Daily Update

Algebraic Topics in Computing: Cryptography

MATH 601, Spring 2020

If you find any **typos** in the Lecture Notes, please email me (`witt@ku.edu`). **Many thanks!**

Virtual Class Notes, Week 13 (April 29 - May 1).

---

1 Elliptic curve cryptography

After Lenstra's breakthrough idea of using elliptic curves to factor large integers, providing a more effective attack on RSA and other cryptosystems than the other current algorithms, many of which we've learned, mathematicians and computer scientists began to investigate how elliptic curves can be applied to *designing secure cryptosystems*. This is our final week of Virtual Class Notes, and we are bringing together the "old" and the "new" by learning about cryptosystems that we are very familiar with (from before the first midterm!), shifted to the setting of elliptic curves!

1.1 Elliptic curve discrete logarithm problem

Recall that the Diffie-Hellman key exchange and the ElGamal cryptosystem rely on the difficulty of solving the discrete logarithm problem $g^x \equiv X \bmod p$ for $x \in \mathbb{Z}$, where p is prime and g, X integers not divisible by p (to ensure a solution, we can take g to be a primitive root modulo p, but in each of these applications, a solution exists by design, regardless). Translating this in the context of the multiplicative group of units $(\mathbb{Z}/p\mathbb{Z})^\times$, this is equivalent to fixing $[g], [X] \in (\mathbb{Z}/p\mathbb{Z})^\times$ and finding an integer solution x to $[g]^x = [X]$ in $(\mathbb{Z}/p\mathbb{Z})^\times$.

Let's consider an analog of the discrete logarithm problem in an elliptic curve group $E(p)$ modulo a prime p. Instead of fixing integers g, X modulo p, let's fix points $P, Q \in E(p)$. In the traditional discrete logarithm problem, the goal is to determine how many times one must *multiply* g by itself to obtain X modulo p (if possible), i.e., what *power* of g equals X:

$$\underbrace{g \cdot g \cdots g}_{x \text{ times}} = g^x \equiv X \bmod p.$$

In our elliptic curve group $E(p)$, we use addition to represent the operation, so an analog would be to ask how many times one must*add* P to itself in order to obtain Q (if possible), i.e., what *multiple of* P equals Q:

$$\underbrace{P + P + \cdots + P}_{m \text{ times}} = nP = Q.$$

Definition 1.1 (Elliptic curve discrete logarithm problem)**.** Given a prime p, fix points P, Q on an elliptic curve $E(p)$. Determine a positive integer n for which

$$nP = Q.$$

Remark 1.2. Note that, in analogy with the traditional discrete logarithm problem where g is not necessarily a primitive root, it is possible that there is no solution to a given elliptic curve discrete logarithm problem; i.e., there exist curves $E(p)$ and points P, Q on $E(p)$ for which $nP \neq Q$ for all integers n. However, we will see that in our applications to cryptography, a solution will always exist, by design.

On the other hand, suppose that a solution n exists, i.e., $nP = Q$. Then if d is the order of P on $E(p)$, we know that $dP = O$, so $(n+dk)P = nP + k(dP) = nP + O = nP$ for every integer $k \geq 0$. Check that all solutions to the discrete logarithm problem have the form $n + dk$ for some $k \in \mathbb{Z}$!

Example 1.3 (Elliptic curve discrete logarithm problem)**.** Let $p = 97$, and fix $P = (74, 13)$ and $Q = (57, 23)$ on the elliptic curve $E(97)$ given by

$$y^2 \equiv x^3 + 11x + 76 \bmod 97.$$

The discrete logarithm problem $nP = Q$ happens to have solution $n = 39$. (Check this using your program for computing multiples of points on an elliptic curve modulo p!).

In fact, $|E(97)| = 107$, which is prime, so given *any* points P, Q on the curve above, there is a solution to the discrete logarithm problem $nP = Q$: By Lagrange's theorem 4.2, the order of any non-identity element of $E(97)$ must have order 107. If any two of the elements

$$P, 2P, 3P, \ldots, 107P = O$$

are equal, say $kP = jP$ for some $1 \leq k < j \leq 107$, then $O = jP + (-kP) = (j-k)P$, and $j - k < 107$, contradicting the fact that the order of P is 107. Hence the list of elements above makes up all elements of $E(97)$, and hence one must be Q, so $nP = Q$ for some $1 \leq n \leq 107$.

Now let $p = 31$, and consider the point $P = (4, 17)$ on the curve $E(31)$ given by

$$y^2 \equiv x^3 + 2x \bmod 31.$$

This curve has order $32 = 2^5$ by Proposition 4.4. In fact, P has order 8, so only the points $Q = P, 2P, 3P, \ldots, 8P = O$ have solutions to the discrete logarithm problem $nP = Q$.

1.2 Elliptic curve Diffie-Hellman key exchange

As the traditional Diffie-Hellman key exchange is designed based on the computational difficulty of solving the discrete logarithm problem when the prime modulus p is very large, the elliptic curve version is relies on the difficulty of solving the elliptic curve version of this problem.

Recall that in the original version of the key exchange, a large prime p and an integer g relatively prime to p (which can be chosen to be a primitive root modulo p) are make public. Two parties, Alice and Bob, want to agree on a shared integer-valued "secret key," that they can then use, e.g., as a tool to send encrypted messages to one another.

Alice chooses a private key $x \in \mathbb{Z}$ and passes $X = g^x \% p$ to Bob, and in turn, Bob chooses secret $y \in \mathbb{Z}$ and passes $Y = g^y \% p$ over a public channel. Their shared key is $k = g^{xy} \% p$, which Alice can find by taking $Y^x \% p$ and Bob can compute as $X^y \% p$, since

$$Y^x \equiv (g^y)^x \equiv g^{xy} \equiv (g^x)^y \equiv X^y \bmod p.$$

Notice how the elliptic-curve version of the Diffie-Hellman key exchange replaces *powers of g* with *multiples of a point P* on an elliptic curve modulo a prime:

Method 1.4 (Elliptic curve Diffie-Hellman key exchange)**.**

Goal: Alice and Bob agree on a shared private key.

Public keys:

- A (large) prime p,
- An elliptic curve $E(p)$ modulo p, and
- A point $P \in E(p)$.

Private keys:

- Alice chooses a secret integer $n > 1$.
- Bob chooses a secret integer $m > 1$.

Process:

1. Alice computes $Q = nP$ and sends it to Bob across a public channel.
2. Bob computes $R = mP$ and sends it to Alice across the public channel.
3. Alice computes nR.
4. Bob computes mQ.

Outcome: The shared private point on $E(p)$ is

$$nR = n(mP) = (nm)P = m(nP) = mQ,$$

and the x-coordinate of this point $(nm)P$ is the **shared secret key**; i.e., $k = x_{(nm)P}$.

Example 1.5 (Elliptic curve Diffie-Hellman key exchange)**.** As in the first part of Example 1.3, let $p = 97$, and $P = (74, 13)$ on the elliptic curve $E(97)$ given by

$$y^2 \equiv x^3 + 11x + 76 \bmod 97.$$

Say Alice chooses the secret key $n = 39$, and Bob's secret key is $m = 52$. Alice computes

$$Q = nP = 39P = (57, 23)$$

(as mentioned in the noted example), and Bob computes

$$R = mP = 52P = (68, 54);$$

they send these values to one another over the public channel.

Then Alice can compute $(nm)P$ as

$$nR = 39 \cdot (68, 54) = (75, 27)$$

and Bob can find it by computing

$$mQ = 52 \cdot (57, 23) = (75, 27).$$

Their points agree! The shared key is then its x-coordinate, 75.

Remark 1.6 (Security of the elliptic curve Diffie-Hellman key exchange)**.** Notice that if an eavesdropper "Eve" can determine n or m by solving either of the elliptic curve discrete logarithm problems (see Definition 1.1)

$$nP = Q \quad \text{or } mP = R$$

on $E(p)$, then she can determine the shared key.

Notice that, although Remark 1.2 shows that if d is the order of P on $E(p)$, then $n' = n + dk$ and $m' = m + dj$ are solutions to the above discrete logarithm problems, respectively, if Eve is able to find one of these solutions and calculate

$$m'Q = (m + dk)(nP) = (nm)P + kn(dP) = (nm)P$$

or similarly, $n'R = (nm)P$ (check that this also holds!), then she has access to the shared secret key. Hence any solution to either elliptic curve discrete logarithms suffices to determine the shared key.

However, in general, it is very difficult to solve the elliptic discrete logarithm problem. The United States National Institute of Standards and Technology has endorsed elliptic curve Diffie–Hellman as a recommended algorithm for key exchange in the National Security Agency's Cryptographic Modernization Program, to be used for both unclassified, and most classified, information.

Remark 1.7 (Shortening the transmission)**.** In fact, for efficiency, Alice and Bob can only send the x-coordinates of the points $Q = nP$ and $R = mP$, respectively, to one another: Suppose that $Q = (x_Q, y_Q)$, and, for instance, Alice only sends Bob the x-coordinate x_Q of Q. Since Q is a multiple of P and $E(p)$ is a group, the coordinates of Q satisfy the elliptic curve equation, say $y^2 \equiv x^3 + ax^2 + b \bmod p$, Bob can compute y_Q as one of the square roots of $x_Q^3 + ax_Q^2 + b$ modulo p. E.g., remember that we derived a formula to determine square roots modulo primes that are congruent to 3 modulo 4, and in general, it is not too difficult to find square roots modulo a prime. Moreover, see Cipolla's algorithm for computations in general (note that unlike our method when $p \equiv 3 \bmod 4$, this is *not* a formula!).

Hence Bob can find $\pm y_Q$, but cannot distinguish which is the actual y-coordinate of Q. If he chooses the wrong one, $-y_Q$, notice that $(x_Q, -y_Q) = -Q$, so he computes

$$m(x_Q, -y_Q) = m(-Q) = -mQ = -m(nP) = -(mn)P = (x_{mnP}, -y_{mnP})$$

which has the same x-coordinate as mnP, the shared secret key!

Finally, we point out that if Bob only sends the x-coordinate of $R = mP$, then by an analogous argument, Alice can determine the x-coordinate of mnP as well.

1.3 Elliptic curve ElGamal cryptosystem

Just as the original ElGamal Cryptosystem can be easily built using the Diffie-Hellman key exchange; the same goes for the elliptic curve versions of these notions.

Suppose that Bob wants to send a secret message to Alice. The message is encoded as a point M on an elliptic curve $E(p)$ modulo p (see Remark 1.11 for more on this).

