bz1[3];
   d0 = gcd(gcd(gcd(d,c1),c2),n);
   a3 = d0*a1*a2/d^2;
   b3 = b2+2*a2*(v*(s-b2)-w*c2)/d;
   c3 = (b3^2-D)/(4*a3);
   f3 = reduce([a3,b3,c3]);
   return(f3);
}
```

Let's try the group out in the case when $D = -23$.

```
? reducedforms(-23)
```

```
[1, 1, 6] ----> [1, 1, 6]
[2, 1, 3] ----> [2, 1, 3]
[3, 1, 2] ----> [2, -1, 3]
[6, 1, 1] ----> [1, 1, 6]
%56 = [[1, 1, 6], [2, -1, 3], [2, 1, 3]]
```

Thus the group has elements $(1,1,6)$, $(2,-1,3)$, and $(2,1,3)$. Since $h_{-23} = 3$, the group must be cyclic of order 3. Let's find the identity element.

```
? composition([1,1,6],[2,-1,3])
%58 = [2, -1, 3]
```

Thus the identity element must be $(1,1,6)$. The element $(2,-1,3)$ is a generator for the group:

```
? composition([2,-1,3],[2,-1,3])
%59 = [2, 1, 3]
? composition([2,-1,3],[2,1,3])
%60 = [1, 1, 6]
```

Chapter 24

Elliptic Curves 1: Introduction

24.1 The Definition

Finally we come to elliptic curves, which I think are the most exciting and central *easily accessibly* objects in modern number theory. There are so many exciting things to tell you about elliptic curves, that the course is suddenly going to move more quickly than before.

Definition 24.1.1. An *elliptic curve* E over a field K is a plane cubic curve of the form

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where $a_1, a_2, a_3, a_4, a_6 \in K$ and

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \neq 0,$$

where

$$b_2 = a_1^2 + 4a_2, \quad b_4 = 2a_4 + a_1a_3, \quad b_6 = a_3^2 + 4a_6.$$

Help! Don't worry, when 2 and 3 are not equal to 0 in K, using completing the square and a little algebra we find a change of coordinates that transforms the above cubic equation into the form

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

and then $\Delta = -16(4a^3 + 27b^2)$. We will consider only elliptic curves of the form $y^2 = x^3 + ax + b$ for a while.

Hey! That's not an ellipse! You're right, elliptic curves are *not ellipses*; they are curves that first arose when 19th century mathematicians studied integral formulas for the arc lengths of ellipses.

In these lectures, I'll give you a glimpse into two main ways in which elliptic curves feature in mathematics. On the left hand, they provide the simplest example of a class of diophantine equations that we still can't totally solve. On the right hand, when K is a finite field (or, more sneakily, a finite ring), elliptic curves can be used as a tool for both making and breaking cryptosystems.

24.2 Linear and Quadratic Diophantine Equations

Consider the following question:

> Let $F(x, y)$ be an irreducible polynomial in two variables over $\mathbb{Q}$. Find all rational numbers x_0, y_0 such that $F(x_0, y_0) = 0$.

When F is linear, this problem is easy. The equation

$$F(x, y) = ax + by + c = 0$$

defines a line, and letting $y = t$, the solutions are

$$\left\{ \left(-\frac{b}{a}t - \frac{c}{a}, t \right) : t \in \mathbb{Q} \right\}.$$

When F is quadratic, the solution is not completely trivial, but it is well understood. In this case, the equation $F = 0$ has infinitely many rational solutions if and only if it has at least one solution. Moreover, it is easy to describe all solutions when there is one. If (x_0, y_0) is a solution and L is a non-tangent line through (x_0, y_0), then L will intersect the curve $F = 0$ in exactly one other point (x_1, x_1). Also $x_1, y_1 \in \mathbb{Q}$ since a quadratic polynomial over $\mathbb{Q}$ with 1 rational root has both roots rational. Thus the rational points on $F = 0$ are in bijection with the slopes of lines through (x_0, y_0).

Chapter 2 of [Kato et al.] is about how to decide whether or not an F of degree 2 has a rational point. The answer is that $F = 0$ has a rational solution if and only if $F = 0$ has a solution with $x_0, y_0 \in \mathbb{R}$ and a solution with $x_0, y_0 \in \mathbb{Q}_p$ for every "p-adic field" $\mathbb{Q}_p$. This condition, though it might sound foreboding, is easy to check in practice. I encourage you to flip through chapter 2 of loc. cit.

24.3 Points on Elliptic Curves

Next suppose that F is an irreducible cubic polynomial. The question of whether or not $F = 0$ has a rational solution is still an *open problem*! We will not consider this problem further until we discuss the Birch and Swinnerton-Dyer conjecture.

Suppose that $F = 0$ has a given rational solution. Then one can change coordinates so that the question of finding the rational solutions to $F = 0$ is equivalent to the problem of finding all rational points on the elliptic curve

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

Recall that when F has degree 2 we can use a given rational point P on the graph of $F = 0$ to find all other rational points by intersecting a line through P with the graph of $F = 0$. The graph of $y^2 = x^3 + ax + b$ looks like

[egg and curvy line] or [curvier line]

Notice that if P is a point on the graph of the curve, then a line through P (usually) intersects the graph in exactly *two* other points. In general, these two other points usually do not have rational coordinates. However, if P and Q are rational points on the graph of $y^2 = x^3 + ax + b$ and L is the line through P and Q, then the third point of intersection with the graph will have rational coordinates. Explicitly, if $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ then the third point of intersection has coordinates[1]

$$x_3 = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2 - x_1 - x_2, \qquad y_3 = \frac{y_2 - y_1}{x_2 - x_1}x_3 - \frac{y_2x_1 - y_1x_2}{x_2 - x_1}.$$

Thus, given two points on E, we can find another. Also, given a single point, we can draw the tangent line to E through that point and obtain a third point.

24.3.1 To Infinity!

At first glance, the above construction doesn't work if $x_1 = x_2$. [draw picture]. Fortunately, there is a natural sense in which the graph of E is missing one point, and when $x_1 = x_2$ this one missing point is the third point of intersection.

The graph of E that we drew above is a graph in the plan $\mathbb{R}^2$. The plane is a subset of the projective plane $\mathbb{P}^2$, which I will define in just a moment. The closure of the graph of $y^2 = x^3 + ax + b$ in $\mathbb{P}^2$ has exactly one extra point, which has rational coordinates, and which we denote by ∞. Formally, $\mathbb{P}^2$ can be viewed as the set of triples (a, b, c) with a, b, c not all 0 modulo the equivalence relation

$$(a, b, c) \sim (\lambda a, \lambda b, \lambda c)$$

for any nonzero λ. Denote by $(a : b : c)$ the equivalence class of (a, b, c). The closure of the graph of $y^2 = x^3 + ax + b$ is the graph of $y^2z = x^3 + axz^2 + bz^3$ and the extra point ∞ is $(0 : 1 : 0)$.

Venerable Problem: Find an algorithm that, given an elliptic curve E over $\mathbb{Q}$, outputs a complete description of the set of rational points (x_0, y_0) on E.

This problem is difficult. In fact, so far it has stumped everyone! There is a conjectural algorithm, but nobody has succeeded in proving that it is really an algorithm, in the sense that it terminates for any input curve E. Several of your profs at Harvard, including Barry Mazur, myself, and Christophe Cornut (who will teach Math 129 next semester) have spent, or will probably spend, a huge chunk of their life thinking about this problem. (Am I being overly pessimistic?)

How could one possible "describe" the set of rational points on E in the first place? In 1923, Louis Mordell proved an amazing theorem, which implies that there is a reasonable way to describe the rational points on E. To state his theorem, we introduce the "group law" on E.

24.4 The Group Law

Consider the set $E(\mathbb{Q}) = \{\infty\} \cup \{(x_0, y_0) : y_0^2 = x_0^3 + ax_0 + b\}$. There is a natural way to endow the set $E(\mathbb{Q})$ with a *group* structure. Here's how it works. First, the element $\infty \in E(\mathbb{Q})$ is the 0 element of the group. Next, suppose P and Q are elements of $E(\mathbb{Q})$. Just like we did earlier, draw the line through P and Q and let $R = (x_3, y_3)$ be the third point of intersection. Define $P + Q = (x_3, -y_3)$. There

[1] It is traditional in a course like ours for me to derive these formulas. I'm not going to, because it's simple algebra and once you see the geometric picture it is easy to carry out. You should do this as an exercise, or read the derivation in [Kato et al.] or [Davenport].

are various special cases to consider, such as when $P = Q$ or the third point of intersection is ∞, but I will let your read about them in [Kato et al.]. It is clear that this binary operation on $E(\mathbb{Q})$ satisfies $P + Q = Q + P$. Also, the inverse of $P = (x_1, y_1)$ is $-P = (x_1, -y_1)$. The only other axiom to check in order to verify that $+$ gives $E(\mathbb{Q})$ an abelian group structure is the associative law. This is simple but *very tedious* to check using only elementary methods[2]. Fortunately, we can coerce the computer algebra system MAGMA into verifying the associative law for us:

```
// The field K = Q(a,b,x0,x1,x2)
K<a,b,x0,x1,x2> := FieldOfFractions(PolynomialRing(Rationals(),5));
// The polynomial ring R = K[y0,y1,y2]
R<y0,y1,y2> := PolynomialRing(K,3);
// A maximal ideal of R.
I := ideal<R | y0^2 - (x0^3+a*x0+b), y1^2 - (x1^3+a*x1+b), y2^2-(x2^3+a*x2+b)>;
// The field L contains three distinct "generic" points on E.
L  := quo<R|I>;
E  := EllipticCurve([L| a,b]);  // The elliptic curve y^2 = x^3 + a*x + b.
P0 := E![L|x0,y0]; P1 := E![L|x1,y1]; P2 := E![L|x2,y2];
lhs := (P0 + P1) + P2;  rhs := P0 + (P1 + P2);
lhs eq rhs;
true     // yeah!
```

24.5 Mordell's Theorem

Theorem 24.5.1 (Mordell). *The group $E(\mathbb{Q})$ is finitely generated.*

This means that there are points $P_1, \ldots, P_r \in E(\mathbb{Q})$ such that every element of $E(\mathbb{Q})$ is of the form $n_1P_1 + \cdots + n_rP_r$ for some $n_1, \ldots n_r \in \mathbb{Z}$. I won't prove Mordell's theorem in this course. You can find an elementary proof of most of it in §1.3 of [Kato et al.].[3]

Example 24.5.2. Consider the elliptic curve E given by $y^2 = x^3+x+1$. Then $E(\mathbb{Q}) \approx \mathbb{Z}$ with generator $(0,1)$. We have $2(0,1) = (-1/4, -9/8)$, $3(0,1) = (72, 611)$, and $4(0,1) = \left(-\frac{287}{1296}, \frac{40879}{46656}\right)$.

[2]The right way to prove that the associate law holds is to develop the theory of algebraic curves and define the group law in terms of divisors; this is way outside the scope of this course.

[3]Matt Baker is teaching a graduate course (255r) this semester, and he is just about to present a proof of Weil's generalization of Mordell's theorem.

Chapter 25

The Elliptic Curve Group Law

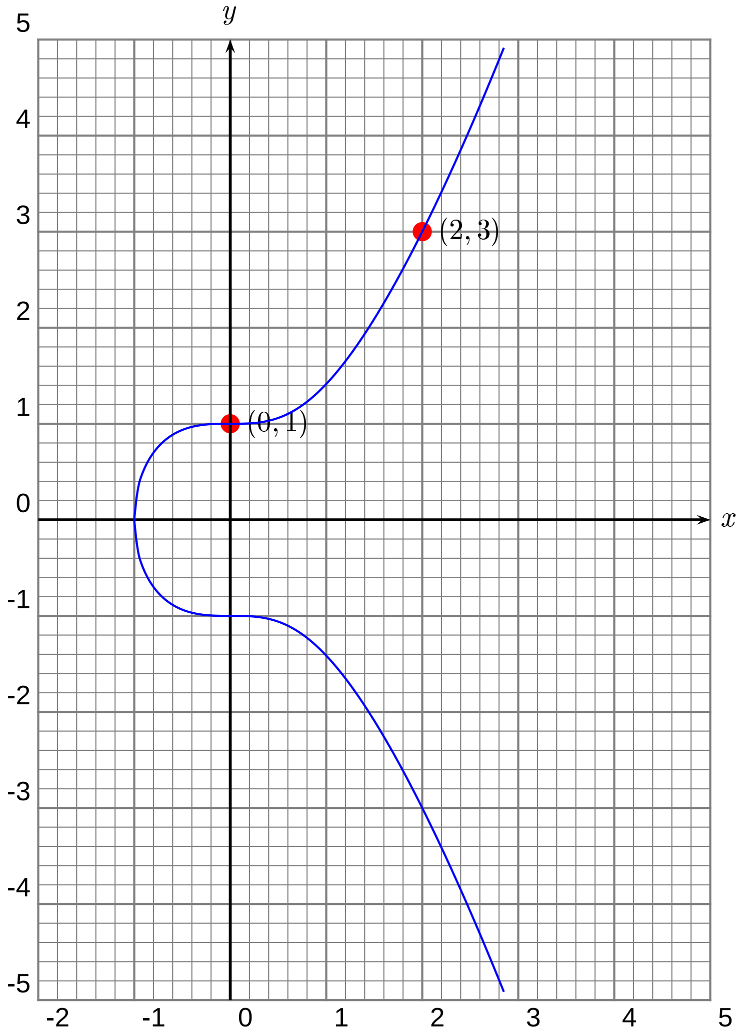
25.1 Some Graphs

Recall that an elliptic curve over a field K (in which 2 and 3 are invertible) can be defined by an equation

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

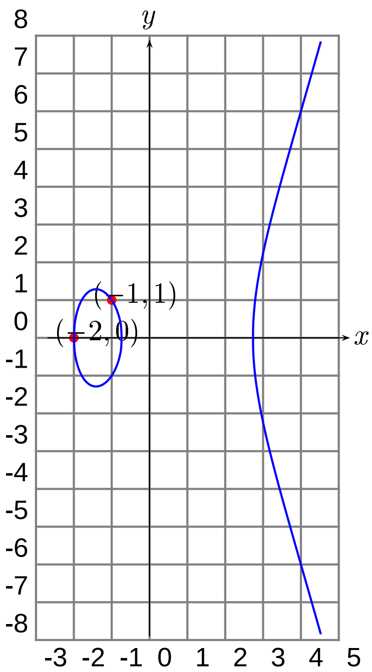
with $a, b \in K$. Here are some examples over $\mathbb{Q}$.

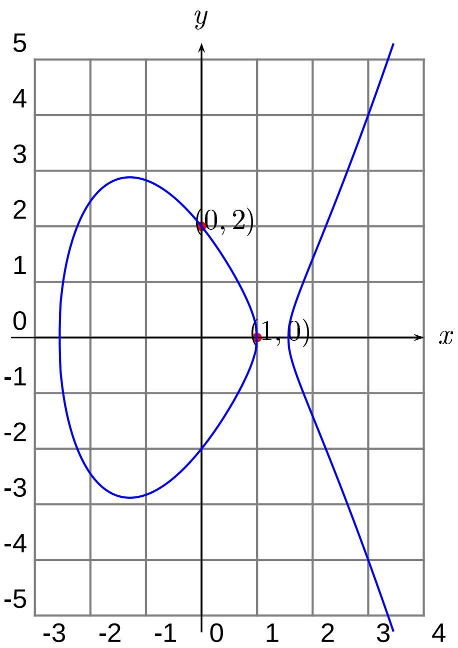
$$y^2 = x^3 + 1, \qquad E(\mathbb{Q}) \approx \mathbb{Z}/6\mathbb{Z}$$



$$y^2 = x^3 - 6x - 4 \qquad \text{and} \qquad y^2 = x^3 - 5x + 4$$

$$E(\mathbb{Q}) \approx (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z} \quad \text{(for both curves)}$$





(Exercise: Add the indicated points.)

25.2 The Point $\mathcal{O}$ at Infinity

The graphs of the previous section are each missing a point at infinity. They are graphs in the plane $\mathbb{R}^2$. The plane is a subset of the projective plane $\mathbb{P}^2$. The "closure" of the graph of $y^2 = x^3 + ax + b$ in $\mathbb{P}^2$ has exactly one extra point $\mathcal{O}$, which has rational coordinates, and which we sometimes call "the point at infinity".

Definition 25.2.1. The *projective plane* $\mathbb{P}^2$ is the set of triples (a, b, c), with a, b, c not all 0, modulo the equivalence relation

$$(a, b, c) \sim (\lambda a, \lambda b, \lambda c)$$

for any nonzero λ. We denote by $(a\!:\!b\!:\!c)$ the equivalence class of (a, b, c).

The "closure" in $\mathbb{P}^2$ of the graph of $y^2 = x^3 + ax + b$ is the graph of

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

and the extra point is $\mathcal{O} = (0\!:\!1\!:\!0)$. All finite points are of the form $(a\!:\!b\!:\!1)$.

For more about the projective plane, see page 28 of [Kato et al.].

25.3 The Group Law is a Group Law

Let E be an elliptic curve of the form $y^2 = x^3 + ax + b$ over a field K. Consider the set

$$E(K) = \{\mathcal{O}\} \cup \left\{(x, y) \in K \times K \,:\, y^2 = x^3 + ax + b\right\}.$$

Recall from the last lecture that there is a natural way to endow the set $E(K)$ with a *group* structure. Here's how it works. First, the element $\mathcal{O} \in E(K)$ is the zero element of the group. Next, suppose P and Q are elements of $E(K)$. Just like we did earlier, let $R = (x_3, y_3)$ be the third point of intersection of E and the line determined by P and Q (try this with the graphs on pages 1 and 2). Define

$$P + Q = (x_3, -y_3).$$

(For what goes wrong if you try to define $P + Q = (x_3, y_3)$, see your homework assignment.) There are various special cases to consider, such as when $P = Q$ or the third point of intersection is $\mathcal{O}$, but I will let you read about them in [Kato et al.].

It is not surprising that this binary operation on $E(K)$ satisfies $P + Q = Q + P$. Also, the inverse of $P = (x_1, y_1)$ is $-P = (x_1, -y_1)$. The only other axiom to check in order to verify that $+$ gives $E(K)$ an abelian group structure is the associative law. This is simple but *tedious* to check using only elementary methods. The right way to prove that the associate law holds is to develop the theory of algebraic curves and define the group law in terms of divisor classes, but this is outside the scope of this course. For fun, we can coerce the amazingly cool (but complicated) computer algebra system MAGMA into verifying the associative law (over $\mathbb{Q}$) for us:

```
// Define the field K = Q(a,b,x0,x1,x2)
K<a,b,x0,x1,x2> := FieldOfFractions(PolynomialRing(Rationals(),5));
// Define the polynomial ring R = K[y0,y1,y2]
R<y0,y1,y2> := PolynomialRing(K,3);
// Define a maximal ideal of R:
I := ideal<R | y0^2 - (x0^3+a*x0+b),
               y1^2 - (x1^3+a*x1+b),
               y2^2 - (x2^3+a*x2+b)>;
// The quotient L = R/I is a field that contains three
// distinct "generic" points on E.
L  := quo<R|I>;
// Define the elliptic curve y^2 = x^3 + a*x + b over L.
E  := EllipticCurve([L| a,b]);
// Let P0, P1, and P2 be three distinct "generic" points on E.
P0 := E![L|x0,y0]; P1 := E![L|x1,y1]; P2 := E![L|x2,y2];
// The algebraic formulas for the group law are built into MAGMA.
lhs := (P0 + P1) + P2;  rhs := P0 + (P1 + P2);
// Verify the associative law.
lhs eq rhs;
true     // Yeah, it works!
```

25.4 An Example Over a Finite Field

Let E be the elliptic curve $y^2 = x^3 + 3x + 3$ over the finite field

$$K = \mathbb{Z}/5\mathbb{Z} = \{0, 1, 2, 3, 4\}.$$

First, we find all points on E using PARI:

```
? for(x=0,4, for(y=0,4, if((y^2-(x^3+3*x+3))%5==0, print1([x,y],"  "))))
[3, 2]  [3, 3]  [4, 2]  [4, 3]
```

Thus $E(K) = \{\mathcal{O}, (3,2), (3,3), (4,2), (4,3)\}$, so $E(K)$ must be a cyclic abelian group of order 5. Let's verify that $E(K)$ is generated by $(3,2)$.

```
? e = ellinit([0,0,0,Mod(3,5),Mod(3,5)])
? ?ellpow    \\ type ?5 for a complete list of elliptic-curve functions
ellpow(e,x,n): n times the point x on elliptic curve e (n in Z).
? x = [3,2];
? for(n=1,5,print(n,"*[3,2] =  ",lift(ellpow(e,x,n))))
1*[3,2] =  [3, 2]
2*[3,2] =  [4, 3]
3*[3,2] =  [4, 2]
4*[3,2] =  [3, 3]
5*[3,2] =  [0]
```

25.5 Mordell's Theorem

Venerable Problem: *Find an algorithm that, given an elliptic curve E over $\mathbb{Q}$, outputs a complete description of the set of rational points (x_0, y_0) on E.*

This problem is difficult. In fact, so far it has stumped everyone! There is a *conjectural algorithm*, but nobody has succeeded in proving that it is really an algorithm, in the sense that it terminates for any input curve E. Several of your profs at Harvard, including Barry Mazur, myself, and Christophe Cornut (who will teach Math 129 next semester) have spent, or might spend, a huge chunk of their life thinking about this problem.

How could one possible "describe" the group $E(\mathbb{Q})$, since it can be infinite? In 1923, Mordell proved that there is always a reasonable way to describe $E(\mathbb{Q})$.

Theorem 25.5.1 (Mordell). *The group $E(\mathbb{Q})$ is finitely generated.*

This means that there are points $P_1, \ldots, P_s \in E(\mathbb{Q})$ such that every element of $E(\mathbb{Q})$ is of the form $n_1P_1 + \cdots + n_sP_s$ for some $n_1, \ldots n_s \in \mathbb{Z}$. I will not prove Mordell's theorem in this course, but see §1.3 of [Kato et al.].

Example 25.5.2. Consider the elliptic curve E given by $y^2 = x^3 - 6x - 4$. Then $E(\mathbb{Q}) \approx (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}$ with generators $(-2,0)$ and $(-1,1)$. We have

$$5(-1,1) = \left(-\frac{131432401}{121462441}, -\frac{1481891884199}{1338637562261}\right).$$

Trying finding that point without knowing about the group law!

$(0,1) + (0,1) = (-1/4, -9/8)$, $(0,1) + (0,1) + (0,1) = (72, 611)$, and $(0,1) + (0,1) + (0,1) + (0,1) = \left(-\frac{287}{1296}, \frac{40879}{46656}\right)$.

Chapter 26

Torsion Points on Elliptic Curves and Mazur's Big Theorem

26.1 Mordell's Theorem

Venerable Problem: *Find an algorithm that, given an elliptic curve E over $\mathbb{Q}$, outputs a complete description of the set of rational points (x_0, y_0) on E.*

This problem is difficult. In fact, so far it has stumped everyone! There is a *conjectural algorithm*, but nobody has succeeded in proving that it is really an algorithm, in the sense that it terminates for any input curve E. Several of your profs at Harvard, including Barry Mazur, myself, and Christophe Cornut (who will teach Math 129 next semester) have spent, or might spend, a huge chunk of their life thinking about variants of this problem.

How could one possible "describe" the group $E(\mathbb{Q})$, since it can be infinite? In 1923, Mordell proved that there is always a reasonable way to describe $E(\mathbb{Q})$.

Theorem 26.1.1 (Mordell). *The group $E(\mathbb{Q})$ is finitely generated.*

This means that there are points $P_1, \ldots, P_s \in E(\mathbb{Q})$ such that every element of $E(\mathbb{Q})$ is of the form $n_1P_1 + \cdots + n_sP_s$ for some $n_1, \ldots n_s \in \mathbb{Z}$. I will not prove Mordell's theorem in this course. See §1.3 of [Kato et al.] for a proof in the special case when E is given by an equation of the form $y^2 = (x-a)(x-b)(x-c)$.

Example 26.1.2. Consider the elliptic curve E given by $y^2 = x^3 - 6x - 4$. Then $E(\mathbb{Q}) \approx (\mathbb{Z}/2\mathbb{Z}) \times \mathbb{Z}$ with generators $(-2, 0)$ and $(-1, 1)$. We have

$$5(-1,1) = \left(-\frac{131432401}{121462441}, -\frac{1481891884199}{1338637562261}\right).$$

Trying finding that point without knowing about the group law!

26.2 Exploring the Possibilities

As E varies over all elliptic curves over $\mathbb{Q}$, what are the possibilities for $E(\mathbb{Q})$? What finitely generated abelian groups occur? Mordell's theorem implies that

$$E(\mathbb{Q}) \approx \mathbb{Z}^r \oplus E(\mathbb{Q})_{\text{tor}},$$

where $E(\mathbb{Q})_{\text{tor}}$ is the set of points of finite order in $E(\mathbb{Q})$ and $\mathbb{Z}^r \approx E(\mathbb{Q})/E(\mathbb{Q})_{\text{tor}}$. The number r is called the *rank* of E.

26.2.1 The Torsion Subgroup

Theorem 26.2.1 (Mazur, April 16, 1976). *Let E be an elliptic curve over $\mathbb{Q}$. Then $E(\mathbb{Q})_{\text{tor}}$ is isomorphic to one of the following 15 groups:*

$$\begin{array}{rl} \mathbb{Z}/n\mathbb{Z} & \textit{for } n \leq 10 \textit{ or } n = 12, \\ (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2n\mathbb{Z}) & \textit{for } n \leq 4. \end{array}$$

As we will see in the next section, all of these torsion subgroups really do occur. Mazur's theorem is very deep, and I can barely begin to hint at how he proved it. The basic idea is to define, for each positive integer N, a curve $Y_1(N)$ with the magnificient property that the points of $Y_1(N)$ with complex coordinates are in natural bijection with the (isomorphism classes of) pairs (E, P), where E is an elliptic curve and P is a point of E of order N. Moreover, $Y_1(N)$ is amazing in that it has a rational point if and only if there is an elliptic curve over $\mathbb{Q}$ with a rational point of order N. I won't define $Y_1(N)$, but here it is for the first few N:

N	A curve that contains $Y_1(N)$
$1-10$, 12	a straight line; these have lots of points!
11	$y^2 + y = x^3 - x^2$
13	$y^2 = x^6 + 2x^5 + x^4 + 2x^3 + 6x^2 + 4x + 1$
14	$y^2 + xy + y = x^3 - x$
15	$y^2 + xy + y = x^3 + x^2$
16	$y^2 = (x-1)(x+1)(x^2 - 2x - 1)(x^2 + 1)$
17	The intersection of the hypersurfaces in $\mathbb{P}^4$ defined by: $ac - b^2 + 5bd - 3be - c^2 - 4cd + 2ce - 4d^2 + 7de - 2e^2$, $ad - bc + bd - be + c^2 - 2cd - 2d^2 + 4de - e^2$, and $ae - be - cd + 2d^2 - 2de + e^2$.
18	$y^2 = x^6 + 4x^5 + 10x^4 + 10x^3 + 5x^2 + 2x + 1$

(Some of the curves in the right hand column have a few obvious rational points, but these points "don't count".)

Mazur proved that if $N = 11$ or $N \geq 13$, then $Y_1(N)$ has no rational points. This result, together with the theory surrounding $Y_1(N)$, yields his theorem.

26.2.2 The Rank

Conjecture 26.2.2. *There exist elliptic curves over $\mathbb{Q}$ of arbitrarily large rank.*

As far as I know, nobody has any real clue as to how to prove Conjecture 26.2.2 (Doug Ulmer recently wrote a paper which gives theoretical evidence). The current "world record" is a curve of rank ≥ 24. It was discovered in January 2000 by Roland Martin and William McMillen of the **National Security Agency**. For security reasons, I won't tell you anything about how they found it.

Theorem 26.2.3. *The elliptic curve*

$$y^2+xy+y = x^3-120039822036992245303534619191166796374x+504224992484910670010801799168082726759443756222911415116$$

over $\mathbb{Q}$ has rank at least 24. The following points $P_1, \ldots, P_{24}$ are independent points on the curve:

$$\begin{aligned}
P_1 &= (2005024558054813068, -16480371588343085108234888252)\\
P_2 &= (-4690836759490453344, -31049883525785801514744524804)\\
P_3 &= (4700156326649806635, -6622116250158424945781859743)\\
P_4 &= (6785546256295273860, -14561809288309785211075204 73)\\
P_5 &= (6823803569166584943, -16859507354771759473517748 17)\\
P_6 &= (7788809602110240789, -6462981622972389783453855713)\\
P_7 &= (27385442304350994620556, 4531892554281655472841805111276996)\\
P_8 &= (54284682060285253719/4, -296608788157989016192182090427/8)\\
P_9 &= (-94200235260395075139/25, -3756324603619419619213452459781/125)\\
P_{10} &= (-3463661055331841724647/576, -439033541391867690041114047287793/13824)\\
P_{11} &= (-6684065934033506970637/676, -473072253066190669804172657192457/17576)\\
P_{12} &= (-956077386192640344198/2209, -2448326762443096987265907469107661/103823)\\
P_{13} &= (-27067471797013364392578/2809, -4120976168445115434193886851218259/148877)\\
P_{14} &= (-25538866857137199063309/3721, -7194962289937471269967128729589169/226981)\\
P_{15} &= (-102632501176025905189433 1/108241, -10008952940674898577361109630032677 73/35611289)\\
P_{16} &= (9351361230729481250627334/1366561, -28697496057486357774753723393062048 32/1597509809)\\
P_{17} &= (10100878635879432897339615/1423249, -53049657762769664510669009414893878 01/1697936057)\\
P_{18} &= (11499655868211022625340735/17522596, -15134357633415411882652302414268264 78043/73349586856)\\
P_{19} &= (110352253665081002517811734/21353641, -46170683330840667140557025454264778 4288/98675175061)\\
P_{20} &= (414280096426033094143668538257/285204544, 266642138924791310663963499787603019833872421/4816534339072)\\
P_{21} &= (36101712290699828042930087436/4098432361, -29952588557667645204633891535871116 70142292/262377541318859)\\
P_{22} &= (45442463408503524215460183165/5424617104, -37160415814701441087215906955546701 56388869/399533898943808)\\
P_{23} &= (983886013344700707678587482584/141566320009, -12661581838771793044916162596039760 5741940953/53264752602346277)\\
P_{24} &= (1124614335716851053281176544216033/152487126016, -37714203831317877163580088877209977295481388540127/59545612760743936)
\end{aligned}$$

Proof. See

`http://listserv.nodak.edu/scripts/wa.exe?A2=ind0005&L=nmbrthry&P=R182`

$\square$

26.3 How to Compute $E(\mathbb{Q})_{\mathrm{tor}}$

The following theorem yields an algorithm to compute $E(\mathbb{Q})_{\mathrm{tor}}$.

Theorem 26.3.1 (Nagell-Lutz). *Suppose that $y^2 = x^3 + ax + b$ (with $a, b \in \mathbb{Z}$) defines an elliptic curve E over $\mathbb{Q}$, let $\Delta = -16(4a^3 + 27b^2)$ be the discriminant, and suppose that $P = (x, y) \in E(\mathbb{Q})_{\mathrm{tor}}$. Then x and y are integers and either $y = 0$, in which case P has order 2, or $y^2 \mid \Delta$.*

Non-proof. I will not prove this theorem. However, you can find a readable proof in Chapter II of Silverman and Tate's *Rational Points on Elliptic Curves.* $\square$

Warning: Nagell-Lutz is NOT an if and only if statement. There are points of infinite order that satisfy the conclusion of Theorem 26.3.1. For example, the point $(1,3)$ on $y^2 = x^3 + 8$ has integer coordinates and $y^2 = 9 \mid \Delta = -16 \cdot 27 \cdot 3^2$. However,

$$(1,3) + (1,3) = \left(-\frac{7}{4}, -\frac{13}{8}\right).$$

Since the coordinates of $(1,3) + (1,3)$ are not integers, it follows from the contrapositive (not converse!) of Nagell-Lutz that $(1,3)$ must be a point of infinite order.

Example 26.3.2. The following is a list of elliptic curves with each possible torsion subgroup. Tom Womack (a graduate student in Nottingham, where Robin Hood lives) has a web page, `http://www.tom.womack.net/maths/torsion.htm`, which contains PARI code that lists infinitely many elliptic curve with each torsion subgroup.

Curve	$E(\mathbb{Q})_{\rm tor}$
$y^2 = x^3 - 2$	$\{0\}$
$y^2 = x^3 + 8$	$\mathbb{Z}/2\mathbb{Z}$
$y^2 = x^3 + 4$	$\mathbb{Z}/3\mathbb{Z}$
$y^2 = x^3 + 4x$	$\mathbb{Z}/4\mathbb{Z}$
$y^2 - y = x^3 - x^2$	$\mathbb{Z}/5\mathbb{Z}$
$y^2 = x^3 + 1$	$\mathbb{Z}/6\mathbb{Z}$
$y^2 = x^3 - 43x + 166$	$\mathbb{Z}/7\mathbb{Z}$
$y^2 + 7xy = x^3 + 16x$	$\mathbb{Z}/8\mathbb{Z}$
$y^2 + xy + y = x^3 - x^2 - 14x + 29$	$\mathbb{Z}/9\mathbb{Z}$
$y^2 + xy = x^3 - 45x + 81$	$\mathbb{Z}/10\mathbb{Z}$
$y^2 + 43xy - 210y = x^3 - 210x^2$	$\mathbb{Z}/12\mathbb{Z}$
$y^2 = x^3 - 4x$	$(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$
$y^2 = x^3 + 2x^2 - 3x$	$(\mathbb{Z}/4\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$
$y^2 + 5xy - 6y = x^3 - 3x^2$	$(\mathbb{Z}/6\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$
$y^2 + 17xy - 120y = x^3 - 60x^2$	$(\mathbb{Z}/8\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$

The `elltors` function in PARI computes torsion subgroups:

```
? ?elltors
elltors(e,{flag=0}): torsion subgroup of elliptic curve e: order, structure,
generators. If flag = 0, use Doud's algorithm; if flag = 1, use Lutz-Nagell.
? e=ellinit([17,-60,-120,0,0]);
? elltors(e)
%4 = [16, [8, 2], [[30, -90], [-40, 400]]]
? e.disc
%5 = 51438240000
? e.disc % 90^2        \\ verify Nagell-Lutz
%6 = 0
? e.disc % 400^2       \\ verify Nagell-Lutz
%7 = 0
```

Chapter 27

Computing with Elliptic Curves

(in PARI)

27.1 Initializing Elliptic Curves

We are concerned primarily with elliptic curves E given by an equation of the form

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

with a and b either rational numbers or elements of a finite field $\mathbb{Z}/p\mathbb{Z}$. If a and b are in $\mathbb{Q}$, we initialize E in PARI using the following command:

```
? E = ellinit([0,0,0,a,b]);
```

If you wish to view a and b as element of $\mathbb{Z}/p\mathbb{Z}$, initialize E as follows:

```
? E = ellinit([0,0,0,a,b]*Mod(1,p));
```

If $\Delta = -16(4a^3 + 27b^2) = 0$ then `ellinit` will complain; otherwise, `ellinit` returns a 19-component vector of information about E. You can access some of this information using the dot notation, as shown below.

```
? E = ellinit([0,0,0,1,1]);
? E.a4
%11 = 1
? E.a6
%12 = 1
? E.disc
%13 = -496
? E.j
%14 = 6912/31
? E5 = ellinit([0,0,0,1,1]*Mod(1,5));
? E5.disc
%15 = Mod(4, 5)
? E5.j
%16 = Mod(2, 5)
```

Here `E.j` is the *j-invariant* of E. It is equal to $\frac{2^8 3^3 a^3}{4a^3+27b^2}$, and has some remarkable properties that I probably won't tell you about.

Most elliptic curves functions in PARI take as their first argument the output of `ellinit`. For example, the function `ellisoncurve(E,P)` takes the output of `ellinit` as its first argument and a point `P=[x,y]`, and returns `1` if `P` lies on `E` and `0` otherwise.

```
? P = [0,1]
? ellisoncurve(E, P)
%17 = 1
? P5 = [0,1]*Mod(1,5)
? ellisoncurve(E5, P)
%18 = 1
```

27.2 Computing in The Group

The following functions implement some basic arithmetic in the group of points on an elliptic curve: `elladd`, `ellpow`, and `ellorder`. The `elladd` function simply adds together two points using the group law. Warning: PARI does *not* check that the two points are on the curve.

```
? P = [0,1]
%2 = [0, 1]
? elladd(E,P,P)
%3 = [1/4, -9/8]
? elladd(E,P,[1,0])    \\ nonsense, since [1,0] isn't even on E!!!
%4 = [0, -1]
? elladd(E5,P5,P5)
%12 = [Mod(4, 5), Mod(2, 5)]
? [1/4,-9/8]*Mod(1,5)
%13 = [Mod(4, 5), Mod(2, 5)]
```

The `ellpow` function computes $nP = P + P + \cdots + P$ (n summands).

```
? ellpow(E,P,2)
%5 = [1/4, -9/8]
? ellpow(E,P,3)
%6 = [72, 611]
? ellpow(E,P,15)
```

```
%7 = [26449452347718826171173662182327682047670541792/9466094804586385762312509661837302961354550401,
4660645813671121765025590267647300672252945873586541077711389394563791/9209928837349924627451415221112259088619760982194656165856492453956 49]
```

27.3 The Generating Function $L(E, s)$

Suppose E is an elliptic curve over $\mathbb{Q}$ defined by an equation $y^2 = x^3 + ax + b$. Then for every prime p that does not divide $\Delta = -16(4a^3 + 27b^2)$, the same equation defines an elliptic curve over the finite field $\mathbb{Z}/p\mathbb{Z}$. As you will discover in problem

3 of homework 9, it can be exciting to consider the package of numbers $\#E(\mathbb{Z}/p\mathbb{Z})$ of points on E over all finite fields. The function `ellap` computes

$$a_p(E) = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}).$$

```
? E = ellinit([0,0,0,1,1]);
? ellap(E,5)
%19 = -3          \\ this should be 5+1 - #points
? E5 = ellinit([0,0,0,1,1]*Mod(1,5));
? for(x=0,4, for(y=0,4, if(ellisoncurve(E5,[x,y]),print([x,y]))))
[0, 1]
[0, 4]
[2, 1]
[2, 4]
[3, 1]
[3, 4]
[4, 2]
[4, 3]
? 5+1 - 9          \\ 8 points above, plus the point at infinity
%22 = -3
```

There is a natural way to extend the definition of a_p to define integers a_n for every integer n. For example, if a_p and a_q are defined as above and p and q are distinct primes, then $a_{pq} = a_p a_q$. Today I won't tell you how to define the a_p when, e.g., $p \mid \Delta$. However, you can compute the numbers a_n quickly in PARI using the function `ellan`, which computes the first few a_n.

```
? ellan(E,15)
%24 = [1, 0, 0, 0, -3, 0, 3, 0, -3, 0, -2, 0, -4, 0, 0]
```

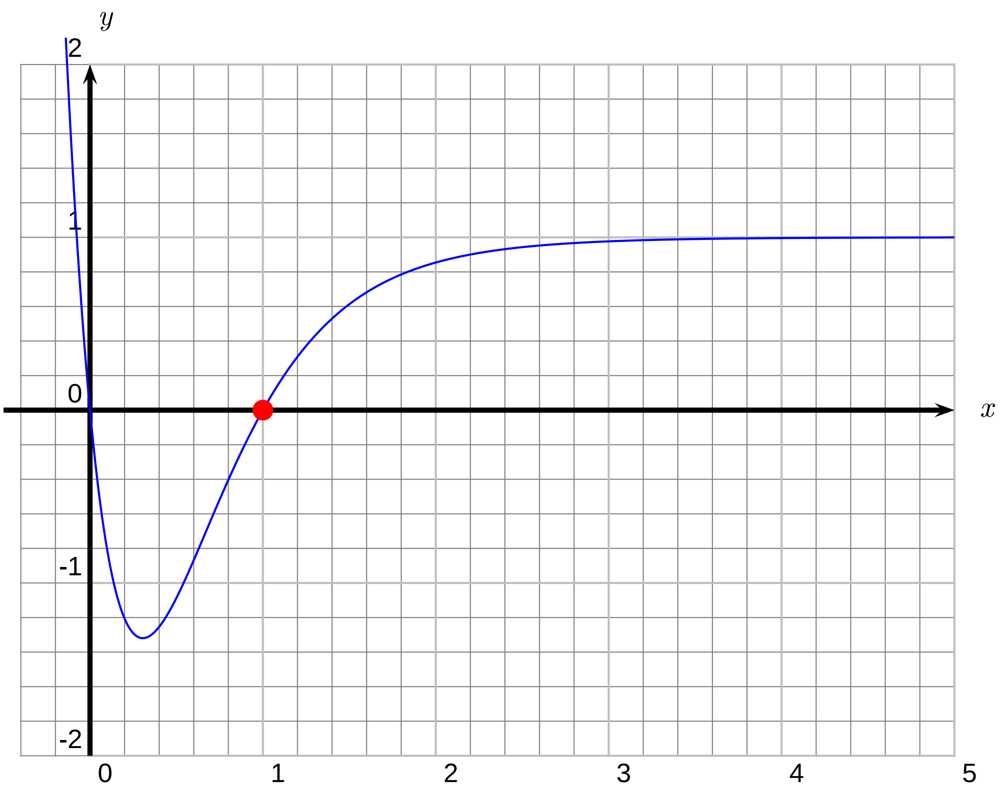
This output means that $a_1 = 1$, $a_2 = a_3 = a_4 = 0$, $a_5 = -3$, $a_6 = 0$, and so on.

When confronted by a mysterious list of numbers, it is a "reflex action" for a mathematician to package them together in a generating function, and see if anything neat happens. It turns out that for the above numbers, a good way to do this is as follows. Define

$$L(E, s) = \sum_{n=1}^{\infty} a_n n^{-s}.$$

This might remind you of Riemann's ζ-function, which is the function you get if you make the simplest generating function $\sum_{n=1}^{\infty} n^{-s}$ of this form.

Using `elllseries(E,s,1)` I drew a graph of $L(E, s)$ for $y^2 = x^3 + x + 1$.



That the value of $L(E, s)$ makes sense at $s = 1$, where the series above doesn't obviously converge, follows from the nontrivial fact that the function

$$f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$$

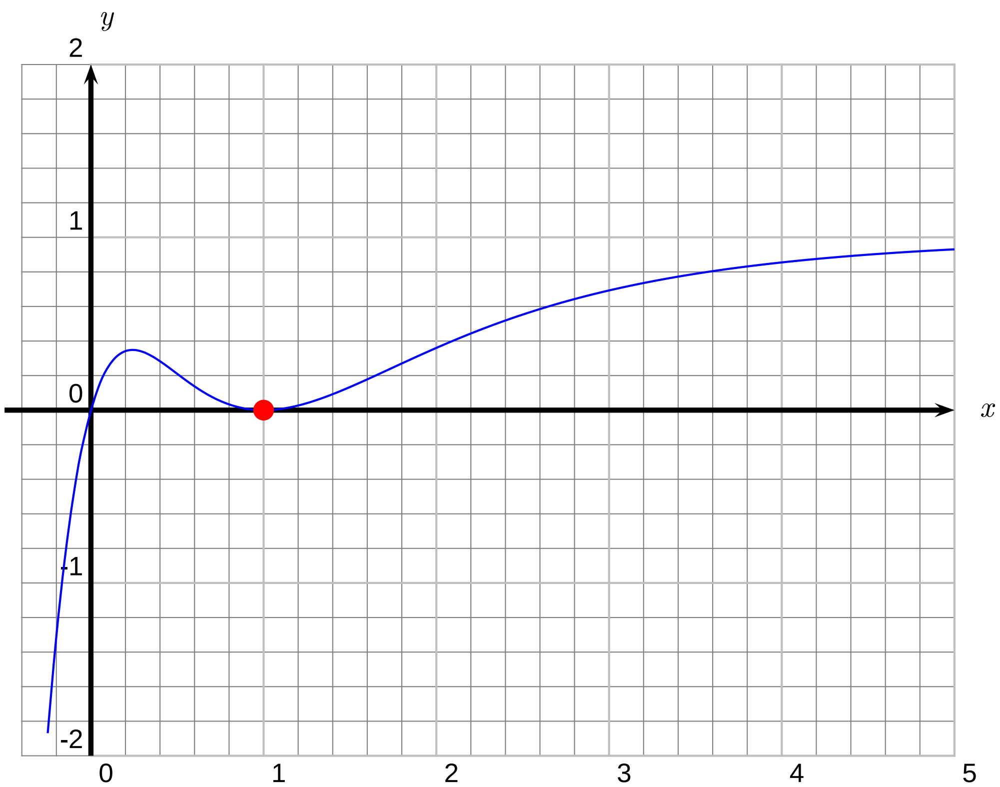
is a *modular form*. Also, keep your eyes on the dot; it plays a central roll in the Birch and Swinnerton-Dyer conjecture, which asserts that $L(E, 1) = 0$ if and only if the group $E(\mathbb{Q})$ is infinite.

27.3.1 A Curve of Rank Two

Let E be the simplest rank 2 curve:

$$y^2 + y = x^3 + x^2 - 2x.$$

The discriminant is 389.

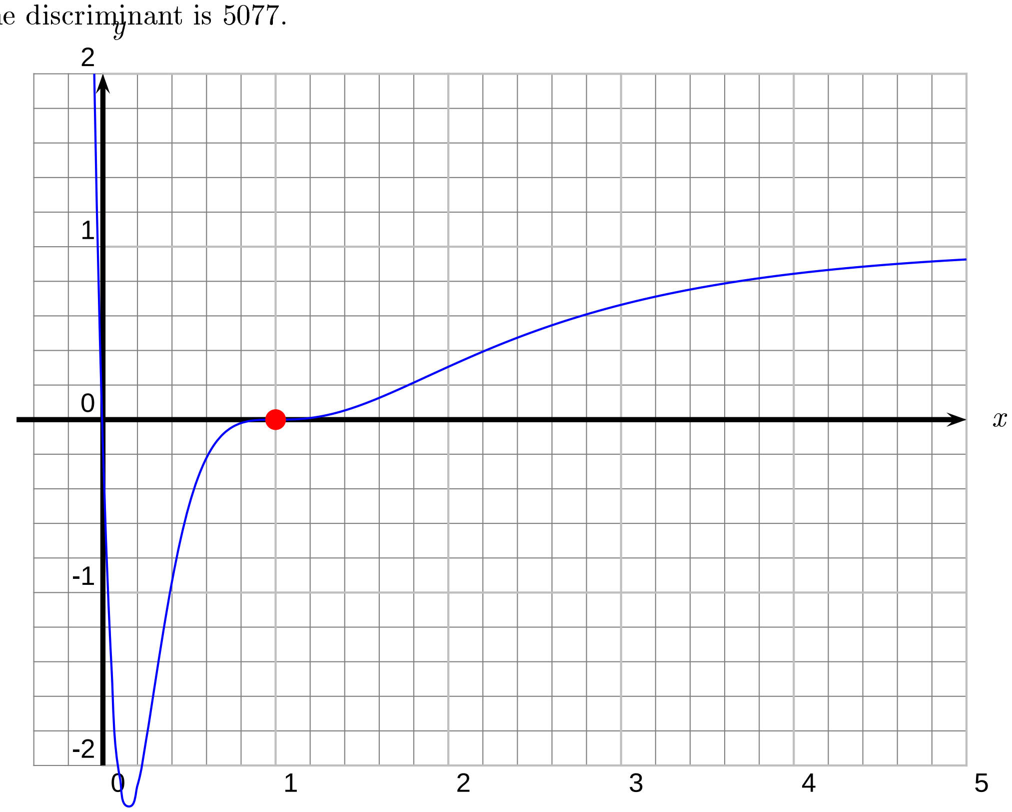


27.3.2 A Curve of Rank Three

Let E be the simplest rank 3 curve:

$$y^2 + y = x^3 - 7x + 6.$$

The discriminant is 5077.

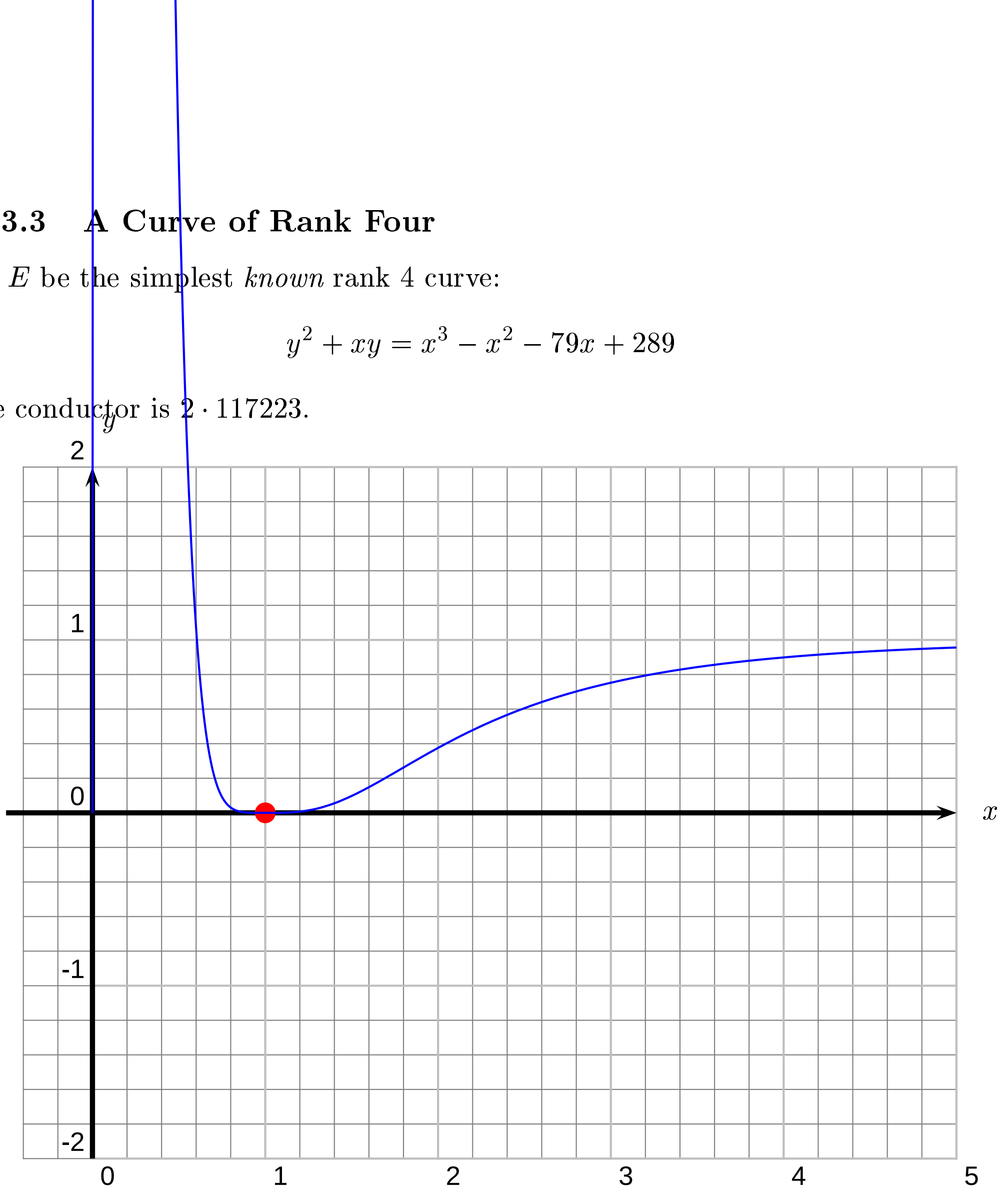


27.3.3 A Curve of Rank Four

Let E be the simplest *known* rank 4 curve:

$$y^2 + xy = x^3 - x^2 - 79x + 289$$

The conductor is $2 \cdot 117223$.



27.4 Other Functions and Programs

You can see a complete list of elliptic-curves functions by typing `?5`:

```
? ?5
elladd          ellak           ellan           ellap
ellbil          ellchangecurve  ellchangepoint  elleisnum
elleta          ellglobalred    ellheight       ellheightmatrix
ellinit         ellisoncurve    ellj            elllocalred
elllseries      ellorder        ellordinate     ellpointtoz
ellpow          ellrootno       ellsigma        ellsub
elltaniyama     elltors         ellwp           ellzeta         ellztopoint
```

I have only described a small subset of these. To understand many of them, you must first learn how to view an elliptic curve as a "donut", that is, as quotient of the complex numbers by a *lattice*, and also as a quotient of the upper half plane.

There is a Maple package called APECS for computing with elliptic curves, which is more sophisticated than PARI in certain ways, especially in connection with algorithms that involve lots of commutative algebra. MAGMA also offers sophisticated features for computing with elliptic curves, which are built in to the standard distribution. I will give a demonstrations of MAGMA in the Basic Notions seminar at 3pm on Monday, December 3 in SC 507. There is also a C++ library called LiDIA that has libraries with some powerful elliptic curves features.

Chapter 28

Elliptic Curve Cryptography

Today's lecture is about an application of elliptic curves to cryptography.

Disclaimer: I do not endorse breaking laws, and give the examples below as a pedagogical tool in the hope of making the mathematics in our course more fun and relevant to everyday life. I don't think I have violated the Digital Millenium Copyright Act, because I have given very few details about Microsoft's actual protocols, and I've given absolutely no source code.

28.1 Microsoft Digital Rights Management

Today I will describe one way to use elliptic curves in cryptography. Our central example will involve version 2 of the Microsoft Digital Rights Management (MS-DRM) system, as applied to `.wma` audio files.

I learned about this protocol from a paper by "Beale Screamer".

28.1.1 Microsoft's Favorite Elliptic Curve

The elliptic curve used in MS-DRM is an elliptic curve over the finite field $k = \mathbb{Z}/p\mathbb{Z}$, where

$$p = 785963102379428822376694789446897396207498568951.$$

As Beale Screamer remarks, this modulus has high nerd appeal because in hexadecimal it is

$$\text{89ABCDEF012345672718281831415926141424F7},$$

which includes counting in hexadecimal, and digits of e, π, and $\sqrt{2}$. The Microsoft elliptic curve E is

$$\begin{aligned} y^2 = x^3 &+ 317689081251325503476317476413827693272746955927x \\ &+ 79052896607878758718120572025718535432100651934. \end{aligned}$$

We have

$$\#E(k) = 785963102379428822376693024881714957612686157429,$$

and the group $E(k)$ is cyclic with generator

$$B = (771507216262649826170648268565579889907769254176,\\ 390157510246556628525279459266514995562533196655).$$

28.1.2 Nikita and Michael

Our heros Nikita and Michael love to share digital music  when they aren't out thwarting terrorists. When Nikita installed Microsoft's content rights management software on her compute, it sneakily generated a private key

$$n = 670805031139910513517527207693060456300217054473,$$

which it very stealthily hid in bits and pieces of files (e.g., `blackbox.dll`, `v2ks.bla`, and `IndivBox.key`). In order for Nikita to play Juno Reactor's latest hit `juno.wma`, her web browser contacts a Microsoft rights management partner. After Nikita gives Microsoft her credit card number, she is allowed to download a license to play `juno.wma`. Microsoft created the license using the ElGamal public-key cryptosystem (see below) in the group $E(k)$. Nikita's license file can now be used to unlock `juno.wma`, but *only* on Nikita's computer. When she shares both `juno.wma` and the license file with Michael, he is very annoyed because he can't play `juno.wma`. This is because Michael's computer doesn't know Nikita's computer's private key (that integer n above), so Michael's computer can't decrypt the license file.

`juno.wma`

28.2 The Elliptic Curve Discrete Logarithm Problem

Definition 28.2.1. If E is an elliptic curve over $\mathbb{Z}/p\mathbb{Z}$ and B is a point on E, then the *discrete log problem* on E to the base B is the following problem: given a point $P \in E$, find an integer n such that $nB = P$, if such an integer exists.

For example, let E be the elliptic curve given by $y^2 = x^3 + x + 1$ over the field $\mathbb{Z}/7\mathbb{Z}$. We have

$$E(\mathbb{Z}/7\mathbb{Z}) = \{\mathcal{O}, (2,2), (0,1), (0,6), (2,5)\}.$$

If $B = (2,2)$ and $P = (0,6)$, then $3B = P$, so $n = 3$ is a solution to the discrete logarithm problem.

To the best of my knowledge, the discrete logarithm problem on E is *really hard* unless $\#E(\mathbb{Z}/p\mathbb{Z})$ is "smooth", i.e., a product of small primes, or E is "supersingular" in the sense that $p \mid \#E(\mathbb{Z}/p\mathbb{Z})$. The Microsoft curve has neither of these deficiencies, and I expect that the discrete logarithm on that curve is quite difficult. This is not the weekness that "Beale Screamer" exploits in breaking MS-DRM.

28.3 ElGamal

How can we set up a public-key cryptosystem using an elliptic curve? The only public-key crytosystem that we've studied so far is the RSA cryptosystem; unfortunately, there is no analogue of RSA for elliptic curves! (Informal Exercise: Think about what goes wrong.)

MS-DRM uses the El Gamal system. Here's how it works. Start with a fixed, publicly known prime p, an elliptic curve E over $\mathbb{Z}/p\mathbb{Z}$, and a point $B \in E(\mathbb{Z}/p\mathbb{Z})$. Michael and Nikita choose random integers m and n, which are kept secret, and compute and publish mB and nB.

In order to send a message P to Michael, Nikita computes a random integer r and sends the pair of points $(rB, P + r(mB))$. To read the message, Michael multiplies rB by his secret key m to get $m(rB) = r(mB)$, and subtracts this from the second point to get

$$P = P + r(mB) - r(mB).$$

As far as I can tell, breaking this cryptosystem requires solving the discrete logarithm problem, so it's very difficult.

The following example is based on an example taken from Beale Screamer's paper.

Example 28.3.1. Nikita's license files contains the pair of points $(rB, P + r(nB))$, where

$$rB = (179671003218315746385026655733086044982194424660,\ 697834385359686368249301282675141830935176314718)$$

and

$$P+r(nB) = (137851038548264467372645158093004000343639118915,\ 1108485892286762240572292302235808150242248756 99).$$

Nikita's computer sneakily loads the secret key

$$n = 670805031139910513517527207693060456300217054473$$

into memory and computes

$$n(rB) = r(nB) = (328901393518732637577115650601768681044040715701,\ 586947838087815993601350565488788846203887988162).$$

It then subtracts this from $P + r(nB)$ to get

$$P = (14489646124220757767,\ 669337780373284096274895136618194604469696830074).$$

That x coordinate, 14489646124220757767, is the top secret magic "content key" that unlocks `juno.wma`.

If Nikita knew the private key n that her computer generated, she could compute P herself and unlock `juno.wma` and share her music with Michael, just like she used to share her favorite CDs with Michael. Beale Screamer found a weakness in Microsoft's system that let him find n:

> "These secret keys are stored in linked lists ... interspersed with the code in the library. The idea is that they can be read by that library, used internally by that library, and never communicated outside the library. Since the `IndivBox.key` file is shuffled in a random way for each client, these keys would be extremely difficult to extract from the file itself.

Fortunately, we don't have to: these keys are part of the object state that is maintained by this library, and since the offset within this object of these secret keys is known, we can let the library itself extract the secret keys! The code for this simply loads up the 'black box' library, has it initialize an instance of the object, and then reads the keys right out of that object. This is clearly a weakness in the code which can be corrected by the DRM software fairly easily, but for now it is the basis of our exploit."

As you can see, Microsoft has undertaken a difficult and interesting problem. *How can Microsoft store data on Nikita's computer in such a way that Nikita can not access it, but Nikita's computer can?*

28.4 Why Use Elliptic Curves?

There are several advantages to using elliptic curves instead of $\mathbb{Z}/p\mathbb{Z}$ for cryptography, though the people at RSA Corporation might disagree. Elliptic curve cryptosystems with smaller key sizes appear to be just as secure as "classical" $\mathbb{Z}/p\mathbb{Z}$ cryptosystems with much larger key sizes, so elliptic curve cryptosystems can be more efficient. Another advantage, which I won't explain at all, is that elliptic curve cryptosystems appear to be vastly more secure over "large finite fields of characteristic 2" than RSA, which is is very important in practical applications. Also, elliptic curves are simply way cooler than $\mathbb{Z}/p\mathbb{Z}$, so they (used to) attract venture capitalists.

Some mobile phones also use elliptic curve cryptography. Do you have an elliptic curve in your pocket right now?

```
/* Base conversion */

function ToDeci(s)
   c := ["0","1","2","3","4","5","6","7","8","9","a","b","c","d","e","f"];
   ans := 0;
   b := 1;
   for i in [1..#s] do
      ans := ans + b*(Index(c,s[#s-i+1])-1);
      b := b*16;
   end for;
   return ans;
end function;


p := 785963102379428822376694789446897396207498568951
k := GF(p);
E := EllipticCurve([317689081251325503476317476413827693272746955927,  79052896607878758
B  := E![k|771507216262649826170648268565579889907769254176, 39015751024655662852527945
rB := E![k|1796710032183157463850266557330860449821944246 60,6978343853596863682493012826
PrnB := E![k|137851038548264467372645158093004000343639118915,1108485892286762240572292 3
n := 670805031139910513517527207693060456300217054473;
```

```
PrnB - n*rB;
(14489646124220757767 : 66933778037328409627489513661819460446969683 0074 : 1)
```

Chapter 29

Using Elliptic Curves to Factor, Part I

In 1987, Hendrik Lenstra published the landmark paper *Factoring Integers with Elliptic Curves*, Annals of Mathematics, **126**, 649–673, which you can download from the Math 124 web page. Lenstra's method is also described in §IV.4 of Silverman and Tate's *Rational Points on Elliptic Curves*, §VIII.5 of [Davenport], and in §10.3 of Cohen's *A Course in Computational Algebraic Number Theory.*

In this lecture and the next, I will tell you about Lenstra's clever algorithm. It shines at finding "medium sized" factors of an integer N, which these days means 10 to 20 decimal digits but probaby not 30 decimal digits. The ECM method is thus not useful for earning money by factoring RSA challenge numbers, but is essential when factoring most integers. It also has small storage requirements. Lenstra writes:

> *"It turns out that ... the elliptic curve method is one of the fastest integer factorization methods that is currently used in practice. The quadratic sieve algorithm still seems to perform better on integers that are built up from two prime numbers of the same order of magnitude; such integers are of interest in cryptography."*

Lenstra's discover of the elliptic curve method was inspired by Pollard's $(p-1)$-method. I will spend most of the rest of this lecture introducing you to it.

29.1 Power-Smoothness

Definition 29.1.1 (Power-smooth). Let B be a positive integer. A positive integer n is *B-power-smooth* if all prime powers dividing n are less than or equal to B. The *power-smoothness* of n is the largest B such that n is B-power-smooth.

The following two PARI functions compute whether or not an integer is B-power-smooth and also the power-smoothness of n.