Method 1.8 (Elliptic curve ElGamal Cryptosystem)**.**

Goal: Bob aims to send a secret message to Alice across a public channel.

Public keys:

- A (large) prime p,
- An elliptic curve $E(p)$ modulo p, and

- A point $P \in E(p)$.

Private keys:

- Alice chooses a secret integer $n > 1$.
- Bob chooses a secret integer $m > 1$.

Process:

1. Bob translates his message into a point M on $E(p)$, in some agreed-upon way.
2. Alice computes $Q = nP$ and sends it to Bob across a public channel.
3. Bob computes
   - $R = mP$, and
   - $S = M + mQ$

   and sends the pair (R, S) to Alice across the public channel.
4. Alice then computes the point $S - nR = S + (-nR)$ on $E(p)$.

Outcome: Alice has recovered the plaintext message M, since

$$S - nR = (M + mQ) - n(mP) = (M + m(nP)) - n(mP) = M.$$

Like the elliptic curve Diffie-Hellman key exchange, to intercept the message M, it suffices to solve either elliptic curve discrete logarithm problem, $Q = nP$ or $R = mP$.

Exercise 1.9. Think about how the elliptic curve ElGamal cryptosystem is analogous to the original one, where our original operation of multiplication is replaced with the addition law on the elliptic curve group!

Example 1.10 (Elliptic curve ElGamal cryptosystem)**.** Again, as in Examples 1.4 and 1.5, let $p = 97$, and $P = (74, 13)$ on the elliptic curve $E(97)$ given by $y^2 \equiv x^3 + 11x + 76 \bmod 97$. Now suppose that Bob wants to send Alice a secret message, translated into the point $M = (7, 60)$ in some agreed-upon way.

Suppose that (as in our previous example of elliptic curve Diffie-Hellman), Alice chooses the private key $n = 39$, and Bob chooses $m = 52$, so Alice computes $Q = nP = 39P = (57, 23)$ and Bob computes $R = mP = 52P = (68, 54)$.

Next, Bob computes

$$S = M + mQ = (7, 60) + 52(57, 23) = (7, 60) + (75, 27) = (81, 13)$$

and sends the pair $(R, S) = ((68, 54), (81, 13))$ to Alice.

Then Alice finds

$$S - nR = (81, 13) - 39(68, 54) = (81, 13) - (75, 27) = (81, 13) + (75, 70) = (7, 60)$$

recovering Bob's secret message M! (Verify these computations using your functions for addition and taking multiples of elliptic curves!)

Remark 1.11 (Turning a message into a point)**.** Unfortunately, there is no "perfect" way to translate a plaintext message (i.e., an integer) into a point on an elliptic curve (e.g., that maximizes efficiency). One way to attempt to do so would be to break the message into "chunks" N less than p. We first hope to find a point whose x-coordinate is N.

If the curve's equation is $y^2 \equiv x^3 + ax + b \bmod p$, then for each chunk N, we compute $N^3 + aN + b$; if this is a square modulo p, let N' be one of its square roots. Then if Bob translates this chunch of the message as the point (N, N') on $E(p)$, Alice will find the message as its x-coordinate. If N does not have a square root, then Bob can append additional digits to the end of N until this new number has a square root modulo p, sending this new point to Alice. When Alice deciphers this message, she will see extra "nonsense" digits at the end of the the message, and will hopefully easily disregard them.

If this doesn't work (i.e., Bob has trouble finding a square modulo p that is small enough to feasibly transmit), he can break the message into smaller or larger chunks and try the same procedure; in this case, there may be more "nonsense" digits at the end of the ciphertext transmission.

Can you think of alternative ways to turn a given message (positive integer) into point(s) on an elliptic curve modulo p?

Virtual Class Notes, Week 12 (April 20 - 24).

2 Lenstra's elliptic curve factoring algorithm

This week, we focus on applying the theory of elliptic curves to attack the security of cryptosystems whose security are based on the difficulty of factoring large integers. *Lenstra's elliptic curve factoring algorithm* is currently the *best algorithm* to find factors that have at most 50 - 60 digits, so, for instance, it is the most efficient algorithm for any factoring problem we have posed in this class! To factor general integers, it is still the *third-best* factoring algorithm in existence.

We stress that this technique is modern, and currently used in practice. Hendrik Lenstra, a Dutch mathematician (who is, even now, likely younger than some of your math or computer science professors!) discovered this technique in 1987. This is even more striking since the mathematical foundations of the cryptosystems we study are very, very old, but there are new and relevant applications of them. Moreover, the mathematics we've build together in our course allows us to execute an in-depth study of these applications!

Lenstra's method is analogous, in many ways, to Pollard's $p-1$ factoring method, and so also to the $p+1$ method. However, there are several features that make Lenstra's method more effective, and we will point these out as we describe the algorithm, and why it works.

2.1 The premise of Lenstra's method: Elliptic curves modulo n

Recall that thus far, when working with elliptic curve groups, we have deliberately only worked over fields, i.e., rings in which all nonzero elements have multiplicative inverses. Why is this? Well, when we add two non-identity points P and Q on a curve, one must compute the slope between them if $P \neq Q$, or the slope of the tangent line to P in the case that $P = Q$ (see the equations we've derived in each case, (3.6.1) and (3.7.1), respectively). In either case, the slope is computed as a "fraction" $M = \frac{r}{s}$. If the denominator s is zero, then we think of this fraction as "infinite," so that the corresponding line is vertical (recall that we are avoiding the degenerate case when both r and s are zero). On the other hand, if s is nonzero, then we compute M as rs^{-1}, which we can only do if s is a unit!

Hence, if we consider an elliptic curve over a ring that is not a field, we do not typically have a well-defined addition law since some slopes necessary to add certain pairs of points may not exist! For instance, consider solutions to the congruence

$$y^2 \equiv x^3 + 1 \bmod 15,$$

so that we are essentially working over the ring $\mathbb{Z}/15\mathbb{Z}$, which is not a field. We can check that $P = (0,4)$ and $Q = (5,6)$ satisfy the above the equation. Toward computing their sum using our current method, we try to compute the slope, modulo 15, as $\frac{6-4}{5-0} \equiv \frac{2}{5}$. However, we know that 5 has no inverse modulo 15, since it is not relatively prime to 15! There is no way to interpret this "slope" as an integer modulo 15, or an element of $\mathbb{Z}/15\mathbb{Z}$.

Similarly, if we seek to double Q to find $2Q$, we try starting to compute the slope of its tangent line, modulo 15, as $\frac{3x^2}{2y}\Big|_{(1,4)} \equiv \frac{3\cdot 25}{12} \equiv \frac{5}{12}$, but again, we run into a similar problem, since $(12,15) = 3 \neq 1$, so 12 also has no inverse modulo 15!

Note that though these problems can occur, sometimes it *is* possible to add some pairs of points on curves "modulo n" for composite n: namely, when the denominator of the slope required is a unit modulo n. As an exercise, check that $2P$ in our example above is well-defined!

2.2 The algorithm

Lenstra's algorithm proceeds by attempting to find *multiples* of a point P on an elliptic curve modulo a composite integer n. Like $2P = P+P$ (if it can be computed), for a positive integer m, we define mP (if it exists, which it definitely does if n is prime) as P added to itself m times:

$$mP = \underbrace{P + P + \cdots + P}_{m \text{ times}}.$$

There are choices to be make when one wants to compute a multiple of P. For instance, to find $4P$, one could first find $2P$, and then double the result to obtain $2(2P)$. On the other hand, one could find $2P$, then $3P = P + 2P$, and then $4P = 3P + P$. Notice that the former method requires only two applications of the group law, and the latter requires three (though when we analyze Lenstra's method, we see why the latter could sometimes be advantageous).

Taking multiples, adding an element to itself a given number of times, is the analog to exponentiation under the operation of multiplication: multiplying an element by itself a given number of times. If desired, we can use an analog of "fast exponentiation"–which we might call "taking fast multiples"–by using base 2 expansions: E.g., if we want to find $681P$ for some point P, one can find that $681 = 1 + 2^3 + 2^5 + 2^7 + 2^9$, so that

$$681P = P + 2^3P + 2^5P + 2^7P + 2^9P$$

One can find $2P, 2^2P = 4P, 2^3P = 8P, \dots, 2^9P$ by successively applying our doubling formula, and then add using associativity.

Method 2.1 (Lenstra's elliptic curve factoring algorithm)**.** Our goal is to find a proper factor of a composite odd integer n.

We start by fixing $a, b \in \mathbb{Z}$, and a solution $P = (x_P, y_P)$ to the equation

$$y^2 \equiv x^3 + ax + b \bmod n.$$

Then start computing a sequence of multiples of P modulo n.

For instance, one can repeatedly double, starting with P:

$$P,\ 2P,\ 2(2P) = 4P,\ 2(4P) = 8P, \dots, 2(2^{k-1}P) = 2^kP, \dots \tag{2.1.1}$$

Alternatively, one can compute successive factorial multiplies of P:

$$P,\ 2P,\ 3(2P) = 6P,\ 4(6P) = 24P, \dots, k\left(((k-1)!)P\right) = (k!)P, \dots \tag{2.1.2}$$

If at any step, the slope $M = \frac{r}{s}$ necessary to compute the next point in the sequence is not well-defined modulo n, i.e., s has no multiplicative inverse modulo n.

Compute $d = (n, s)$, which is greater than 1. If $d \neq n$, then d is a proper factor of n.

We will often refer to the point P above as the "starting point" of choice in Lenstra's algorithm (there is a double meaning of "point" here!).

You may notice that, especially using (2.1.2), that Lenstra's algorithm has some similarities with Pollard's $p - 1$ factoring algorithm, where after fixing some integer a, one iteratively computes the least nonnegative residue of a to $k!$ modulo the composite integer n, for $k = 1, 2, 3, \dots$ and finds certain greatest common divisors. Here, we *multiply*, instead of add, a by itself some number of times. One can think of Lenstra's method as an analog of

Pollard's $p-1$ method (or the $p+1$ method), where, in the background, the group $(\mathbb{Z}/p\mathbb{Z})^\times$ (or $\mathbb{F}_{p^2}$, respectively) is replaced with an elliptic curve group $E(p)$.

In general, it is not necessary to follow one of the two processes of taking multiples as above. For efficiency, it is advantageous to use multiples kP where k is the products of small integers, since finding very large multiples of a point becomes more computationally taxing.

Question 2.2. How do we pick a (random) point on a (random) elliptic curve to start with in Lenstra's algorithm?

One can first pick (random) integers x_P, y_P and a, and after setting $b = (y_P^2 - x_P^3 - ax_P) \,\%\, n$, the desired equation

$$y_P^2 \equiv x_P^3 + ax_P + b \bmod n$$

holds. In fact, b is the only integer modulo n that will satisfy the above equation!

Example 2.3 (Repeated doubling in Lenstra's factoring algorithm)**.** Take $n = 899$, and fix the starting point $P = (10, 11)$ on the curve

$$y^2 \equiv x^3 + 2x \bmod 899.$$

We proceed using the convention (2.1.1) in Lenstra's algorithm, repeatedly doubling points. Using the doubling formula (3.7.1), we compute that

$$\begin{aligned} 2P &\equiv (109, 428) && \bmod 889 \\ 4P &\equiv 2(2P) \equiv (194, 371) && \bmod 889 \\ 8P &\equiv 2(4P) \equiv (806, 31) && \bmod 889 \end{aligned}$$

However, to compute $16P = 2(8P)$, we find that since $\frac{dy}{dx} = \frac{3x^2+2}{2y}$, the slope at $8P = (806, 31)$ should be congruent to $\frac{3\cdot 806^2+2}{2\cdot 31}$ modulo n, which has denominator $2 \cdot 31 \equiv 62 \bmod 899$. Using the Euclidean algorithm, we can find that $(899, 62) = 31$.doubling Hence 31 is a proper factor of $n = 899$! Dividing out, we have that $899 = 29 \cdot 31$.

Example 2.4 (Lenstra's factoring algorithm using factorials)**.** Take $n = 517$, and fix the starting point $P = (3, 6)$ on the curve

$$y^2 \equiv x^3 + 9 \bmod 517.$$

We proceed using the convention (2.1.2) in Lenstra's algorithm, finding consecutive factorial multiples of our starting point. Using equation (3.6.1) and (3.7.1), we compute

$$\begin{aligned} 2P &\equiv (96, 431) && \bmod 517 \\ 6P &\equiv 3(2P) \equiv 2(2P) + 2P \equiv (352, 129) + (96, 431) \equiv (227, 495) && \bmod 517 \end{aligned}$$

Now, to compute $24P = 4(6P)$, we can proceed by computing $2(2(6P))$, or, for instance, $3\cdot(6P)+6P$. Let's do the former. To first find $2(6P)$, we attempt to find the slope of the tangent line to $6P = (227, 495)$ at this point. Since $\frac{dy}{dx} = \frac{3x^2}{2y}$, at our point $\frac{3\cdot 227^2}{2\cdot 495}$ modulo n, which has denominator $2 \cdot 495 = 990 \equiv 473 \bmod 517$. Using the Euclidean algorithm, we find that $(517, 473) = 11$. Hence 11 is a proper factor of $n = 517$. Dividing out, we have that $517 = 11 \cdot 47$.

Remark 2.5 (Effectiveness of Lenstra's method)**.** Suppose that P satisfies $y^2 \equiv x^3 + ax + b \bmod n$. Then it necessarily satisfies the same equation modulo p for every prime factor p of n, i.e., modulo p, P in the corresponding elliptic curve group $E(p)$.

Assume that we obtain the (not useful) factor $d = n$ by attempting to compute the multiple mP of P. Then n is a divisor of the denominator s of the slope $M = \frac{r}{s}$ between between $(k-1)P$ and P, or more generally, between kP and $(m-k)P$ for any $1 \leq k \leq m-1$ (as usual, we mean the slope of the tangent line when the points are equal). Hence every prime factor p of n is a factor of s, which means that in $E(p)$, the line between $(m-1)P$ and P (or kP and $(m-k)P$, respectively) is vertical, and the sum of these two points is the identity O. (Note that it is possible that we are unlucky and $E(p)$ is singular, but if we cannot compute the slope of the tangent line at a point on such a curve, its denominator is still 0 modulo p, i.e., a multiple of p.) Hence the order of P in the group $E(p)$ is m (notice that if it were smaller, our algorithm would have terminated earlier.)

Hence, if Lenstra's algorithm results in the factor $d = n$ after taking the multiple mP, then m must by a multiple of the order of P on $E(p)$ for *every prime factor* p of n! Since the orders of elliptic curves modulo p vary and are in some sense "close to" (e.g., see Hasse's bound, Theorem 4.3), it is unlikely that this will happen.

For instance, consider Example 2.3, where we used repetitive doubling to attempt to find a factor of $n = 899$. Why did we come across a *proper* factor? Well, we now know that $n = 29\cdot 31$ By Proposition 4.4, since $31 \equiv 3 \bmod 4$, the elliptic curve $E(31)$ given by our congruence $y^2 \equiv x^3 + 2x \bmod 31$ has order $32 = 2^5$, so that by Lagrange's theorem (Theorem 4.2) every non-identity point on $E(31)$ has order $2, 2^2, 2^3$, or 2^5. On the other hand, it is a fact that $|E(29)| = 26 = 2 \cdot 13$, so if a non-identity element on $E(29)$ does not have order 2, then it must have order 13 or 26. Therefore, in applying Lenstra's algorithm by repeatedly doubling a point P that has order greater than 2 modulo 17, i.e., computing 2^kP for $k = 2, 3, \dots$ modulo n, then we will arrive at the order of P on $E(31)$ before we come across the order of P on $E(29)$, so we will obtain the proper factor 31. Indeed, this is what happened; in fact, the order of $P = (10, 11)$ on $E(31)$ is the highest possible value, $2^4 = 16$!

In Example **??**, where took factorial multiples of a point P on the curve $y^2 \equiv x^3 + 3 \bmod 517$ to factor $n = 517 = 11 \cdot 47$, Proposition 4.4 tells us that since $11 \equiv 2 \bmod 3$ and $47 \equiv 2 \bmod 3$, $|E(11)| = 12 = 2^2 \cdot 3$ and $E|(47)| = 48 = 2^4 \cdot 3$. Since $12 \mid 4!$, we know that we will obtain a factor in at most 4 steps, and we did!

Notice that by the above discussion, if for all prime divisors p of n, our starting point P has order that is not a multiple of 2, then successive doubling will never yield a factor; the algorithm will not terminate! Hence, though the factorial method is slower (we must do more than simply applying the doubling formula repeatedly), in general it is more likely to be successful.

Question 2.6. If we apply Lenstra's algorithm, and it fails to produce a proper factor of n, must we move on to an alternative factoring method?

There are a few ways that this can happen, and two of these are described in Remark 2.5: The algorithm produces the factor $d = n$, which is not helpful, or the algorithm cannot terminate. Alternatively, we could continue the process of computing multiples of a point for some time, and the algorithm has not reached a point where a factor is produced; perhaps we have used significant computational time/power. This might appear similar to the previous scenario.

One of the features of Lenstra's algorithm that makes it so amazingly effective is that in any of these cases, one can simply choose a different point on a different curve and re-start the algorithm!

Virtual Class Notes, Week 11 (April 13 - 17).

Recall that to put a group structure on an elliptic curve, we needed to add an extra point O "at infinity." Before we start to go into more depth on elliptic curve groups, we make a few more refinements to our current definition.

3 A revised definition of an elliptic curve group

To start, we will only work over fields F in which $2 \neq 0$ and $3 \neq 0$ in F, where 2 denotes the element $1+1$, and 3 is $1+1+1$. This means that, among the fields that we've used in this class, we will avoid $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$, as well as $\mathbb{F}_4$ and $\mathbb{F}_9$. On the other hand, the fields $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{F}_{p^2}$ are OK if $p \geq 5$, as are $\mathbb{R}$, $\mathbb{Q}$, and $\mathbb{C}$. In our immediate discussion, we will point out instances where we are using the fact that $2, 3 \neq 0$ in our base field F.

3.1 Singular curves

Next, we observe that there is a gap in our current definition of the elliptic curve group law! Recall that if a line is tangent to a point P on an elliptic curve, and passes through another point R, then $2P = -R$ in the group–we consider P to have multiplicity 2, so $P+P+R = O$ (see the third and fourth graphs in Figure 6.5.1–in the latter, $R = O$). This allows us to add any point P on an elliptic curve to itself, i.e., compute $2P$, using its tangent line–unless the tangent line does not exist! This happens if and only if we cannot define the slope of the elliptic curve at P, that is, the curve is not differentiable at P.

Notice that if an elliptic curve is given by $y^2 = f(x)$, then implicit differentiation allows us to find that $2y\frac{dy}{dx} = f'(x)$. If $y \neq 0$. the derivative necessarily exists and equals $\frac{dy}{dx} = \frac{f'(x)}{2y}$ (note that f is differentiable everywhere since it is a polynomial, and that we are also using our assumption that $2 \neq 0$ in F here in dividing by 2!). However, if $y = 0$, two things can happen. Recall from calculus that if $f'(x) \neq 0$, then we can think of the formula $\frac{f'(x)}{2y}$ as being "infinite"; more precisely, as $y \to 0$, the fraction approaches $\pm\infty$. The fourth graph in Figure 6.5.1 illustrates an example where $y = 0$ and the tangent line to the point intersecting the elliptic curve has "infinite" slope, meaning that it is vertical.

The final case is when both $y = 0$ and $f'(x) = 0$, in which case the fraction $\frac{f'(x)}{2y}$ has the form $\frac{0}{0}$, and the derivative $\frac{dy}{dx}$ does not exist. In this case, we have not defined the point $P + P = 2P$, since there is no tangent line to the elliptic curve at P. Therefore, we have not given a well-defined addition law on elliptic curves with this property, which we call *singular*:

Definition 3.1. We call an elliptic curve **singular** if it has a point where the derivative $\frac{dy}{dx}$ does not exist (and is not infinite); otherwise, it is **nonsingular**, or **smooth**.

Our next goal is the understand when an elliptic curve is singular, so that we can quickly and easily identify the smooth cases in which our group law makes sense. We will apply the following fact, and use the following definition:

Remark 3.2. Given a polynomial $g(x)$ with coefficients in a field F, an element $a \in F$ is a root of $g(x)$ if and only if $g(x) = (x-a)h(x)$ for some polynomial $h(x)$ with coefficients in F.

Definition 3.3. Fix a polynomial $g(x)$ with coefficients in a field F, and an element $a \in F$. We say that $g(x)$ has an **n-th root** at $x = a$ if $g(x) = (x-a)^n h(x)$ for some polynomial $h(x)$ with coefficients in F. We often say $g(x)$ has a **double root** at $x = a$ if $n = 2$, and a **triple root** if $n = 3$.