```
{ispowersmooth(n, B) =   \\ true if and only if n is B-powersmooth
```

```
    local(F,i);
    F = factor(n);
    for(i=1,matsize(F)[1],if(F[i,1]^F[i,2]>B,return(0)));
    return(1);
}

{powersmoothness(n) =    \\ the powersmoothness of n.
    local(F,L,i);
    F = factor(n);
    L = 1;
    for(i=1,matsize(F)[1],L=max(L,F[i,1]^F[i,2]));
    return(L);
}
```

29.2 Pollard's $(p-1)$-Method

Let N be an integer that we wish to factor. Choose a positive integer B (usually $\leq 10^6$ in practice). The Pollard $(p-1)$-method hunts for prime divisors p of N such that $p-1$ is B-power-smooth. Here is the strategy. Suppose that $p \mid N$ and $a > 1$ is an integer that is prime to p. By Fermat's Little Theorem,

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

Assume that further that $p-1$ is B-power-smooth and let $m = \operatorname{lcm}(1,2,3,\ldots B)$. Then $B \mid m$, so $p-1 \mid m$, and so

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

Thus

$$p \mid \gcd(a^m - 1, N) > 1.$$

Usually $\gcd(a^m - 1, N) < N$ also, and when this is the case we have split N. In the unlikely case when $\gcd(a^m - 1, N) = N$, then $a^m \equiv 1 \pmod{q^r}$ for every prime power divisor of N. In this case, repeat the above steps but with a smaller choice of B (so that m is smaller). Also, it's a good idea to check from the start whether or not N is not a perfect power M^r, and if so replace N by M.

In practice, we don't know p. We choose a B, then an a, cross our fingers, and proceed. If we split N, great! If not, increase B or change a and try again.

For fixed B, this algorithm works when N is divisible by a prime p such that $p-1$ is B-power-smooth. How many primes p have the property that $p-1$ is B-power-smooth? Is this very common or not? Using the above two functions, we find that roughly 15% of primes p between 10^{15} and $10^{15}+10000$ are such that $p-1$ is 10^6 power-smooth.

```
\\ Count the number of B-power-smooth numbers an interval.
{cnt(B)= s=0;t=0;
    for(p=10^15, 10^15+10000,
        if(isprime(p),
            t++;if(ispowersmooth(p-1,B),s++)
```

```
    )
  );
  s/t*1.0
}
? cnt(10^6)
%5 = 0.1482889733840304182509505703
```

Thus the Pollard $(p-1)$-method with $B = 10^6$ is blind to 85% of the primes around 10^{15}. There are nontrivial theorems about densities of power-smooth numbers, but I will not discuss them today.

29.3 Pollard's Method in Action!

We now illustrate the Pollard $(p-1)$-method through several examples.

Example 29.3.1. Let $N = 5917$. We try to use the Pollard $p-1$ method with $B = 5$ to split N. We have $m = \mathrm{lcm}(1,2,3,4,5) = 60$. Take $a = 2$. We have

$$2^{60} - 1 \equiv 3416 \pmod{5917}, \qquad \text{(can compute quickly!)}$$

so

$$\gcd(2^{60} - 1, 5917) = \gcd(3416, 5917) = 61.$$

Wow, we found a prime factor of N!

In PARI, these computations are carried out as follows:

```
{lcmfirst(B) =  \\ compute the lcm of 1,2,3,...,B
    local(L,i);
    L=1;
    for(i=2,B,L=lcm(L,i));
    return(L);}
? lcmfirst(5)
%8 = 60
? Mod(2,5917)^60 - 1
%9 = Mod(3416, 5917)
? gcd(3416,5917)
%10 = 61
```

Example 29.3.2. Let $N = 779167$. First try $B = 5$ and $a = 2$:

$$2^{60} - 1 \equiv 710980 \pmod{N},$$

and $\gcd(2^{60} - 1, N) = 1$. Thus no prime divisor p of N has the property that $p - 1$ is 5-power-smooth. Next, we try $B = 15$. We have $m = \mathrm{lcm}(1, 2, \ldots, 15) = 360360$, and

$$2^{360360} - 1 \equiv 584876 \pmod{N},$$

so

$$\gcd(2^{360360} - 1, N) = 2003,$$

and we have split N!

Example 29.3.3. Let $N = 61\cdot 71$. Then both $61-1 = 60 = 2^2\cdot 3\cdot 5$ and $71-1 = 2\cdot 5\cdot 7$ are 7-power-smooth, so Pollard's $(p-1)$-method with any $B \geq 7$ will fail, but in a confidence-inspiring way. Suppose $B = 7$, so $m = \mathrm{lcm}(1, 2, \ldots, 7) = 420$. Then

$$2^{420} - 1 \equiv 0 \pmod{N},$$

so $\gcd(2^{420} - 1, N) = N$, and we get nothing. If we shrink B to 5, then Pollard works:

$$2^{60} - 1 \equiv 1464 \pmod{N},$$

and $\gcd(2^{60} - 1, N) = 61$, so we split N.

29.4 Motivation for the Elliptic Curve Method

Fix an integer B. If $N = pq$ with p and q prime and neither $p-1$ nor $q-1$ a B-power-smooth number, then the Pollard $(p-1)$-method is extremely unlikely to work. For example, let $B = 20$ and suppose that $N = 59 \cdot 101 = 5959$. Note that neither $59 - 1 = 2\cdot 29$ nor $107 - 1 = 2 \cdot 53$ is B-power-smooth. With $m = \mathrm{lcm}(1, 2, 3, \ldots, 20) = 232792560$, we have

$$2^m - 1 \equiv 5944 \pmod{N},$$

and $\gcd(2^m - 1, N) = 1$, so we get nothing.

As remarked above, the problem is that $p-1$ is not 20-power-smooth for either $p = 59$ or $p = 101$. However, notice that $p - 2 = 3 \cdot 19$ is 20-power-smooth! If we could somehow replace the group $(\mathbb{Z}/p\mathbb{Z})^*$, which has order $p-1$, by a group of order $p-2$, and compute a^m for an element of this *new* group, then we might easily split N. Roughly speaking, this is what Lenstra's elliptic curve factorization method does; it replaces $(\mathbb{Z}/p\mathbb{Z})^*$ by an elliptic curve E over $\mathbb{Z}/p\mathbb{Z}$. The order of the group $E(\mathbb{Z}/p\mathbb{Z})$ is $p + 1 \pm s$ for some nonnegative integer $s < 2\sqrt{p}$ (any s can occur). For example, if E is the elliptic curve

$$y^2 = x^3 + x + 54$$

over $\mathbb{Z}/59\mathbb{Z}$ then $E(\mathbb{Z}/59\mathbb{Z})$ is cyclic of order 57. The set of numbers $59 + 1 \pm s$ for $s \leq 15$ contain numbers with very small power-smoothness.

I won't describe the elliptic curve factorization method until the next lecture. The basic idea is as follows. Suppose that we wish to factor N. Choose an integer B. Choose a random point P and a random elliptic curve $y^2 = x^3 + ax + b$ "over $\mathbb{Z}/N\mathbb{Z}$" that goes through P. Let $m = \mathrm{lcm}(1, 2, \ldots, B)$. Try to compute mP working modulo N and using the group law formulas. If at some point it is necessary to divide modulo N, but division is not possible, we (usually) find a nontrivial factor of N. Something going wrong and not being able to divide is analogous to a^m being congruent to 1 modulo p.

More details next time!

29.5 The Elliptic Curve Method

```
{isalmostpowersmooth(p,B)= local(r);
```

```
for(r=p+1-floor(2*sqrt(p)),p+1+floor(2*sqrt(p)),
if(ispowersmooth(r,B), return(1)) ) }
cnt(B)=s=0;t=0;for(p=10^15,10^15+10000,
if(isprime(p),t++;if(isalmostpowersmooth(p,B),s++,print("BAD
",p));print(s/t*1.0))); s/t*1.0
```

29.6 The Method in Action!

Chapter 30

Using Elliptic Curves to Factor, Part II

I constructed $N = 800610470601655221392794180058088102053408423$ by multiplying together five random (and promptly forgotten) primes p with the property that $p-1$ is not B-power-smooth for $B = 10^8$. Since N is a product of five not-too-big primes, N begs to be factored using the elliptic curve method.

30.1 The Elliptic Curve Method (ECM)

The following description of the algorithm is taken from Lenstra's paper [*Factoring Integers with Elliptic Curves*, Annals of Mathematics, **126**, 649–673], which you can download from the Math 124 web page.

Cohen and Lenstra

"The new method is obtained from Pollard's $(p-1)$-method by replacing the multiplicative group by the group of points on a random elliptic curve. To find a non-trivial divisor of an integer $n > 1$, one begins by selecting an elliptic curve E over $\mathbb{Z}/n\mathbb{Z}$, a point P on E with coordinates in $\mathbb{Z}/n\mathbb{Z}$, and an integer k as above [$k = \mathrm{lcm}(2, 3, \ldots, B)$]. Using the addition law of the curve, one next calculates the multiple $k \cdot P$ of P. One now hopes that there is a prime divisor p of n for which $k \cdot P$ and the neutral element $\mathcal{O}$ of the curve become the same modulo p; if E is given by a homogeneous Weierstrass equation $y^2z = x^3 + axz^2 + bz^3$, with $\mathcal{O} = (0 : 1 : 0)$, then this is equivalent to the z-coordinate of $k \cdot P$ being divisible by p. Hence one hopes to find a non-trivial factor of n by calculating the greatest common divisor of this z-coordinate with n."

If the above algorithm fails with a specific elliptic curve E, there is an option that is unavailable with Pollard's $(p-1)$-method. We may repeat the above algorithm with a different choice of E. The number of points on E over $\mathbb{Z}/p\mathbb{Z}$ is of the form $p+1-t$ for some t with $|t| < 2\sqrt{p}$, and the algorithm is likely to succeed if $p+1-t$ is B-power-smoth.

Suppose that $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ are nonzero points on an elliptic curve $y^2 = x^3 + ax + b$ and that $P \neq \pm Q$. Let $\lambda = (y_1 - y_2)/(x_1 - x_2)$ and

$\nu = y_1 - \lambda x_1$. Recall that $P + Q = (x_3, y_3)$ where

$$x_3 = \lambda^2 - x_1 - x_2 \qquad \text{and} \qquad y_3 = -\lambda x_3 - \nu.$$

If we do arithmetic on an elliptic curve modulo N and at some point we can not compute λ because we can not compute the inverse modulo N of $x_1 - x_2$, then we (usually) factor N.

30.2 Implementation and Examples

For simplicity, we use an elliptic curve of the form

$$y^2 = x^3 + ax + 1,$$

which has the point $P = (0, 1)$ already on it.

The following tiny PARI function implements the ECM. It generates an error message along with a usually nontrivial factor of N exactly when the ECM succeeds.

```
{ECM(N, m)= local(E);
   E = ellinit([0,0,0,random(N),1]*Mod(1,N));
   print("E: y^2 = x^3 + ",lift(E[4]),"x+1,  P=[0,1]");
   ellpow(E,[0,1]*Mod(1,N),m);  \\ this fails if and only if we win!
}
```

The following two functions are also useful:

```
{lcmfirst(B) =
   local(L,i); L=1; for(i=2,B,L=lcm(L,i));
   return(L);
}
numpoints(a,p) = return(p+1 - ellap(ellinit([0,0,0,a,1]),p));
```

First we will try the program on a small integer N, then we will try it on the N at the top of this lecture. (`ECM` uses the random function, so the results of your run may differ from the one below.)

```
? N = 5959;           \\ This number motivated the ECM last time.
\\ Recall what happened when we tried to factor 5959 using the p-1 method.
? m = lcmfirst(20);   \\ B = 20.
? Mod(2,N)^m-1
%108 = Mod(5944, 5959)
? gcd(5944,5959)
%109 = 1               \\ bummer!
\\ Now we try the ECM:
? ECM(N,m)
E: y^2 = x^3 + 1201x+1,  P=[0,1]
%112 = [Mod(666, 5959), Mod(3229, 5959)]
? ECM(N,m)
E: y^2 = x^3 + 1913x+1,  P=[0,1]
  ***   impossible inverse modulo: Mod(101, 5959).
```

```
\\ Wonderful!! There's a factor--------/\
? factor(numpoints(1913,101))
%120 =
[2 4]                          \\ #E(Z/101) is 16-power-smooth,
[7 1]                          \\ so ECM sees 101.
? factor(numpoints(1913,59))
%119 =
[2 1]                          \\ #E(Z/59) is 29-power-smooth,
[29 1]                         \\ so ECM doesn't see 59.

\\ Here's the view from another angle:
? E = ellinit([0,0,0,1752,0]*Mod(1,5959));
? P = [0,1]*Mod(1,5959);
? ellpow(E,P,2)
%127 = [Mod(4624, 5959), Mod(1495, 5959)]
? ellpow(E,P,3)
%128 = [Mod(3435, 5959), Mod(1031, 5959)]
? ellpow(E,P,4)
%129 = [Mod(803, 5959), Mod(5856, 5959)]
? ellpow(E,P,8)
%133 = [Mod(1347, 5959), Mod(2438, 5959)]
? ellpow(E,P,m)
  ***   impossible inverse modulo: Mod(101, 5959).
```

Now we are ready to try the big integer N from the begining of the lecture.

```
? N = 800610470601655221392794180058088102053408423;
? B = 100;
? m = lcmfirst(B);
? ECM(N,m);
E: y^2 = x^3 + 273687051132207711452727265152539544370874547x+1,  P=[0,1]
... many tries ..
? ECM(N,m);
E: y^2 = x^3 + 174264237886300715545169749498695137077020788x+1,  P=[0,1]
? B=1000;  \\ give up and try a bigger B.
? m=lcmfirst(B);
? ECM(N,m);
E: y^2 = x^3 + 652986182461202633808585244537305097270008449x+1,  P=[0,1]
... many tries ...
? ECM(N,m);
E: y^2 = x^3 + 755060727645891482095225151281965348765197238x+1,  P=[0,1]
? B=10000; \\ try an even bigger B
? m=lcmfirst(B);
? ECM(N,m);
E: y^2 = x^3 + 722355978919416556225676818691766898771312229x+1,  P=[0,1]
? ECM(N,m);
E: y^2 = x^3 + 124781379199538996805045456359983628056546634x+1,  P=[0,1]
? ECM(N,m);
```

```
E: y^2 = x^3 + 3503107156272519792781442715947445140523646663x+1,  P=[0,1]
? ECM(N,m);
E: y^2 = x^3 + 3963850014650323091382382956262041054794730 7x+1,  P=[0,1]
  ***   impossible inverse modulo: Mod(1004320322301182911,
                                800610470601655221392794180058088102053408423).
```

Thus $N = N_1 \cdot N_2 = 1004320322301182911 \cdot 797166454590134548773760793$. One checks that neither N_1 nor N_2 is prime. Next we try ECM on each:

```
? N1 = 1004320322301182911; N2 = N / N1;
? ECM(N1,m);
E: y^2 = x^3 + 725771039569085210x+1,  P=[0,1]
  ***   impossible inverse modulo: Mod(1406051123, 1004320322301182911).
? ECM(N2,m);
E: y^2 = x^3 + 573369475441522110156437806x+1,  P=[0,1]
  ***   impossible inverse modulo: Mod(2029256729,
                                            797166454590134548773760793).
```

Now

$N = N_{1,1} \cdot N_{1,2} \cdot N_{2,1} \cdot N_{2,2} = 1406051123 \cdot 714284357 \cdot 2029256729 \cdot 392836669307471617,$

and one can check that $N_{1,1}$, $N_{1,2}$, $N_{2,1}$ are prime but that $N_{2,2}$ is composite. Again, we apply ECM:

```
? N22 = 392836669307471617
%173 = 392836669307471617
? ECM(N22,m)
E: y^2 = x^3 + 133284810657519512x+1,  P=[0,1]
%174 = [0]
? ECM(N22,m)
E: y^2 = x^3 + 368444010842952211x+1,  P=[0,1]
%175 = [Mod(236765299763600601, 392836669307471617),
                       Mod(63845045623767003, 392836669307471617)]
? ECM(N22,m)
E: y^2 = x^3 + 245772885854824846x+1,  P=[0,1]
%176 = [0]
? ECM(N22,m)
E: y^2 = x^3 + 33588046732320063x+1,  P=[0,1]
  ***   impossible inverse modulo: Mod(615433499, 392836669307471617).
```

This time it took a long time to factor $N_{2,2}$ because m is too large so we often get both factors. A smaller m would have worked more quickly. In any case, we discover that the prime factorization is

$$N = 1406051123 \cdot 714284357 \cdot 2029256729 \cdot 615433499 \cdot 638308883.$$

30.3 How Good is ECM?

According to Henri Cohen (page 476 of *A Course in Computational Algebraic Number Theory*):

“Unique among modern factoring algorithms however, it is sensitive to the size of the prime divisors . In other words, its running time depends on the size of the smallest prime divisor p of N, and not on N itself. Hence, it can be profitably used to remove “small” factors [...]. Without too much trouble, it can find prime factors having 10 to 20 decimal digits [with B around 10^4]. On the other hand, it very rarely finds prime factors having more than 30 decimal digits.

Chapter 31

Fermat's Last Theorem and Modularity of Elliptic Curves

In this lecture I will sketch an outline of the proof of Fermat's last theorem, then give a rigorous account of what it means for an elliptic curve to be "modular".

The are several exercises below. They are optional, but if you do them and give them to Grigor, I suspect that he would look at them (whether or not you do the exercises will not directly affect your course grade in any way).

31.1 Fermat's Last Theorem

Theorem 31.1.1. *Let $n > 2$ be an integer. If $a, b, c \in \mathbb{Z}$ and*

$$a^n + b^n = c^n,$$

then $abc = 0$.

Proof (sketch). First reduce to the case when $n = \ell$ is a prime greater than 3 (see Exercise 31.1.2). Suppose that

$$a^\ell + b^\ell = c^\ell$$

with $a, b, c \in \mathbb{Z}$ and $abc \neq 0$. Permuting (a, b, c), we may suppose that b is even and that we have $a \equiv 3 \pmod 4$. Following Gerhard Frey, consider the elliptic curve E defined by

$$y^2 = x(x - a^\ell)(x + b^\ell).$$

The discriminant of E is $2^4(abc)^{2\ell}$ (see Exercise 31.1.3 below).

Andrew Wiles and Richard Taylor [Annals of Math., May 1995] proved that E must be "modular". This means that there is a "modular form"

$$f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z}$$

of "level $N = abc$" such that for all primes $p \nmid abc$,

$$a_p = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}).$$

Ken Ribet [Inventiones Math., 1991] used that the discriminant of E is a perfect ℓth power (away from 2) to prove that there is a cuspidal modular form

$$g(z) = \sum_{n=1}^{\infty} b_n e^{2\pi i n z}$$

of "level 2" such that

$$a_p \equiv b_p \pmod{\ell} \qquad \text{for all } p \nmid abc.$$

This is a contradiction because the space of "cuspidal modular forms" of level 2 has dimension 0 (see Section 31.3.1). □

Exercise 31.1.2. Reduce to the prime case. That is, show that if Fermat's last theorem is true for prime exponents, then it is true.

Exercise 31.1.3. Prove that $y^2 = x(x - a^\ell)(x + b^\ell)$ has discriminant $2^4(abc)^{2\ell}$.

The rest of this lecture is about the words in the proof that are in quotes.

31.2 Holomorphic Functions

The complex *upper half plane* is the set

$$\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}.$$

A *holomorphic function* $f : \mathfrak{h} \to \mathbb{C}$ is a function such that for all $z \in \mathfrak{h}$ the derivative

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists. Holomorphicity is a very strong condition because $h \in \mathbb{C}$ can approach 0 in many ways.

Example 31.2.1. Let $\mathrm{SL}_2(\mathbb{Z})$ denote the set of 2×2 integers matrices with determinant 1. If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, then the corresponding *linear fractional transformation*

$$\gamma(z) = \frac{az + b}{cz + d}$$

is a holomorphic function on $\mathfrak{h}$. (Note that the only possible pole of γ is $-\frac{d}{c}$, which is not an element of $\mathfrak{h}$.)

For future use, note that if $f : \mathfrak{h} \to \mathbb{C}$ is a holomorphic function, and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, then

$$f|_\gamma(z) = f(\gamma(z))(cz + d)^{-2}$$

is again a holomorphic function.

Example 31.2.2. Let $q(z) = e^{2\pi i z}$. Then q is a holomorphic function on $\mathfrak{h}$ and $q' = 2\pi i q$. Moreover, q defines a surjective map from $\mathfrak{h}$ onto the punctured open unit disk $D = \{z \in \mathbb{C} : 0 < |z| < 1\}$.

31.3 Cuspidal Modular Forms

Let N be a positive integer and consider the set

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \,:\, N \mid c \right\}.$$

Definition 31.3.1 (Cuspidal Modular Form). A *cuspidal modular form* of level N is a holomorphic function $f : \mathfrak{h} \to \mathbb{C}$ such that

1. $f|_\gamma = f$ for all $\gamma \in \Gamma_0(N)$,
2. for every $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,
$$\lim_{z\to\infty} f(\gamma(z)) = 0,$$
and
3. f has a Fourier expansion:
$$f = \sum_{n=1}^{\infty} a_n q^n.$$

Exercise 31.3.2. Prove that condition 3 is implied by conditions 1 and 2, so condition 3 is redundant. [Hint: Since $\gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$, condition 1 implies that $f(z+1) = f(z)$, so there is a function $F(q)$ on the open punctured unit disc such that $F(q(z)) = f(z)$. Condition 2 implies that $\lim_{q\to 0} F(q) = 0$, so by complex analysis F extends to a holomorphic function on the full open unit disc.]

Definition 31.3.3. The *q-expansion* of f is the Fourier expansion $f = \sum_{n=1}^{\infty} a_n q^n$.

Exercise 31.3.4. Suppose that $f \in S_2(\Gamma_0(N))$. Prove that

$$f(z)dz = f(\gamma(z))d(\gamma(z))$$

for all $\gamma \in \Gamma_0(N)$. [Hint: This is simple algebraic manipulation.]

Exercise 31.3.5. Let $S_2(\Gamma_0(N))$ denote the set of cuspidal modular forms of level N. Prove that $S_2(\Gamma_0(N))$ forms a $\mathbb{C}$-vector space under addition.

31.3.1 The Dimension of $S_2(\Gamma_0(N))$

The dimension of $S_2(\Gamma_0(N))$ is

$$\dim_{\mathbb{C}} S_2(\Gamma_0(N)) = 1 + \frac{\mu}{12} - \frac{\nu_2}{4} - \frac{\nu_3}{3} - \frac{\nu_\infty}{2},$$

where $\mu = N\prod_{p\mid N}(1+1/p)$, and $\nu_2 = \prod_{p\mid N}\left(1 + \left(\frac{-4}{p}\right)\right)$ unless $4 \mid N$ in which case $\nu_2 = 0$, and $\nu_3 = \prod_{p\mid N}\left(1 + \left(\frac{-3}{p}\right)\right)$ unless $2 \mid N$ or $9 \mid N$ in which case $\nu_3 = 0$, and $\nu_\infty = \sum_{d\mid N} \varphi(\gcd(d, N/d))$. For example,

$$\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = 1 + \frac{3}{12} - \frac{1}{4} - \frac{0}{3} - \frac{2}{2} = 0,$$

and

$$\dim_{\mathbb{C}} S_2(\Gamma_0(11)) = 1 + \frac{12}{12} - \frac{0}{4} - \frac{0}{3} - \frac{2}{2} = 1.$$

One can prove that the vector space $S_2(\Gamma_0(11))$ has basis

$$f = q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 = q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 + \cdots.$$

Exercise 31.3.6. Compute the dimension of $S_2(\Gamma_0(25))$.

31.4 Modularity of Elliptic Curves

Let E be an elliptic curve defined by a Weierstrass equation $y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Q}$. For each prime $p \nmid \Delta = -16(4a^3 + 27b^2)$, set

$$a_p = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}).$$

Definition 31.4.1 (Modular). E is *modular* if there exists a cuspidal modular form

$$f(z) = \sum_{n=1}^{\infty} b_n q^n \in S_2(\Gamma_0(\Delta))$$

such that $b_p = a_p$ for all $p \nmid \Delta$.

At first glance, modularity appears to be a bizarre and unlikely property for an elliptic curve to have. When poor Taniyama (and Shimura) first suggested in 1955 that every elliptic curve is modular, people were dubious. But Taniyama was right. The proof of that conjecture is one of the crowning achievements of number theory.

Theorem 31.4.2 (Breuil, Conrad, Diamond, Taylor, Wiles).

Every elliptic curve over $\mathbb{Q}$ is modular.

Wiles

Chapter 32

The Birch and Swinnerton-Dyer Conjecture, Part 1

The next three lectures will be about the Birch and Swinnerton-Dyer conjecture, which is considered by many people to be the most important accessible open problem in number theory. Today I will guide you through Wiles's Clay Math Institute paper on the Birch and Swinnerton-Dyer conjecture.

On Friday, I will talk about the following open problem, which is a frustrating specific case of the Birch and Swinnerton-Dyer conjecture. Let E be the elliptic curve defined by

$$y^2 + xy = x^3 - x^2 - 79x + 289.$$

Denote by $L(E,s) = \sum_{n=1}^{\infty} a_n n^{-s}$ the corresponding L-series, which extends to a function everywhere. The graph of $L(E,s)$ for $s \in (0,5)$ is given on the next page. It can be proved that $E(\mathbb{Q}) \approx \mathbb{Z}^4$ by showing that

$$(8,7),\ \left(\frac{120}{27}, \frac{29}{27}\right),\ \left(\frac{70}{8}, \frac{81}{8}\right),\ \text{ and }\ \left(\frac{564}{8}, \frac{665}{64}\right)$$

generate a "subgroup of finite index" in $E(\mathbb{Q})$. The Birch and Swinnerton-Dyer Conjecture then predicts that

$$\operatorname{ord}_{s=1} L(E,s) = 4,$$

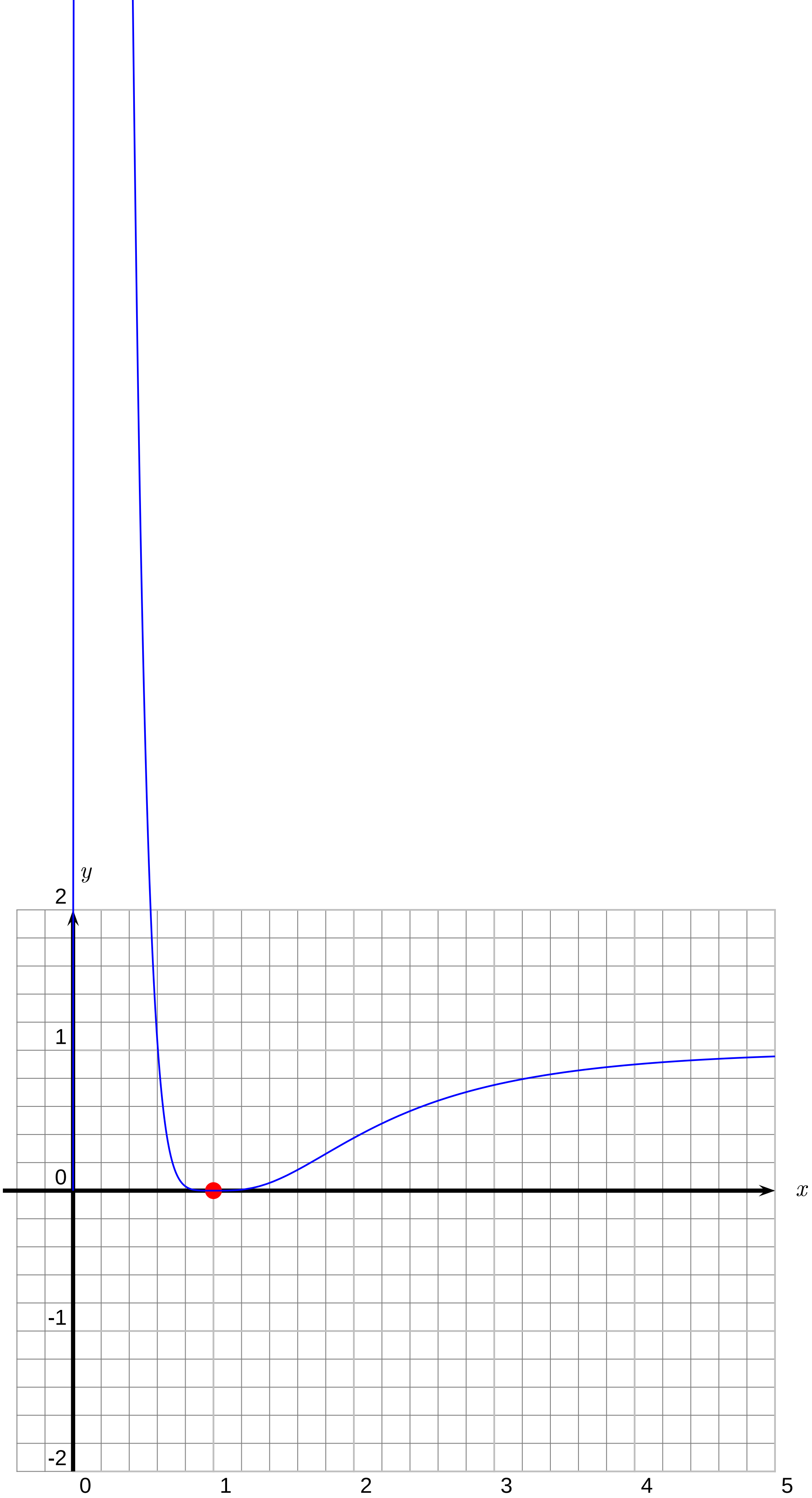
which looks plausible from the shape of the graph on the next page. It is relatively easy to prove that the following is equivalent to showing that $\operatorname{ord}_{s=1} L(E,s) = 4$:

Open Problem: *Prove that $L''(E,1) = 0$.*

If you could solve this open problem, people like Gross, Tate, Mazur, Zagier, Wiles, me, etc., would be **very** excited. The related problem of giving an example of an L-series with $\operatorname{ord}_{s=1} L(E,s) = 3$, was solved as a consequence of a very deep theorem of Gross and Zagier, and resulting in an effective solution to Gauss's class number problem.

John Tate gave a talk about the BSD conjecture for the Clay Math Institute. I strongly encourage you to watch it online at

`http://www.msri.org/publications/ln/hosted/cmi/2000/cmiparis/index-tate.html`



The L-series of the "simplest" known elliptic curve of rank 4.

Chapter 33

The Birch and Swinnerton-Dyer Conjecture, Part 2

33.1 The BSD Conjecture

Let E be an elliptic curve over $\mathbb{Q}$ given by an equation

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

with $a, b \in \mathbb{Z}$. For $p \nmid \Delta = -16(4a^3 + 27b^2)$, let $a_p = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z})$. Let

$$L(E, s) = \prod_{p \nmid \Delta} \frac{1}{1 - a_p p^{-s} + p^{1-2s}}.$$

Theorem 33.1.1 (Breuil, Conrad, Diamond, Taylor, Wiles). *$L(E, s)$ extends to an analytic function on all of $\mathbb{C}$.*

Conjecture 33.1.2 (Birch and Swinnerton-Dyer). *The Taylor expansion of $L(E, s)$ at $s = 1$ has the form*

$$L(E, s) = c(s-1)^r + \text{higher order terms}$$

with $c \neq 0$ and $E(\mathbb{Q}) \approx \mathbb{Z}^r \times E(\mathbb{Q})_{\text{tor}}$.

A special case of the conjecture is the assertion that $L(E, 1) = 0$ if and only if $E(\mathbb{Q})$ is infinite. The assertion "$L(E, 1) = 0$ implies that $E(\mathbb{Q})$ is infinite" is the part of the conjecture that secretely motives much of my own research.

33.2 What is Known

On page 5 of Wiles's paper, he discusses the history of the following theorem.

Theorem 33.2.1 (Gross, Kolyvagin, Zagier, et al.). *Suppose that*

$$L(E, s) = c(s-1)^r + \text{higher order terms}$$

with $r \leq 1$. Then the Birch and Swinnerton-Dyer conjecture is true for E, that is, $E(\mathbb{Q}) \approx \mathbb{Z}^r \oplus E(\mathbb{Q})_{\text{tor}}$.

I suspect that most elliptic curves satisfy the hypothesis of the above theorem, i.e., they have rank 0 or 1. For example, almost 96% of the "first 78198" elliptic curves have $r \leq 1$. I suspect that the curves with $r > 1$ have "density" 0 amongst all elliptic curves. This doesn't mean that we are done. In practice it is often the curves with $r > 1$ that are interesting and useful, and experts can still be observed saying "almost nothing is known about the Birch and Swinnerton-Dyer conjecture".

33.3 How to Compute $L(E, s)$ with a Computer

33.3.1 Best Models

Let E be an elliptic curve over $\mathbb{Q}$, defined by a Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

There are many choices of Weierstrass equations that define an elliptic curve that is "essentially the same" as E. E.g., you found others by completing the square. Among all of these, there is a best possible model, which is the one with smallest discriminant. It can be computed in PARI as follows:

```
? E = ellinit([0,0,0,-43,166]);
? E.disc
%61 = -6815744
? E = ellchangecurve(E,ellglobalred(E)[2])
%62 = [1, -1, 1, -3, 3, ...]
? E.disc
%63 = -1664
```

Thus $y^2 + xy + y = x^3 - x^2 - 3x + 3$ is a "better" model than $y^2 = x^3 - 43x + 166$.

WARNING: Some of the elliptic curves functions in PARI will *LIE* if you give as input an elliptic curve that is defined by a model that isn't the best possible. These devious liars include `elltors, ellap, ellak`, and `elllseries`.

33.3.2 Formula for $L(E, s)$

As mentioned before, the PARI function `elllseries` can compute $L(E, s)$. I figured out how this function works, and explain it below.

Because E is modular, one can show that we have the following rapidly-converging series expression for $L(E, s)$, for $s > 0$:

$$L(E, s) = N^{-s/2} \cdot (2\pi)^s \cdot \Gamma(s)^{-1} \cdot \sum_{n=1}^{\infty} a_n \cdot (F_n(s-1) - \varepsilon F_n(1-s))$$

where

$$F_n(t) = \Gamma\left(t+1, \frac{2\pi n}{\sqrt{N}}\right) \cdot \left(\frac{\sqrt{N}}{2\pi n}\right)^{t+1}.$$

Here

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt$$

is the Γ-*function* (e.g., $\Gamma(n) = (n-1)!$), and

$$\Gamma(z, \alpha) = \int_{\alpha}^{\infty} t^{z-1} e^{-t} dt$$

is the *incomplete* Γ-*function*. The number N is called the *conductor* of E and is very similar to the discriminant of E; it is only divisible by primes that divide the best possible discriminant of E. You can compute N using the PARI command `ellglobalred(E)[1]`.

As usual, for $p \nmid \Delta$, we have

$$a_p = p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}),$$

and for $r \geq 2$,

$$a_{p^r} = a_{p^{r-1}} a_p - p a_{p^{r-2}},$$

and $a_{nm} = a_n a_m$ if $\gcd(n, m) = 1$ (I won't define the a_p when $p \mid \Delta$, but it's not difficult.) Finally, ε depends only on E and is either $+1$ or -1. I won't define ε either, but you can compute it in PARI using `ellrootno(E)`.

At $s = 1$, the formula can be massively simplified, and we have

$$L(E, 1) = (1 + \varepsilon) \cdot \sum_{n=1}^{\infty} \frac{a_n}{n} e^{-2\pi n/\sqrt{N}}.$$

This sum converges rapidly, because $e^{-2\pi n/\sqrt{N}} \to 0$ quickly as $n \to \infty$.

Monday's lecture will be filled with numerical examples and numerical evidence for the Birch and Swinnerton-Dyer conjecture. Wednesday's lecture will be a review for the take-home **FINAL EXAM**.

Chapter 34

The Birch and Swinnerton-Dyer Conjecture, Part 3

34.1 A Rationality Theorem

In the last lecture, I mentioned that it can be surprisingly difficult to say anything precise about $L(E, s)$, even with the above formulas. For example, it is a very deep theorem of Gross and Zagier that for the elliptic curve $y^2 + y = x^3 - 7x + 6$ we have

$$L(E, s) = c(s-1)^3 + \text{ higher order terms,}$$

and nobody has any idea how to prove that there is an elliptic curve with

$$L(E, s) = c(s-1)^4 + \text{ higher order terms.}$$

Fortunately, it is possible to decide whether or not $L(E, 1) = 0$.

Theorem 34.1.1. *Let $y^2 = x^3 + ax + b$ be an elliptic curve. Let*

$$\Omega_E = 2^n \int_\gamma^\infty \frac{dx}{\sqrt{x^3 + ax + b}},$$

where γ is the largest real root of $x^3 + ax + b$, and $n = 0$ if $\Delta(E) < 0$, $n = 1$ if $\Delta(E) > 0$. Then

$$\frac{L(E, 1)}{\Omega_E} \in \mathbb{Q},$$

and the denominator is ≤ 24.

In practice, one computes Ω_E using the "Arithmetic-Geometric Mean", *NOT* numerical integration. In PARI, Ω_E is approximated by `E.omega[1]*2^(E.disc>0)`.

Remark 34.1.2. I don't know if the denominator is ever really as big as 24. It would be a fun student project to either find an example, or to understand the proof that the quotient is rational and prove that 24 can be replaced by something smaller.

Example 34.1.3. Let E be the elliptic curve $y^2 = x^3 - 43x + 166$. We compute $L(E, 1)$ using the above formula and observe that $L(E, 1)/\Omega_E$ appears to be a rational number, as predicted by the theorem.

```
? E = ellinit([0,0,0,-43,166]);
? E = ellchangecurve(E, ellglobalred(E)[2]);
? eps = ellrootno(E)
%77 = 1
? N = ellglobalred(E)[1]
%78 = 26
? L = (1+eps) * sum(n=1,100, ellak(E,n)/n * exp(-2*Pi*n/sqrt(N)))
%79 = 0.6209653495490554663758626727
? Om = E.omega[1]*2^(E.disc>0)
%80 = 4.346757446843388264631038710
? L/Om
%81 = 0.1428571428571428571428571427
? contfrac(L/Om)
%84 = [0, 7]
? 1/7.0
%85 = 0.1428571428571428571428571428
? elltors(E)
%86 = [7, [7], [[1, 0]]]
```

Notice that in this example, $L(E,1)/\Omega_E = 1/7 = 1/\#E(\mathbb{Q})$. This is shadow of a more refined conjecture of Birch and Swinnerton-Dyer.

Example 34.1.4. In this example, we verify that $L(E,1) = 0$ computationally.

```
? E=ellinit([0, 1, 1, -2, 0]);
? L1 = elllseries(E,1)
%4 = -6.881235151133426545894712438 E-29
? Omega = E.omega[1]*2^(E.disc>0)
%5 = 4.980425121710110150642715583
? L1/Omega
%6 = 1.795732353252503036074927634 E-20
```

34.2 Approximating the Rank

Fix an elliptic curve E over $\mathbb{Q}$.

The usual method to *approximate* the rank is to find a series that rapidly converges to $L^{(r)}(E,1)$ for $r = 0,1,2,3,\ldots$, then compute $L(E,1)$, $L'(E,1)$, $L^{(2)}(E,1)$, etc., until one appears to be nonzero. You can read about this method in §2.13 of Cremona's book *Algorithms for Elliptic Curves*. For variety, I will describe a slightly different method that I've played with recently, which uses the formula for $L(E,s)$ from the last lecture, the definition of the derivative, and a little calculus.

Proposition 34.2.1. *Suppose that*

$$L(E,s) = c(s-1)^r + \text{ higher terms.}$$

Then

$$\lim_{s\to 1}(s-1)\cdot\frac{L'(E,s)}{L(E,s)} = r.$$

Proof. Write

$$L(s) = L(E,s) = c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots .$$

Then

$$\begin{aligned}
\lim_{s\to 1}(s-1)\cdot\frac{L'(s)}{L(s)} &= \lim_{s\to 1}(s-1)\cdot\frac{rc_r(s-1)^{r-1} + (r+1)c_{r+1}(s-1)^r + \cdots}{c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots}\\
&= r\cdot\lim_{s\to 1}\frac{c_r(s-1)^r + \frac{(r+1)}{r}c_{r+1}(s-1)^{r+1} + \cdots}{c_r(s-1)^r + c_{r+1}(s-1)^{r+1} + \cdots}\\
&= r.
\end{aligned}$$

□

Thus the rank r is "just" the limit as $s \to 1$ of a certain (smooth) function. We know this limit is an integer. But, for example, for the curve

$$y^2 + xy = x^3 - x^2 - 79x + 289$$

nobody has succeeded in proving that this integer limit is 4. (One can prove that the limit is either 2 or 4.)

Using the definition of derivative, we *heuristically* approximate $(s-1)\frac{L'(s)}{L(s)}$ as follows. For $|s-1|$ small, we have

$$\begin{aligned}
(s-1)\frac{L'(s)}{L(s)} &= \frac{s-1}{L(s)}\cdot\lim_{h\to 0}\frac{L(s+h)-L(s)}{h}\\
&\approx \frac{s-1}{L(s)}\cdot\frac{L(s+(s-1)^2)-L(s)}{(s-1)^2}\\
&= \frac{L(s^2-s+1)-L(s)}{(s-1)L(s)}
\end{aligned}$$

Question 34.2.2. Does

$$\lim_{s\to 1}(s-1)\cdot\frac{L'(s)}{L(s)} = \lim_{s\to 1}\frac{L(s^2-s+1)-L(s)}{(s-1)L(s)}?$$

In any case, we can use this formula in PARI to "approximate" r.

```
? E = ellinit([ 0, 1, 1, -2, 0 ]);
? r(E,s) = L1=elllseries(E,s); L2=elllseries(E,s^2-s+1); (L2-L1)/((s-1)*L1);
? r(E,1.01)
%8 = 2.004135342473941928617680057
? r(E,1.001)
%9 = 2.000431337547225544819319104
\\ One can prove that 2 is the correct limit.
```

Now let's try the mysterious curve $y^2 + xy = x^3 - x^2 - 79x + 289$ of rank 4:

```
? E=ellinit([ 1,-1,0,-79,289]);
? r(E,1.001)          \\ takes 6 seconds on PIII 1Ghz
%1 = 4.002222374519085610896440642
? r(E,1.00001)
%2 = 4.000016181256911064613006133
```

It certainly looks like $\lim_{s\to 1} r(s) = 4$. We know for a fact that $\lim_{s\to 1} r(s) \in \mathbb{Z}$, and if only there were a good way to bound the error we could conclude that the limit is 4. But this has stumped people for years, and maybe it is impossible without a very deep result that somehow interprets this limit in a different way. This problem has totally stumped the experts for years. We desperately need a new idea!!

If one of you wants to do a reading or research project on this problem in the next year or two, let me know. One could draw pictures of $L^{(3)}(E,s)$ or investigate the analogous problem for other more accessible L-series.