In particular, notice that a polynomial with a triple root at $x = a$ also has a double root there.

The following lemma gets us closer to giving a simple characterization of when when an elliptic curve is singular. Notice that although the limit definition of a derivative does not make sense over an arbitrary field (e.g., think about $\mathbb{Z}/p\mathbb{Z}$), we can still formally define derivatives of polynomial using the power rule $\frac{d}{dx}x^n = nx^{n-1}$ for $n \geq 0$. Note that in this setting of polynomials, the product and chain rules follow from this.

Lemma 3.4. *Let $g(x)$ be a polynomial over a field F, with root $a \in F$. Then $g(x)$ has a double root at $x = a$ if and only if $x = a$ is also a root of the derivative $g'(x)$.*

Proof. Suppose that m is the maximal number of factors $x - a$ in $g(x)$, so that $g(x) = (x-a)^m h(x)$, where $m \geq 1$ and $h(a) \neq 0$. Then by the product rule,

$$g'(x) = m(x-a)^{m-1}h(x) + (x-a)^m h'(x).$$

First assume that $g(x)$ has a double root at $x - a$, so that $m \geq 2$. Then $m - 1 \geq 1$ and $0^m = 0^{m-1} = 0$, so $g'(a) = m \cdot 0^{m-1} \cdot h(a) + 0^m \cdot h'(x) = 0$. On the other hand, if $m = 1$, then $g(x) = (x-a)h(x)$, so $g'(x) = h(x) + (x-a)h'(x)$, and $g'(a) = h(a) + (a-a) \cdot h'(x) = h(a) \neq 0$ by our assumption on $h(x)$. □

Since the derivative $\frac{dy}{dx}$ at a point on an elliptic curve (x_0, y_0) does not exist if and only if $y_0 = 0$ and $f'(x_0) = 0$, after applying Lemma 3.4, we have proved the following proposition!

Proposition 3.5. *An elliptic curve $y^2 = f(x)$ is singular if and only if it contains a point $P = (x_0, 0)$ such that x_0 is a double root of $f(x)$.*

From now on, we will only consider elliptic curve groups over curves that are nonsingular, so that the group law is well-defined.

3.2 Shifting to simplify the elliptic curve equation

Our final alteration is out of convenience, not necessity. Notice that if we shift an elliptic curve $y^2 = f(x)$ horizontally, to the right by σ, we get another one, $y^2 = f(x - \sigma)$. This new elliptic curve has the same shape as the original one (just shifted), and a line passes through three points on the original curve if and only if a line (with shifted x-coordinates) passes through the corresponding points on the other. Hence the group structures on the two curves are the same, at least after renaming the points.

We use this fact to simplify our equation for an elliptic curve. Recall that up until this point, we considered curves of the form $y^2 = f(x)$, where $f(x) = x^3 + ax^2 + bx + c$ and $a, b, c \in F$. Under our assumption that $3 \neq 0$ in F, $a/3$ (i.e., $a \cdot 3^{-1}$) is a well-defined element of F. Notice that if we shift the curve to the right by $a/3$, we obtain the curve

$$\begin{aligned} y^2 = f\left(x - \frac{a}{3}\right) &= \left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c \\ &= \left(x^3 - 3x^2 \cdot \frac{a}{3} + 3x \cdot \frac{a^2}{9} - \frac{a^3}{27}\right) + a\left(x^2 - \frac{2ax}{3} + \frac{a^2}{9}\right) + bx - \frac{ab}{3} + c \\ &= x^3 + \left(b - \frac{a^2}{3}\right)x + \left(\frac{2a^3}{27} - \frac{ab}{3} + c\right) \end{aligned}$$

In particular, the the coefficient of x^2 is zero! Hence, since we are concerned with the group structure of an elliptic curve, we can restrict ourselves to studying equations of the form $y^2 = f(x)$, where $f(x)$ is a monic cubic polynomial in x whose coefficient of x^2 is zero. Notice that in our revised definition, we reuse "a" and "b" to mean coefficients of different terms than in our original definition.

We now finalize our definition of an elliptic curve group.

Definition 3.6 (Elliptic curve group)**.** Fix a field F for which $2, 3 \neq 0$, and elements $a, b, c \in F$ for which

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

is nonsingular. Let E denote all points (x, y) satisfying the above equation, along with a point O. Then $(E, +)$ is a group under the following axioms:

1. The identity is O, so that $P + O = P = O + P$ for all $P \in E$.

2. Given $P = (x_P, y_P) \in E$, its inverse is $-P = (x_P, -y_P)$.

3. Given $P = (x_P, y_P), Q = (x_Q, y_Q) \in E$ for which $x_P \neq x_Q$,

$$P + Q = -R$$

where $R = (x_R, y_R)$ is the point on the line L through P and Q, which is considered to be the tangent line to P if $P = Q$, with coordinates

$$x_R = M^2 - x_P - x_Q \quad \text{and} \quad y_R = M(x_R - x_P) + y_P \tag{3.6.1}$$

where M is the slope of L.

Note that we found the coordinates of R using our derivation (6.3.3), but now with the coefficient of x^2 in $f(x)$, previously called “a,” set to 0 in our new notation.

Soon, we will often want to double points on elliptic curves. Applying $Q = P$ to (3.6.1) so that $M = \frac{dy}{dx}|_P$, we obtain the following formula:

Method 3.7 (Doubling formula for points on elliptic curves)**.** If $P = (x_P, y_P)$ is a point on an elliptic curve, let $M = \frac{dy}{dx}\big|_P$. Then by (3.6.1), $2P = -R$, where $R = (x_R, y_R)$, where $x_R = M^2 - 2x_P$ and $y_R = M \cdot (x_R - x_P) + y_P$. Hence $2P = -R = (x_{2P}, y_{2P})$, where

$$x_{2P} = M^2 - 2x_P \quad \text{and} \quad y_{2P} = M(x_P - x_{2P}) - y_P \tag{3.7.1}$$

Whenever we want to use the group law on an elliptic curve, we need to know that the curve is not singular, i.e., the right-hand side of the equation doesn’t have a double or triple root. A useful tool to do this is the *discriminant.*

Given a *quadratic* (rather than cubic, our focus here) polynomial $ax^2 + bx + c$ in variable x, you might recall that its discriminate is the value $b^2 - 4ac$. Notice that the roots of a quadratic equation over the real numbers (which can be complex), $\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$, coincide and equal $-\frac{b}{2a}$ if and only if its discriminant is 0. Hence the vanishing of the discriminant characterizes whether the polynomial has a double root. Notice that, up to a unit, the discriminant is the square of the difference of its roots.

If a monic cubic equation has roots x_1, x_2, and x_3, then its discriminant is (a constant multiple of) the polynomial $((x_1 - x_2)(x_1 - x_3)(x_2 - x_3))^2$. The square ensures that there is no sign ambiguity, and it is clear that this number vanishes if and only if one term in the product equals 0–the cubic has a double (or triple) root. As a (tedious) exercise, you can check, by expanding and setting coefficients equal, that using our equational conventions, the value above is (a constant multiple of) the expression in the description below:

Remark 3.8 (Discriminant of an elliptic curve)**.** Given elements a, b in a field F for which $2, 3 \neq 0$, the **discriminant** of $x^3 + ax + b$ is

$$\Delta = 4a^3 + 27b^2.$$

Moreover, $\Delta = 0$ if and only if the elliptic curve $y^2 = x^3 + ax + b$ is singular, i.e., the right-hand side has a double or triple root.

Notice that if we were working over a field where $2 = 0$ or $3 = 0$, the vanishing of the discriminant would simply be equivalent to the vanishing of b, or of a, respectively.

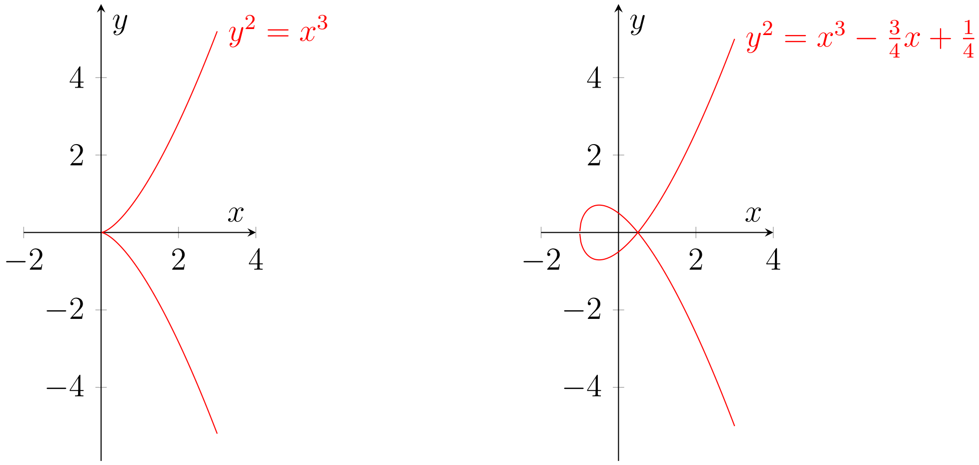


Figure 3.8.1: Some singular elliptic curves over $\mathbb{R}$

Example 3.9 (Singular elliptic curves)**.** Notice that in both graphs appearing in Figure 3.8.1, there is a point where the curve is not smooth. For $y^2 = x^3$, there is a cusp at the origin, and for $y^2 = x^3 - \frac{3}{4}x + \frac{1}{4}$, there is a point where the tangent line is not well-defined–it looks as if there are "two" possible tangent lines there. Indeed, x^3 has a triple root at $x - 0$, and $x^3 - \frac{3}{4}x + \frac{1}{4} = (x+1)\left(x - \frac{1}{2}\right)^2$ has a double root at $x = \frac{1}{2}$.

We can also verify that these curves are singular via the discriminant: For the first curve, $a = b = 0$, so $\Delta = 0 + 0 = 0$. For the second, $a = -3/4$ and $b = 1/4$, so $\Delta = 4 \cdot (-3/4)^3 + 27 \cdot (1/4)^2 = -3^3/4^2 + 3^3/4^2 = 0$ as well. On the other hand, Example 6.7 with equation $y^2 = x^3 + 1$ has $a = 0$ and $b = 1$, and appears smooth. Its discriminant is $27 = 3^3$, which is nonzero as an element of $\mathbb{R}$ (and also in all other fields where $3 \neq 0$).