```
? E=ellinit([0,0,1,-7,6]);
? r(E,s) = L1=elllseries(E,s); L2=elllseries(E,s^2-s+1); (L2-L1)/((s-1)*L1);
? r(E,1.001)
%2 = 3.0011441049856192065044448552
```

Chapter 35

Homework

35.1 Primes and the Euclidean Algorithm

1. Let p be a prime number and r and integer such that $1 \leq r < p$. Prove that p divides the binomial coefficient
$$\frac{p!}{r!(p-r)!}.$$
You may not assume that this coefficient is a integer.

2. Compute the following gcd's using a pencil and the Euclidean algorithm:
$$\gcd(15, 35), \quad \gcd(247, 299), \quad \gcd(51, 897), \quad \gcd(136, 304)$$

3. Using mathematical induction to prove that
$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2},$$
then find a formula for
$$1 - 2 + 3 - 4 + \cdots \pm n = \sum_{a=1}^{n} (-1)^{a-1} a.$$

4. What was the most recent prime year? I.e., which of $2001, 2000, \ldots$ was it?

5. Use the Euclidean algorithm to find integers x, $y \in \mathbb{Z}$ such that
$$2261x + 1275y = 17.$$
[I did not tell you how to do this; see §1.8 of Davenport's book.]

6. Factor the year that you should graduate from Harvard as a product of primes. E.g., frosh answer $2005 = 5 \times 401$.

7. Write a PARI program to print "Hello Kitty" five times.

8. Let $f(x) \in \mathbb{Z}[x]$ be a polynomial with integer coefficients. Formulate a conjecture about when the set $\{f(a) : a \in \mathbb{Z} \text{ and } f(a) \text{ is prime }\}$ is infinite. Give computational evidence for your conjecture.

9. Is it easy or hard for PARI to compute the gcd of two random 2000-digit numbers?

10. Prove that there are infinitely many primes of the form $6x - 1$.

11. (a) Use PARI to compute

$$\pi(2001) = \#\{\text{ primes } p \leq 2001\}.$$

    (b) The prime number theorem predicts that $\pi(x)$ is asymptotic to $x/\log(x)$. How close is $\pi(2001)$ to $2001/\log(2001)$?

35.2 Congruences

1. Find complete sets of residues modulo 7, all of whose elements are (a) nonnegative, (b) odd, (c) even, (d) prime.

2. Find an integer x such that $37x \equiv 1 \pmod{101}$.

3. What is the order of 5 modulo 37?

4. Let $n = \varphi(7!)$. Compute the prime factorization of n.

5. Find $x, y \in \mathbb{Z}$ such that

$$6613x + 8947y = 389.$$

6. Find an $x \in \mathbb{Z}$ such that $x \equiv -4 \pmod{17}$ and $x \equiv 3 \pmod{23}$.

7. Compute $7^{100} \pmod{389}$.

8. Find a number a such that $0 \leq a < 111$ and

$$(102^{70} + 1)^{35} \equiv a \pmod{111}.$$

   (See Problem 2.05 on page 217 of Davenport.)

9. Prove that if $n > 4$ is composite then

$$(n-1)! \equiv 0 \pmod{n}.$$

10. For what values of n is $\varphi(n)$ odd?

11. Find your own 100-digit number n such that $a^{n-1} \equiv 1 \pmod{n}$ for $a = 2, 3, 5$.

12. Seven thieves try to share a hoard of gold bars equally between themselves. Unfortunately, six bars are left over, and in the fight over them, one thief is killed. The remaining six thieves, still unable to share the bars equally since two are left over, again fight, and another is killed. When the remaining five share the bars, one bar is left over, and it is only after yet another thief is killed that an equal sharing is possible. What is the minimum number of bars which allows this to happen?

13. An elderly woman goes to a market where a horse tramples her basket crushing her eggs. The horse's honest rider offers to pay for the damages and asks her how many eggs she had brought. She doesn't remember the exact number, but recalls that when she had taken them out two at a time, there was one egg left. The same happened when she picked them out three, four, five, and six at a time, but when she took them out seven at a time two were left. What is the smallest number of eggs she could have had?

35.3 Public-Key Cryptography

1. (3 points) You and Nikita wish to agree on a secret key using the Diffie-Hellman protocol. Nikita announces that $p = 3793$ and $g = 7$. Nikita secretely chooses a number $n < p$ and tells you that $g^n \equiv 454 \pmod{p}$. You choose the random number $m = 1208$. Tell me what the secret key is!

2. (4 points) This problem concerns encoding phrases using numbers.

   (a) Find the number that corresponds to `VE␣RI␣TAS`, where we view this string as a number in base 27 using the encoding of Section 2 of Lecture 9. (Note that the left-most "digit", `V`, is the least significant digit, and ␣ denotes a blank space.)

   (b) What is the longest sequence of letters (and space) that can be stored using a number that is less than 10^{20}?

3. (4 points) You see Michael and Nikita agree on a secret key using the Diffie-Hellman key exchange protocol. Michael and Nikita choose $p = 97$ and $g = 5$. Nikita chooses a random number n and tells Michael that $g^n \equiv 3 \pmod{97}$, and Michael chooses a random number m and tells Nikita that $g^m \equiv 7 \pmod{97}$. Crack their code: What is the secret key that Nikita and Michael agree upon? What is n? What is m?

4. (2 points) Using the RSA public key is $(n, e) = (441484567519, 238402465195)$, encrypt the year that you will graduate from Harvard.

5. (6 points) In this problem, you will "crack" an RSA cryptosystem.

   (a) What is the secret decoding number d for the RSA cryptosystem with public key $(n, e) = (5352381469067, 4240501142039)$?

   (b) The number 3539014000459 encrypts an important question using the RSA cryptosystem from part (a). What is the question? (After decoding, you'll get a number. To find the corresponding word, see Section 2 of Lecture 9.)

6. (4 points) Suppose Michael creates an RSA cryptosystem with a very large modulus N for which the factorization of N cannot be found in a reasonable amount of time. Suppose that Nikita sends messages to Michael by representing each alphabetic character as an integer between 0 and 26 (`A` corresponds to 1, `B` to 2, etc., and a space ␣ to 0), then encrypts each number *separately* using Michael's RSA cryptosystem. Is this method secure? Explain your answer.

7. (6 points) Nikita creates an RSA cryptosystem with public key

$$(n, e) = (1433811615146881, 329222149569169).$$

In the following two problems, show the steps you take. Don't simply factor n directly using the `factor` function in PARI.

   (a) Somehow you discover that $d = 116439879930113$. Show how to use the probabilistic algorithm of Lecture 10 to use d to factor n.

   (b) In part (a) you found that the factors p and q of n are very close. Show how to use the "Fermat Factorization" method of Lecture 10 to factor n.

35.4 Primitive Roots and Quadratic Reciprocity

1. (2 points) Calculate the following symbols by hand: $\left(\frac{3}{97}\right)$, $\left(\frac{5}{389}\right)$, $\left(\frac{2003}{11}\right)$, and $\left(\frac{5!}{7}\right)$.

2. (3 points) Prove that $\left(\dfrac{3}{p}\right) = \begin{cases} 1 & \text{if } p \equiv 1, 11 \pmod{12}, \\ -1 & \text{if } p \equiv 5, 7 \pmod{12}. \end{cases}$

3. (3 points) Prove that there is no primitive root modulo 2^n for any $n \geq 3$.

4. (6 points) Prove that if p is a prime, then there is a primitive root modulo p^2.

5. (5 points) Use the fact that $(\mathbb{Z}/p\mathbb{Z})^*$ is cyclic to give a direct proof that $\left(\frac{-3}{p}\right) = 1$ when $p \equiv 1 \pmod 3$. [Hint: There is an $c \in (\mathbb{Z}/p\mathbb{Z})^*$ of order 3. Show that $(2c+1)^2 = -3$.]

6. (6 points) If $p \equiv 1 \pmod 5$, show directly that $\left(\frac{5}{p}\right) = 1$ by the method of Exercise 5. [Hint: Let $c \in (\mathbb{Z}/p\mathbb{Z})^*$ be an element of order 5. Show that $(c+c^4)^2 + (c+c^4) - 1 = 0$, etc.]

7. (4 points) For which primes p is $\displaystyle\sum_{a=1}^{p-1} \left(\frac{a}{p}\right) = 0$?

8. (4 points) Artin conjectured that the number of primes $p \leq x$ such that 2 is a primitive root modulo p is asymptotic to $C\pi(x)$ where $\pi(x)$ is the number of primes $\leq x$ and C is a fixed constant called Artin's constant. Using a computer, make an educated guess as to what C should be, to a few decimal places of accuracy. Explain your reasoning. (Note: Don't try to prove that your guess is correct.)

35.5 Continued Fractions

1. (3 points) Draw some sort of diagram that illustrates the partial convergents of the following continued fractions:

   (i) $[13, 1, 8, 3]$

   (ii) $[1, 1, 1, 1, 1, 1, 1, 1]$

   (iii) $[1, 2, 3, 4, 5, 6, 7, 8]$

2. (5 points) If $c_n = p_n/q_n$ is the nth convergent of the continued fraction $[a_0, a_1, \ldots, a_n]$ and $a_0 > 0$, show that

$$[a_n, a_{n-1}, \ldots, a_1, a_0] = \frac{p_n}{p_{n-1}}$$

   and

$$[a_n, a_{n-1}, \ldots, a_2, a_1] = \frac{q_n}{q_{n-1}}.$$

   (Hint: In the first case, notice that $\frac{p_n}{p_{n-1}} = a_n + \frac{p_{n-2}}{p_{n-1}} = a_n + \frac{1}{\frac{p_{n-1}}{p_{n-2}}}$.)

3. (4 points) There is a function $j(\tau)$, denoted by `ellj` in PARI, which takes as input a complex number τ with positive imaginary part, and returns a complex number called the "j-invariant of the associated elliptic curve". Suppose that τ is *approximately* $-0.5 + 0.3281996289i$ and that you know that $j = j(\tau)$ is a rational number. Use continued fractions and PARI to compute a reasonable guess for the rational number $j = \texttt{ellj}(\tau)$. (Hint: In PARI $\sqrt{-1}$ is represented by `I`.)

4. (3 points) Evaluate each of the following infinite continued fractions:

   (i) $[\overline{2, 3}]$

   (ii) $[2, \overline{1, 2, 1}]$

   (iii) $[0, \overline{1, 2, 3}]$

5. (3 points) Determine the infinite continued fraction of each of the following numbers:

   (i) $\sqrt{5}$

   (ii) $\dfrac{1 + \sqrt{13}}{2}$

   (iii) $\dfrac{5 + \sqrt{37}}{4}$

6. (i) (4 points) For any positive integer n, prove that $\sqrt{n^2+1} = [n, \overline{2n}]$.
   (ii) (2 points) Find a convergent to $\sqrt{5}$ that approximates $\sqrt{5}$ to within four decimal places.

7. (4 points) A famous theorem of Hurwitz (1891) says that for any irrational number x, there exists infinitely many rational numbers a/b such that

$$\left| x - \frac{a}{b} \right| < \frac{1}{\sqrt{5}b^2}.$$

   Taking $x = \pi$, obtain three rational numbers that satisfy this inequality.

8. (3 points) The continued fraction expansion of e is

$$[2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, \ldots].$$

   It is a theorem that the obvious pattern continues indefinitely. Do you think that the continued fraction expansion of e^2 also exhibits a nice pattern? If so, what do you think it is?

9. (i) (4 points) Show that there are infinitely many even integers n with the property that both $n+1$ and $\frac{n}{2}+1$ are perfect squares.
   (ii) (3 points) Exhibit two such integers that are greater than 389.

10. (7 points) A primitive Pythagorean triple is a triple x, y, z of integers such that $x^2 + y^2 = z^2$. Prove that there exists infinitely many primitive Pythagorean triples x, y, z in which x and y are consecutive integers.

35.6 Binary Quadratic Forms

1. (3 points) Which of the following numbers is a sum of two squares? Express those that are as a sum of two squares.

$$-389, \;\; 12345, \;\; 91210, \;\; 729, \;\; 1729, \;\; 68252$$

2. (i) (4 points) Write a PARI program that takes a positive integer n as input and outputs a sequence `[x,y,z,w]` of integers such that $x^2+y^2+z^2+w^2 = n$. (Hint: Your program does not have to be efficient.)
   (ii) (2 point) Write 2001 as a sum of three squares.

3. (3 points) Find a positive integer that has a least three different representations as the sum of two squares, disregarding signs and the order of the summands.

4. (5 points) Show that a natural number n is the sum of two integer squares if and only if it is the sum of two rational squares.

5. (6 points) Mimic the proof of the main theorem of Lecture 21 to show that an odd prime p is of the form $8m+1$ or $8m+3$ if and only if it can be written as $p = x^2 + 2y^2$ for some choice of integers x and y. (Hint: Use the formula for the quadratic residue symbol $\left(\frac{-2}{p}\right)$ from Lecture 13.)

6. (4 points) A *triangular number* is a number that is the sum of the first m integers for some positive integer m. If n is a triangular number, show that all three of the integers $8n^2$, $8n^2+1$, and $8n^2+2$ can be written as a sum of two squares.

7. (3 points) Prove that of any four consecutive integers, at least one is not representable as a sum of two squares.

8. (4 points) Show that $13x^2+36xy+25y^2$ and $58x^2+82xy+29y^2$ are each equivalent to the form x^2+y^2, then find integers x and y such that $13x^2+36xy+25y^2=389$.

9. (4 points) What are the discriminants of the forms $199x^2-162xy+33y^2$ and $35x^2-96xy+66y^2$? Are these forms equivalent?

35.7 Class Groups and Elliptic Curves

1. (10 points) For any negative discriminant D, let C_D denote the finite abelian group of equivalence classes of primitive positive definite quadratic forms of discriminant D. Use the PARI program `forms.gp` from lecture 24 (download it from my web page) to compute representatives for C_D and determine the structure of C_D as a produce of cyclic groups for each of the following five values of D:
$$D=-155,-231,-660,-12104,-10015.$$

2. (6 points) Draw a beautiful graph of the set $E(\mathbb{R})$ of real points on each of the following elliptic curves:

   (i) $y^2=x^3-1296x+11664$,

   (ii) $y^2+y=x^3-x$,

   (iii) $y^2+y=x^3-x^2-10x-20$.

3. (4 points) A rational solution to the equation $y^2-x^3=-2$ is $(3,5)$. Find a rational solution with $x\neq 3$ by drawing the tangent line to $(3,5)$ and computing the third point of intersection.

35.8 Elliptic Curves I

1. (3 points) Consider the elliptic curve $y^2+xy+y=x^3$ over $\mathbb{Q}$. Find a linear change of variables that transforms this curve into a curve of the form $Y^2=X^3+aX+b$ for rational numbers a and b.

2. (6 points) Let E be the elliptic curve over the finite field $K=\mathbb{Z}/5\mathbb{Z}$ defined by the equation
$$y^2=x^3+x+1.$$

   (i) List all 9 elements of $E(K)$.

   (ii) What is the structure of the group $E(K)$, as a product of cyclic groups?

3. (8 points) Let E be an elliptic curve over $\mathbb{Q}$. Define a binary operation $\boxplus$ on E as follows:

$$P \boxplus Q = -(P+Q).$$

Thus the $\boxplus$ of P and Q is the third point of intersection of the line through P and Q with E.

(i) Lists the axiom(s) of a group that fail for $E(\mathbb{R})$ equipped with this binary operation. (The group axioms are "identity", "inverses", and "associativity".)

(ii) Under what conditions on $E(\mathbb{Q})$ does this binary operation define a group structure on $E(\mathbb{Q})$? (E.g., when $E(\mathbb{Q}) = \{\mathcal{O}\}$ this binary operation does define a group.)

4. (6 points) Let $g(t)$ be a quartic polynomial with distinct (complex) roots, and let α be a root of $g(t)$. Let $\beta \neq 0$ be any number.

(i) Prove that the equations

$$x = \frac{\beta}{t-\alpha}, \qquad y = x^2 u = \frac{\beta^2 u}{(t-\alpha)^2}$$

give an "algebraic transformation" between the curve $u^2 = g(t)$ and the curve $y^2 = f(x)$, where $f(x)$ is the cubic polynomial

$$f(x) = g'(\alpha)\beta x^3 + \frac{1}{2}g''(\alpha)\beta^2 x^2 + \frac{1}{6}g'''(\alpha)\beta^3 x + \frac{1}{24}g''''(\alpha)\beta^4.$$

(ii) Prove that if g has distinct (complex) roots, then f also has distinct roots, and so $u^2 = g(t)$ is an elliptic curve.

5. (8 points) In this problem you will finally find out exactly why elliptic curves are called "elliptic curves"! Let $0 < \beta \leq \alpha$, and let C be the ellipse

$$\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1.$$

(i) Prove that the arc length of C is given by the integral

$$4\alpha \int_0^{\pi/2} \sqrt{1 - k^2 \sin^2\theta}\, d\theta$$

for an appropriate choice of constant k depending on α and β.

(ii) Check your value for k in (i) by verifying that when $\alpha = \beta$, the integral yields the correct value for the arc length of a circle.

(iii) Prove that the integral in (i) is also equal to

$$4\alpha \int_0^1 \sqrt{\frac{1-k^2t^2}{1-t^2}}\, dt = 4\alpha \int_0^1 \frac{1-k^2t^2}{\sqrt{(1-t^2)(1-k^2t^2)}}\, dt.$$

(iv) Prove that if the ellipse E is not a circle, then the equation

$$u^2 = (1-t^2)(1-k^2t^2)$$

defines an elliptic curve (cf. the previous exercise). Hence the problem of determining the arc length of an ellipse comes down to evaluating the integral

$$\int_0^1 \frac{1-k^2t^2}{u} dt$$

on the "elliptic" curve $u^2 = (1-t^2)(1-k^2t^2)$.

6. (8 points) Suppose that $P = (x, y)$ is a point on the cubic curve

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

(i) Verify that the x coordinate of the point $2P$ is given by the duplication formula

$$x(2P) = \frac{x^4 - 2ax^2 - 8bx + a^2}{4y^2}.$$

(ii) Derive a similar formula for the y coordinate of $2P$ in terms of x and y.

(iii) Find a polynomial in x whose roots are the x-coordinates of the points $P = (x, y)$ satisfying $3P = \mathcal{O}$. [Hint: The relation $3P = \mathcal{O}$ can also be written $2P = -P$.]

(iv) For the particular curve $y^2 = x^3 + 1$, solve the equation in (iii) to find all of the points satisfying $3P = \mathcal{O}$. Note that you will have to use complex numbers.

35.9 Elliptic Curves II

1. (10 points) Let Φ be the set of the 15 possible groups of the form $E(\mathbb{Q})_{\text{tor}}$ for E an elliptic curve over $\mathbb{Q}$ (see Lecture 27). For each group $G \in \Phi$, if possible, find a finite field $k = \mathbb{Z}/p\mathbb{Z}$ and an elliptic curve E over k such that $E(k) \approx G$. (Hint: It is a fact that $|p + 1 - \#E(\mathbb{Z}/p\mathbb{Z}))| \leq 2\sqrt{p}$, so you only have to try finitely many p to show that a group G does not occur as the group of points on an elliptic curve over a finite field.)

2. (6 points) Many number theorists, such as myself one week ago, incorrectly think that Lutz-Nagell works well in practice. Describe the steps you *would* take if you were to use the Lutz-Nagell theorem (Lecture 27) to compute the torsion subgroup of the elliptic curve E defined by the equation

$$y^2 + xy = x^3 - 8369487776175x + 9319575518172005625,$$

then tell me why it would be *very* time consuming to actually carry these steps out. Find the torsion subgroup of E using the `elltors` command in PARI. Does `elltors` use the Lutz-Nagell algorithm by default?

3. (6 points) Let E be the elliptic curve defined by the equation $y^2 = x^3 + 1$.

(i) For each prime p with $5 \leq p < 30$, describe the group of points on this curve having coordinates in the finite field $\mathbb{Z}/p\mathbb{Z}$. (You can just give the order of each group.)

(ii) For each prime in (i), let N_p be the number of points in the group. (Don't forget the point infinity.) For the set of primes satisfying $p \equiv 2 \pmod 3$, can you see a pattern for the values of N_p? Make a general conjecture for the value of N_p when $p \equiv 2 \pmod 3$.

(iii) Prove your conjecture.

4. (6 points) Let p be a prime and let E be the elliptic curve defined by the equation $y^2 = x^3 + px$. Use Lutz-Nagel to find all points of finite order in $E(\mathbb{Q})$.

5. (4 points)

   (i) Let E be an elliptic curve over the real numbers $\mathbb{R}$. Prove that $E(\mathbb{R})$ is not a finitely generated abelian group.

   (ii) Let E be an elliptic curve over a finite field $k = \mathbb{Z}/p\mathbb{Z}$. Prove that $E(k)$ is a finitely generated abelian group.

35.10 Elliptic Curves III

1. (5 points) Make up a simple example that illustrates how to use the ElGamal elliptic curve cryptosystem (see Lecture 29). You may mention Nikita and Michael if you wish. Be very clear about what you are illustrating so that the grader can effortlessly understand your example.

2. (5 points) Make up an example that illustrates an interesting aspect of the Pollard $(p-1)$ factorization method.

3. (5 points) Make up an example that illustrates something that you consider an interesting aspect of Lenstra's elliptic curves factorization method.

4. (10 points) Let R be a ring. We say that Fermat's last theorem is false in R if there exists $x, y, z \in R$ and $n \in \mathbb{Z}$ with $n \geq 3$ such that $x^n + y^n = z^n$ and $xyz \neq 0$. For which prime numbers p is Fermat's last theorem false in the ring $\mathbb{Z}/p\mathbb{Z}$?[1]

[1] This problem was on the dreaded Harvard graduate school qualifying examination this year. Every one of the students who took that exam got this problem right.