In fact, we can remove the point on a singular elliptic curve whose derivative does not exist (think about why there can only be one!), and use the remaining points to define a

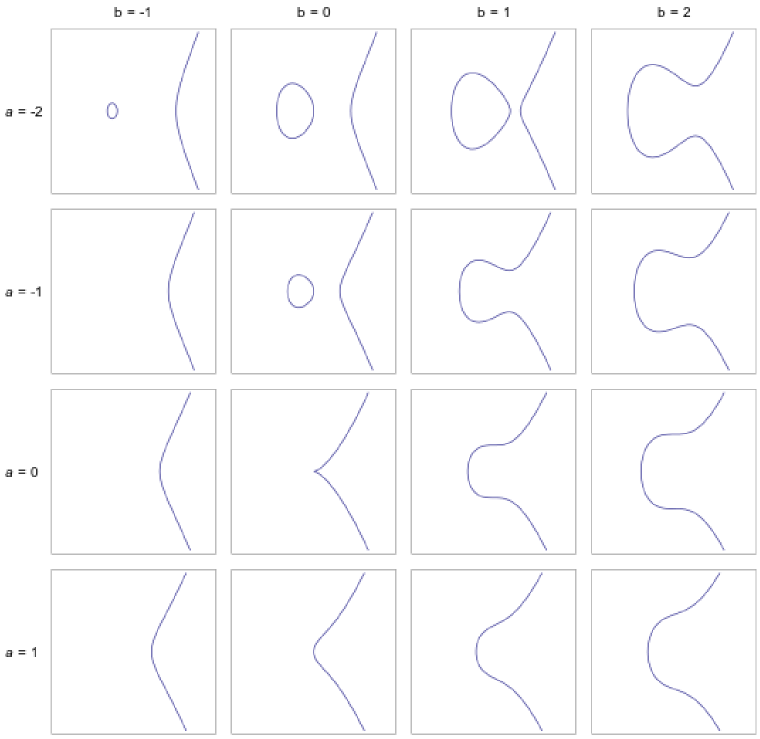


Figure 3.9.1: Elliptic curves over $\mathbb{R}$ as a, b vary

group law. However, the groups obtained in these cases can be "degenerate," and have the same structure as some well-understood groups. As specified in our final definition above, we will stick to the nonsingular case when discussing elliptic curve groups.

Indeed, the graphs of nonsingular elliptic curves over the real numbers appear smooth. Check out Figure 3.9.1, which illustrates how elliptic curves can change as the coefficients change; notice the singular case $a = b = 0$ that also appeared in Figure 3.8.1.

4 Elliptic curves modulo a prime

Since cryptography is a discrete science requiring finite data, our applications will involve elliptic curves over *finite fields*. Our focus is when the base field F is $\mathbb{Z}/p\mathbb{Z}$, where p is a prime and $p \neq 2, 3$. These elliptic curves are often denoted $E(p)$ to clarify the prime, and besides infinite point O, points on $E(p)$ can be considered as coordinate pairs (x, y) of integers modulo p satisfying an equation of the form

$$y^2 \equiv x^3 + ax + b \bmod p$$

where a and b are fixed integers. We often refer to an elliptic curve $E(p)$ as an **elliptic curve modulo** p.

Consider the familiar curve $y^2 = x^3 + 1$, which we determined in Example 3.9 is nondegenerate over any field we consider. The point $P = (2, 2)$ is in the elliptic curve group $E(5)$ defined by this curve over $\mathbb{Z}/5\mathbb{Z}$ since $4 \equiv 8 + 1 \bmod 5$.

To get oriented with calculations on $E(5)$, let's begin to find multiples of P. We first find that $\frac{dy}{dx} \equiv \frac{3x^2}{2y}$, which at $P = (2,2)$ equals $\frac{12}{4} \equiv \frac{2}{4}$ mod 5, and since $4^{-1} \equiv 4$ mod 5, this equals $2 \cdot 4 \equiv 3$ mod 5. Now applying the doubling formula (3.7.1) The x-coordinate of $2P$ is $3^2 - 2 \cdot 2 \equiv 0$ mod 5, and the y-coordinate is $3(2-0) - 2 = 4$. Then $2P = (0,4)$.

Now, let's find $3P = P + 2P$. The slope between $P = (2,2)$ and $2P = (0,4)$ is $\frac{4-2}{0-2} \equiv \frac{2}{3} \equiv 2 \cdot 2 = 4$ mod 5. The addition formula (3.6.1) yields $x_{3P} \equiv 4^2 - x_P - x_{2P} \equiv x_{2P} = 1 - 2 - 0 \equiv 4$ mod 5 and $y_{3P} = 4(x_P - x_{3P}) - y_P = 4(2-4) - 2 \equiv 0$ mod 5. Hence $3P = (4,0)$. Try finding $4P$ by writing this as $P + 3P$, and then as $2 \cdot (2P)$, and verifying that you get the same answer!

It is useful to extend our definition of the order of units in $\mathbb{Z}/p\mathbb{Z}$ to elements of an arbitrary group (though not every element of a group has a well-defined order).

Definition 4.1. The **order of an element** a in a group G, if is the smallest positive integer n, if it exists, for which $a^n = e$, or if the operation is addition,

$$na = \underbrace{a + a + \cdots + a}_{n \text{ times}} = e.$$

If no such $n \geq 1$ satisfying this condition exists, the order of a is not defined.

The **order of a group** G, denoted $|G|$, is the number of elements in the group. In particular, the order of an infinite group is infinity.

The order of i in the group $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$ is 4 since $i^2 = -1, i^3 = i^2 \cdot i = -i$, and $i^4 = (i^2)^2 = (-1)^2 = 1$. The order of 2 in $\mathbb{Z}$ (with operation addition) does not exist, because $n \cdot 2 \neq 0$ for all $n \geq 1$. The orders of both groups, $\mathbb{C}$ and $\mathbb{R}$, are infinity.

The order of a point P on an elliptic curve is the smallest integer $n \geq 1$ for which $nP = O$. For instance, our work just before Definition 4.1 shows that the order of $P = (2,2)$ on $E(5)$ is at least 4, since none of $P, 2P, 3P$ equal O.

Suppose that G is a finite group under multiplication, and $|G| = n$. Then for $a \in G$, two of the $n+1$ elements

$$g^0 = 1, g, g^2, \ldots, g^n$$

must coincide. Suppose that $g^i = g^j$ for $0 \leq j < i \leq n$. Then $g^{i-j} = g^i(g^j)^{-1} = g^j(g^j)^{-1} = 1$. Hence every element of a finite group has a well-defined order.

The orders of elements of a finite group, and of the group, are related. Recall that we used a specialized version of Lagrange's theorem to study the termination of the $p+1$ factoring algorithm; here is a more general statement.

Theorem 4.2 (Lagrange's theorem)**.** *The order of an element in a finite group divides the order of the group.*

Recall that we have already proved Lagrange's theorem for $(\mathbb{Z}/p\mathbb{Z})^\times$, showing that the order of any element a must be a divisor of $p-1$. The general proof is analogous to our proof in this special case.

There are at most p^2+1 points on an elliptic curve group modulo p, as there are p choices each for the x- and y- coordinates of a point, and we also have the identity O in the group. In particular, for any curve and any prime p, $|T(p)| \leq p^2+1 < \infty$.

How can we determine the number of points on $E(p)$? For instance, can we find the order of $E(5)$ without actually checking whether all $5 \cdot 5 = 25$ possible pairs satisfy the congruent equation $y^2 \equiv x^3+1 \bmod 5$? Consider the table below:

$x \ \% \ 5$	0	1	2	3	4
$x^3 \ \% \ 5$	0	1	3	2	4
$(x^3+1) \ \% \ 5$	1	2	4	3	0

In particular, all least nonnegative residues of integers modulo 5 appear exactly once in the third row! The only squares modulo 5 are $0^2 \equiv 0$, $1^2 \equiv 4^2 \equiv 1$, and $2^2 \equiv 3^2 \equiv 4$. Since we are looking for solutions $y^2 \equiv x^3+1$, *only x-values for which the third rows are squares modulo* 5 can contribute points. Each will contribute two points, the square roots modulo the 5, unless the value is 0, which only has one square root, 0, modulo 5.

$x \ \% \ 5$	0	1	2	3	4
$x^3 \ \% \ 5$	0	1	3	2	4
$(x^3+1) \ \% \ 5$	1	2	4	3	0
$y \ \% \ 5$	± 1		± 2		0

We conclude that the points (x, y) on the curve are the following:

$$(0,1),(0,4),(2,2),(2,3),(4,0),O.$$

Adding the identity O, the order of $E(5)$ is 6.

In contrast, there are 12 points on the elliptic $E(7)$ given by the same congruence equation, i.e., $y^2 \equiv x^3+1 \bmod 7$. The elements are, in general, different: $(2,2)$ is on $E(5)$ but not on $E(7)$, and vice versa for $(2,4)$.

In general, it can be difficult to determine the number of points on an elliptic curve. Though we do not have all the tools to prove it, the following bound on the order of elliptic curves, due to Hasse, can be very useful:

Theorem 4.3 (Hasse's bound)**.** *Given an elliptic curve $E(p)$ modulo p, its order satisfies:*

$$p - 2\sqrt{p} + 1 \leq |E(p)| \leq p + 2\sqrt{p} + 1$$

Let's apply Hasse's bound to the curve $E(5)$ that we've been working with. Since $2 \cdot \sqrt{5} \approx 4.8$, we have that $5 + 2\sqrt{5} + 1 \approx 10.5$ and $5 - 2\sqrt{5} + 1 \approx 1.5$. Since the order is an integer, $2 \leq |T(5)| \leq 10$, a pretty broad range. Recall that we determined that in our case, $|T(5)| = 6$.

Notice that Hasse's bound does not depend on the field elements a, b that define the elliptic curve congruence $y^2 \equiv x^3 + ax + b \bmod p$, and it cannot precisely determine the order of the elliptic curve group. In some cases, after focusing on elliptic curves modulo p with specific formulas, we can determine the exact order of an elliptic curve modulo p.

Proposition 4.4. *Fix a prime p and an elliptic curve $E(p)$ modulo p satisfying one of the following conditions:*

(1) $p \equiv 2 \bmod 3$, and $E(p)$ is defined by the equation $y^2 \equiv x^3 + b \bmod p$, where $p \nmid b$.

(2) $p \equiv 3 \bmod 4$, and $E(p)$ is defined by the equation $y^2 \equiv x^3 + ax \bmod p$, where $p \nmid a$.

Then $|E(p)| = p + 1$.

Notice that in both cases, the elliptic curves are nonsingular: For (1), the discriminant equals $27b^2$, which is nonzero modulo p since $p \neq 3$ and $b \not\equiv 0 \bmod p$. The discriminant for (2) is $4a^2$, which is again nonzero modulo p since $p \neq 2$ and $a \not\equiv 0 \bmod p$.

Please refer to Propositions 48 and 49 in Savin for the proofs of these statements. They are quite interesting, and rely on properties of squares modulo primes! Notice that our example $y^2 \equiv x^3 + 1$ satisfies (1) in Proposition 4.4, and $|E(5)|$ satisfies the conclusion, $|E(5)| = 5 + 1 = 6$. In fact, our logic above, in counting points on $E(5)$, extends to a general proof of (1) for $E(p)$ of the given form. The proof of (2) is broken into two cases, based on whether $-b$ is a square modulo p.

Virtual Class Notes, Week 10 (April 6 - 10).

5 The quadratic sieve factoring algorithm

Along with Pollard's $p-1$ factoring algorithm and the $p+1$ factoring algorithm, we present another factoring algorithm called the *quadratic sieve method*. The quadratic sieve factoring algorithm relies on the fact that if an integer has a square root modulo a composite odd integer n, then the Chinese remainder theorem guarantees that it has at least two pairs of square roots. (E.g., think about our application of this fact when "flipping coins over the telephone.") In other words, it is possible that $a \not\equiv \pm b \bmod n$, but $a^2 \equiv b^2 \bmod n$. Each step of the proof of the following lemma is likely familiar to you at this point:

Lemma 5.1. *Given a composite integer n, suppose that a and b are integers such that $a \not\equiv b \bmod n$ and $a \not\equiv -b \bmod n$, but*

$$a^2 \equiv b^2 \bmod n.$$

Then $(a-b, n)$ and $(a+b, n)$ are proper factors of n.

Proof. Since $a^2 \equiv b^2 \bmod n$, we know that n is a divisor of $a^2 - b^2 = (a-b)(a+b)$. However, $n \nmid (a-b)$ and $n \nmid (a+b)$; otherwise $a \equiv b \bmod n$ or $a \equiv -b \bmod n$, respectively. Then at least one prime factor of n divides $a-b$, and a different one must divide $a+b$, so n is not relatively prime to $a-b$ nor $a+b$. Moreover, neither $(n, a-b)$ nor $(n, a+b)$ equals n since n is not a divisor of $a-b$ nor $a+b$, so both of these greatest common divisors are proper factors of n. □

We can apply Lemma 5.1 to factor a composite integer n, as long as we can find two integers whose squares are congruent modulo n, but they are not congruent, nor negatives of one another, modulo n. For instance, it is clear that $a = 4$ is a solution to

$$x^2 \equiv 16 \bmod 45,$$

but $b = 14$ is also a solution since $14^2 = 196 \equiv 16 \bmod 45$. Hence $4^2 \equiv 14^2 \bmod 45$, so that 45 divides $b^2 - a^2 = 14^2 - 4^2 = (14-4)(14+4) = 10 \cdot 18$. We find that $(45, 10) = 5$ and $(45, 18) = 9$, which, indeed, are proper factors of 45!

The concept just described can be applied to factor a composite integer; before describing this algorithm, consider the following observation about squares and prime factorizations. As an exercise, justify it.

Remark 5.2 (Unique factorization of squares)**.** Given an integer x, suppose that its prime factorization is $p_1^{e_1} p_2^{e_2} \cdots p_t^{e_t}$, where the p_i are distinct primes, and the e_i are nonnegative integers. Then n is a square if and only if all the e_i are even.

Now we are ready to introduce a factoring algorithm relying on the principles discussed thus far.

Method 5.3 (Quadratic sieve factoring algorithm.)**.** Our goal is to find a proper factor of a composite odd integer n. Let m denote the smallest integer greater than $\sqrt{n}$. Start computing

$$\begin{aligned} x_0 &= m^2 \ \% \ n \\ x_1 &= (m+1)^2 \ \% \ n \\ x_2 &= (m+2)^2 \ \% \ n \\ &\ \vdots \\ x_k &= (m+k)^2 \ \% \ n \\ &\ \vdots \end{aligned}$$

and at each step, find the prime factorization of x_k. Continue until one of the following conditions holds:

(1) All exponents in the prime factorization of x_k are even, so that $x_k = X^2$ for some integer X. In this case, let $Y = (m+k)^2$.

(2) All exponents in the prime factorization of some *product* of the x_i computed so far are all even; i.e., this product equals X^2 for some integer X. In this case, let Y denote the product of the $m+i$ for which x_i is in this original product.

In either case, compute $(X-Y,n)$ and $(X+Y,n)$ of n via the Euclidean algorithm. If not 1 nor n, each is a proper factor of n.

Notice that if m is the smallest integer greater than $\sqrt{n}$, then m^2 will be larger than n, but not by much. Hence its least nonnegative residue should be fairly small, so that its prime factors should not be very large, and finding its prime factorization should not be computationally difficult. Similarly, its should be possible to factor the first few of $(m+1)\ \%\ n, (m+2)\ \%\ n, \ldots$ We hope that the algorithm terminates in relatively few steps, giving a proper factor of n.

Example 5.4 (Quadratic sieve, Case 1)**.** Consider $n = 4183$. Since $64 < \sqrt{n} < 65$, we have that $m = 65$. We begin finding x_k for $k = 0, 1, 2, \ldots$, and at each step, finding its prime factorization:

$$\begin{array}{llll}
x_0 \equiv 65^2 \equiv 4225 \equiv 42 & \bmod 4183 & & 42 = 2 \cdot 3 \cdot 7 \\
x_1 \equiv 66^2 \equiv 4356 \equiv 173 & \bmod 4183 & & 173 \text{ is prime} \\
x_2 \equiv 67^2 \equiv 4489 \equiv 306 & \bmod 4183 & & 306 = 2 \cdot 3^2 \cdot 17 \\
x_3 \equiv 68^2 \equiv 4624 \equiv 441 & \bmod 4183 & & 441 = 3^2 \cdot 7^2
\end{array}$$

Since $441 = (3 \cdot 7)^2 = 21^2$, we have that $68^2 \equiv 21^2 \bmod 4183$. Since $68 - 21 = 47$ and $68 + 21 = 89$, we use the Euclidean algorithm to find their greatest common divisors with n; $(47, 4183) = 47$ and $(89, 4183) = 89$. Both are proper factors of n. In fact, you can check that both are prime, and $n = 47 \cdot 89$.

Example 5.5 (Quadratic sieve, Case 2)**.** Let $n = 4033$, so that $63 < \sqrt{n} < 64$, we have that $m = 64$. Again, we begin finding x_k for $k = 0, 1, 2, \ldots$, and at each step, finding its prime factorization:

$$\begin{array}{llll}
x_0 \equiv 64^2 \equiv 4096 \equiv 63 & \bmod 4033 & & 63 = 3^2 \cdot 7 \\
x_1 \equiv 65^2 \equiv 4225 \equiv 192 & \bmod 4033 & & 192 = \boxed{2^6 \cdot 3} \\
x_2 \equiv 66^2 \equiv 4356 \equiv 323 & \bmod 4033 & & 323 = 17 \cdot 19 \\
x_3 \equiv 67^2 \equiv 4489 \equiv 456 & \bmod 4033 & & 456 = 2^3 \cdot 3 \cdot 19 \\
x_4 \equiv 68^2 \equiv 4624 \equiv 591 & \bmod 4033 & & 591 = 3 \cdot 197 \\
x_5 \equiv 69^2 \equiv 4761 \equiv 728 & \bmod 4033 & & 728 = 2^3 \cdot 7 \cdot 13 \\
x_6 \equiv 70^2 \equiv 4900 \equiv 867 & \bmod 4033 & & 867 = \boxed{3 \cdot 17^2}
\end{array}$$

Though none of the least nonnegative residues is a square (i.e., has only even powers of primes), we see that the product

$$x_1 \cdot x_6 = (2^6 \cdot 3)(3 \cdot 17^2) = 2^6 \cdot 3^2 \cdot 17^2 = (2^3 \cdot 3 \cdot 17)^2 = 408^2$$

is a square! Since $x_1 \cdot x_6 = 65 \cdot 70 = 4550$, so that $4550 - 408 = 4142$ and $4550 + 408 = 4958$. We find that $(4142, 4033) = 109$ and $(4985, 4033) = 37$ are proper factors of n! You can check that each factor we found is prime, and their product is n.

6 Introduction to elliptic curves.

Now, we transition to a completely new topic, studying the solutions to certain polynomial equations. In fact, the points satisfying these equations form a group (after a small modification), and like the groups of units of $\mathbb{Z}/p\mathbb{Z}$ and the circle group $T(p)$, for p prime, these groups can be applied to cryptography in interesting and powerful ways. In fact, some of the more recent cryptosystems in use rely on the group structure of a so-called *elliptic curve*.

Recall that a *field* is a commutative ring in which all nonzero elements are units; $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields, as are $\mathbb{Z}/p\mathbb{Z}$ and $\mathbb{F}_{p^2}$, for p prime.

Preliminary definition 6.1 (Elliptic curve)**.** Given a field F and $a, b, c \in F$, an **elliptic curve** E over F is the set of all points (x, y) satisfying

$$y^2 = x^3 + ax^2 + bx + c.$$

Example 6.2 (Elliptic curve)**.** Let E denote the elliptic curve over $\mathbb{R}$ given by

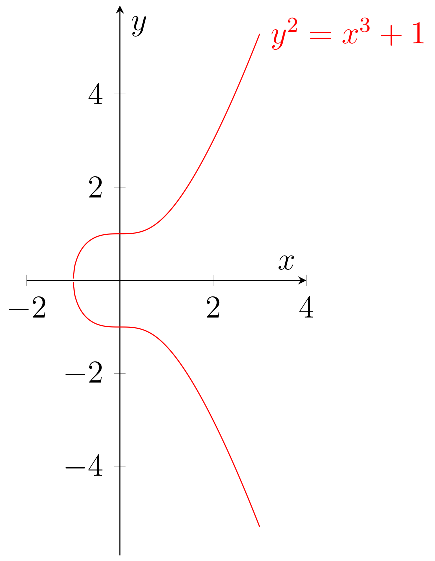
$$y^2 = x^3 + 1.$$

Since $y^2 \geq 0$, we have that $x^3 + 1 \geq 0$ for all points (x, y) on E, so $x^3 \geq -1$, i.e., $x \geq -1$.

The graph is symmetric about the x-axis since $(-y)^2 = x^3 + 1$; i.e., if $(x, y) \in E$, then $(x, -y) \in E$ as well. The points with positive y-coordinates make up a "positive branch," and those with negative y-coordinate form the "negative branch."

Moreover, implicitly differentiating the equation, we find that $2y\frac{dy}{dx} = 3x^2$, so $\frac{dy}{dx} = \frac{3x^2}{2y}$. This derivative be thought of as infinite if the denominator vanishes, i.e., $y = 0$, in which case $0 = 0^2 = x^3 + 1$, so $x = -1$. We see that the tangent line to $(-1, 0)$ is indeed vertical. We also notice that if $y \geq 0$, then $\frac{dy}{dx} \geq 0$, i.e., the positive branch is increasing, and if if $y \leq 0$, then $\frac{dy}{dx} \leq 0$, so the negative branch is decreasing (which would also come for free from the symmetry we've noticed).

In fact, the graph has the following shape; as an exercise, try finding its inflection points!



The group law of an elliptic curve relies on the fact that **a line L typically intersects an elliptic curve E at exactly three points (but not always)**. To show this let's fix the following conventions for our immediate discussion:

Setup 6.3. Let E be the elliptic curve over a field F given by

$$y^2 = f(x), \text{ where } f(x) = x^3 + ax^2 + bx + c.$$

Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be points on the intersection of E with a line L.

Notice that "most" lines are not vertical (e.g., zero is only one of many possibly slopes for a line among all possible values). Let's first address this exceptional case.

Assume that L is vertical, so that $x_1 = x_2 = \gamma$ for some $\gamma \in F$. Then L has the simple equation $x = \gamma$, and any point on $E \cap L$ satisfies

$$y^2 = f(\gamma) = \gamma^3 + a\gamma^2 + b\gamma + c.$$

Notice that $f(\gamma) \in F$, so that there are only two possible y-values of points on $E \cap L$, $\pm\sqrt{f(x)}$. These must be the y-coordinates of P and Q. Hence in this case, after possibly renaming the points, the only points on $E \cap L$ are P and Q, and they have the form $P = (\gamma, f(\gamma))$ and $Q = (\gamma, -f(\gamma))$. Moreover, if $f(\gamma) = 0$, then $P = Q$. We have discovered the following.

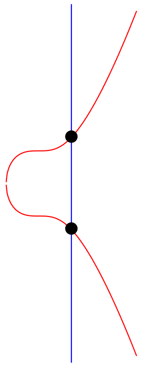
Case 1 (L vertical). Suppose that L is vertical and $P = (x_1, y_1)$ lies on $E \cap L$. Then

(a) If $y_1 \neq 0$, there are exactly two points on $E \cap L$, P and $Q = (x_1, -y_1)$.

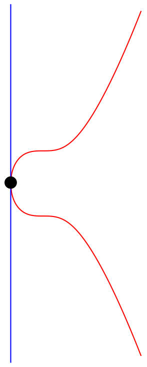
(b) If $y_1 = 0$, $P = Q$, so that there is only one point $P = (x_1, 0)$ on $E \cap L$.

Now assume that L is not vertical; i.e., $x_1 \neq x_2$. Then the slope of L equals

$$M = \frac{y_2 - y_1}{x_2 - x_1}$$



An example of Case 1(a)



An example of Case 1(b)

where this quotient denotes $(y_2 - y_1) \cdot (x_2 - x_1)^{-1}$ in the field F. (Note that we will continue to use fractional notation to mean the product of a numerator with the inverse of a denominator.)

In this case, let B denote the y-intercept of the L (which exists since L is not vertical), so that L has equation $y = Mx + B$. Then these points (x, y) in the intersection of $E \cap L$ are those that satisfy

$$(Mx + B)^2 = y^2 = f(x) = x^3 + ax^2 + bx + c,$$

and since the left-hand side of this equation equals $M^2x^2 + 2MBx + B^2$, these points are those whose x-coordinates satisfy

$$x^3 + (a - M^2)x^2 + (b - 2MB)x + (c - B^2) = 0. \tag{6.3.1}$$

Let $h(x)$ denote the left-hand side of (6.3.1), so that the points in $E \cap L$ are those whose x-coordinate satisfies $h(x) = 0$. Since $P, Q \in E \cap L$, x_1 and x_2 must both be roots of $h(x)$. Since $x_1 \neq x_2$, this means that $(x - x_1)(x - x_2)$ is a factor of $h(x)$. Applying polynomial long division to find the quotient of $h(x)$ by this factor, the result must be $x - x_3$ for some $x_3 \in F$, since $h(x)$ is a monic (its leading coefficient equals 1) cubic. Then

$$h(x) = (x - x_1)(x - x_2)(x - x_3) \tag{6.3.2}$$

and x_3 is a root of $h(x)$, so x_3 is the x-coordinate of a point $R = (x_3, y_3)$ in $E \cap L$. In fact, there can only be one such point, else the line passing through them would be vertical.

The coefficient of x^2 in a monic cubic polynomial is the negative of the sum of its zeros (check this!). By (6.3.1), this coefficient for $h(x)$ equals $a - M^2$, while the corresponding sum is $x_1 + x_2 + x_3$ by (6.3.2). Therefore, $a - M^2 = -(x_1 + x_2 + x_3)$, and

$$x_3 = M^2 - a - x_1 - x_2.$$

Since L passes through x_1 and x_3 and has slope M, $\frac{y_3 - y_1}{x_3 - x_1} = M$, and

$$y_3 = M(x_3 - x_1) + y_1 = M(M^2 - a - 2x_1 - x_2) + y_1.$$

We conclude that if L is not vertical, then

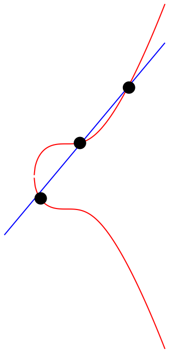
$$\begin{aligned} R =&(x_3, y_3), \text{ where} \\ x_3 &= M^2 - a - x_1 - x_2 \\ y_3 &= M(x_3 - x_1) + y_1 = M(M^2 - a - 2x_1 - x_2) + y_1. \end{aligned} \tag{6.3.3}$$

lies on $E \cap L$, and no other points besides P and Q lie on this intersection. We summarize our conclusions:

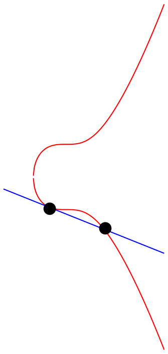
Case 2 (L not vertical). Suppose that L is not vertical and $P \neq Q$ lie on $E \cap L$. Then so does R as defined in (6.3.3), and no other points are in the intersection. Hence

(a) If R is distinct from P and Q, then exactly three points, P, Q, and R, lie on $E \cap L$.

(b) If $R = P$ or $R = Q$, then exactly two points lie on $E \cap L$, namely P and Q.



An example of Case 2(a), the **typical case**.



An example of Case 2(b)

In fact, **Case 2(a) is the "typical" case**, meaning that for most lines intersecting an elliptic curve, they intersect the curve at exactly three points. We've already see that Case 1 is a special one, and Case 2(b) actually only occurs if L happens to be tangent to P. Let's formalize what happens in the two cases if L is tangent to E.

Lemma 6.4. *Suppose that the line L is tangent the elliptic curve E at the point $P = (x_1, y_1)$. If L is vertical, then P is on the only point on $E \cap L$. Otherwise, the only other point on $E \cap L$ is R as defined in* (6.3.3)*, but with $x_2 = x_1$ and $M = \frac{dy}{dx}\big|_P = \frac{f'(x_1)}{2y_1}$.*

Proof. In assuming that L is tangent to E at P, notice that it is necessarily the case that either $\frac{dy}{dx}$ exists at P, or that the tangent line is vertical.

Implicitly differentiating the formula $y^2 = f(x)$ for E, we find $2y\frac{dy}{dx} = f'(x)$. Hence $\frac{dy}{dx}\big|_P = \frac{f'(x_1)}{2y_1}$ unless $y_1 = 0$, in which case the tangent line is vertical. If L is vertical, then we know from Case 1 above that P is the only point on $E \cap L$.

If L is not vertical, i.e., $y_1 \neq 0$, then take another point $Q = (x_2, y_2)$ on $E \cap L$. Then if $R = (x_3, y_3)$ is the point on E that intersects the line through P and Q, its coordinates satisfy (6.3.3).

Consider what happens when Q approaches P. In this case, $x_2 \to x_1$ and $y_2 \to y_1$, and the slope between P and Q approaches the slope of the tangent line, $\frac{dy}{dx}\big|_P$. Hence as $Q \to P$, i.e., L approaches the tangent line to P, R approaches the point as in (6.3.3), but with x_2 replaced with x_1 and M replaced with $D = \frac{dy}{dx}\big|_P$. More specifically,

$$\begin{aligned}\lim_{Q\to P} x_3 &= \lim_{Q\to P}\left(M^2 - a - x_1 - x_2\right) = D^2 - a - 2x_1\\ \lim_{Q\to P} y_3 &= \lim_{Q\to P}\left(M(M^2 - a - 2x_1 - x_2) + y_1\right) = D^2\left(D^2 - a - 3x_1 - x_2\right) + y_1.\end{aligned}$$

Hence these are the coordinates of the other point on $E \cap L$. □

Finally, we can define a group law on the elliptic curve E, assuming that at each point of E, either $\frac{dy}{dx}$ exists, or $\frac{dy}{dx}$ is infinite, i.e., the tangent line is vertical. The group relies on the fact that a line usually intersects an elliptic curve at three points; however, we have seen that there are exceptions to this statement. To rectify this, we add a "point at infinity" to the group, which serves as the identity of the group. We use addition as the operation on an elliptic curve group, so we call this extra point "O.'

From now on, we will always include the point O in our elliptic curves, so that they have this underlying group structure.

Preliminary definition 6.5 (Elliptic curve group)**.** Given a field F and elements $a, b, c \in F$, consider the set E of all points (x, y) satisfying $y^2 = x^3 + ax^2 + bx + c$, along with the point at infinity, denoted "O." Then $(E, +)$ forms a group under the following axioms:

1. The infinite point O is the identity.
2. The inverse (negative) of a point $P = (x, y)$ in E is $-P = (x, -y)$.
3. If a line intersects points P, Q, and R in E, then

$$P + Q + R = O,$$

so $P + Q = -R$.

4. If a line is tangent to P, then we consider the point to have multiplicity greater than 1 on the line. If there is another point R on this line, then P is considered to have multiplicity 2 and

$$P + P + R = 2P + R = O,$$

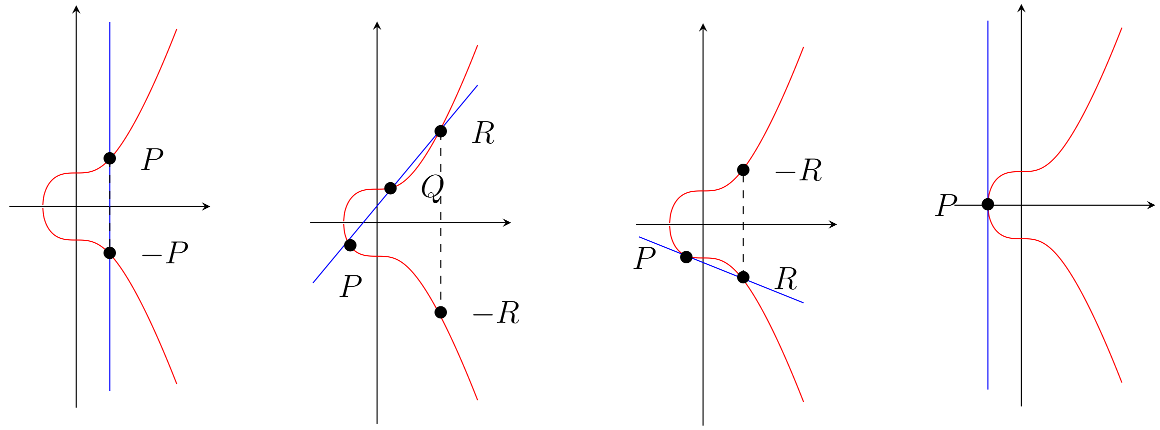
so $2P = -R$, and the formula (6.3.3) will confirm this. If there is no other point on the line, then the computation, and P has multiplicity 3 on the line so that

$$P + P + P = 3P = O$$

and $2P = -P$, and the formula (6.3.3) will confirm this. .

5. If a line is vertical, we consider the infinite point O to be on the line. Hence if points P, Q in E are on the line, then $P + Q + O = O$, so $P = -Q$.

The following graphs illustrate features of the group law:



$-P$ is the reflection of P about the x-axis.

$P + Q = -R$

$2P = -R$

$2P = O$ and $-P = P$

Figure 6.5.1: Examples illustrating the elliptic curve group structure

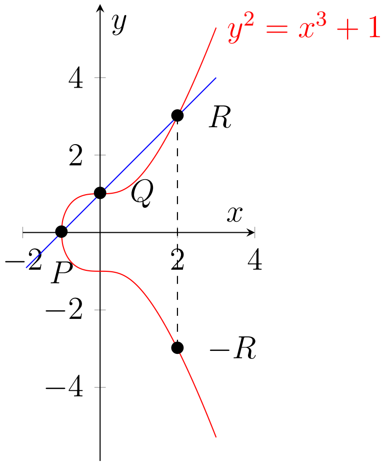
Remark 6.6. It is possible that

Hint: If you find that a line is tangent to some point S on the elliptic curve, and there are no other points on the intersection of the curve with the tangent line, then S is considered to have multiplicity 3 on E, i.e., $S + S + S = 3S = O$. To verify this, either use the addition formula applies to $S + S$, and you should find that $S + S = -S$, or you can plug the equation for the tangent line into the elliptic curve equation, an check that the x-coordinate is triple root of the resulting cubic equation.

Example 6.7 (Elliptic curve group law)**.** Consider the elliptic curve group $(E, +)$ built from Example 6.2, where

$$E = \{(x, y) \in \mathbb{R}^2 \mid y^2 = x^3 + 1\} \cup \{O\}.$$

Then $P = (-1, 0)$ and $Q = (0, 1)$ satisfy the equation above. Then $-Q = (0, -1)$. Moreover, $P + Q + R = O$, where R is the other point on E and the line passing through P and Q; this line has slope $M = \frac{1-0}{0-(-1)} = 1$. Using (6.3.3), since $M = 1$ and $a = 0$, we find that $R = (1^2 - (-1), 1 \cdot (1^2 - 2(-1))) = (2, 3)$. Since O is the identity, we conclude that $P + Q = -R = (2, -3)$.



Virtual Class Notes, Week 9 (March 30 - April 3).

Finding methods for factoring large integers is fundamental to attacks on certain cryptosystems. For example, in order to break the RSA cryptosystem, one must factor the public modulus into its two prime factors.

This week we study two algorithms that push primality testing further, by not only determining that a given integer n is composite, but by finding a **proper factor** $k > 0$, meaning $k \neq 1$ and $k \neq n$. Then we can divide n by k to find another proper factor j, for which $n = kj$. If n only has two prime factors (like in RSA), we've found them!

For a general composite integer n, though, k and j may not be prime, but we can determine whether each is prime using the Miller-Rabin test. If they are both prime, then we have obtained the prime factorization of n. If not, we can then apply the factoring algorithm to whichever are not prime to attempt to factor further. Repeating this process, we may be able to determine the prime factorization of any composite integer n.

For instance, if we apply a factoring algorithm to $n = 3819$, and find that 57 is a factor. Dividing out, we find that $n = 57 \cdot 67$. We can then apply the factoring algorithm to 57 and 67; suppose we find that $57 = 3 \cdot 19$, but the algorithm fails to find a factor of 67. Hence $3819 = 3 \cdot 19 \cdot 67$, and we can easily check that each of these factors is prime.

7 Pollard's $p - 1$ factoring algorithm

Our first factoring method is called *Pollard's $p - 1$ factoring algorithm*. Suppose that we want to factor an integer $n > 1$. If the algorithm succeeds, this method is pretty efficient in finding a factor of n in the case that n has at least one prime factor p for which the prime factorization of $p - 1$ consists of small primes. For example, the integer $n = 15\,023$ factors as

$83 \cdot 181$. Notice that $83 - 1 = 82 = 2 \cdot 41$, so its prime factorization has a fairly large prime 41. However, $181 - 1 = 182 = 2^2 \cdot 3^2 \cdot 5$, so all its factors are small primes.

In fact, suppose that n has a prime factor p, and B is a positive integer for which $(p-1)|B!$, where $B!$ is the product $B \cdot (B-1) \cdots 3 \cdot 2 \cdot 1$. If Pollard's $p-1$ method finds the factor p of n, or a multiple of it, then it finds one in at most B steps. In our example of $n = 15\,023$ above, since $p = 181$ is a prime factor of n and $p - 1 = 182 = 2^{\cdot}3^2 \cdot 5$, we find that $p - 1$ divides

$$6! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = (3 \cdot 2) \cdot 5 \cdot (2 \cdot 2) \cdot 3 \cdot 2 \cdot 1 = 2^4 \cdot 3^2 \cdot 5$$

(but notice that $(p-1) \nmid 5!$), so Pollard's $p-1$ method will take at most $B = 6$ steps to find the factor 181 if it does so.

Let's see how the algorithm runs:

Method 7.1 (Pollard's $p-1$ factoring algorithm)**.** Our goal is to find a proper factor of a composite integer n. Choose an integer a, $1 < a < n$. If n and a are not relatively prime, then (a, n) is a proper factor of n. Otherwise, successively compute the following:

$$\begin{aligned}
&a_2 \ \% \ n, \text{ where } a_2 = a^2 \\
&a_3 \ \% \ n, \text{ where } a_3 = a_2^3 = (a^2)^3 = a^{2\cdot 3} \\
&a_4 \ \% \ n, \text{ where } a_4 = a_3^4 = (a^{2\cdot 3})^4 = a^{2\cdot 3\cdot 4} \\
&\quad\vdots \\
&a_k \ \% \ n, \text{ where } a_k = a_{k-1}^k = (a^{2\cdot 3\cdots(k-1)})^k = a^{k!} \\
&\quad\vdots
\end{aligned}$$

At each step, compute $(a_k - 1, n)$, and continue until $(a_k - 1, n) \neq 1$. If this greatest common divisor is not n, then it is a proper factor of n. If $(a_k - 1, n) = n$, then the algorithm fails, and one can start the algorithm again with another value of a.

In the algorithm, it is preferable to choose a to be a fairly small integer, so that the computations are more efficient.

Example 7.2 (Pollard's $p-1$ factoring algorithm)**.** Let's use our example of $n = 15\,023$ from above, and choose $a = 2$. Then we start by computing

$$\begin{array}{lll}
a_2 \equiv 2^2 \equiv 4 & \bmod 15\,032 & \text{and compute } (3, 15\,032) = 1 \\
a_3 \equiv a_2^3 \equiv 4^3 \equiv 64 & \bmod 15\,032 & \text{and compute } (63, 15\,032) = 1 \\
a_4 \equiv a_3^4 \equiv 64^4 \equiv 11\,548 & \bmod 15\,032 & \text{and compute } (11\,547, 15\,032) = 1 \\
a_5 \equiv a_4^5 \equiv 11\,548^5 \equiv 5924 & \bmod 15\,032 & \text{and compute } (5923, 15\,032) = 1 \\
a_6 \equiv a_5^6 \equiv 5924^6 \equiv 5431 & \bmod 15\,032 & \text{and compute } (5430, 15\,032) = 181
\end{array}$$

Hence we found the factor 181 of n, and dividing out, we can find $n = 181 \cdot 83$. As an exercise, apply the Miller-Rabin test to show that each of these factors is prime!

Remark 7.3 (Efficiency of Pollard's $p-1$ factoring algorithm.)**.** Suppose that p is a prime factor of n, and $(p-1) \mid B!$. Then if $(a,n)=1$, we know that $(a,p)=1$, so that by Fermat's little theorem, $a^{p-1} \equiv 1 \bmod p$. Now, $B! = (p-1)k$ for some integer k, so $a_B \equiv a^{B!} \equiv (a^{p-1})^k \equiv 1 \bmod p$. Therefore, $p \mid (a_B - 1)$, so $(a_B - 1, n)$ is a multiple of p. If it is a multiple besides n, we have found a proper factor!

Remember that in our example of $n = 15\,023$ with prime factor $p = 181$ of n, we computed earlier that $p-1$ is a divisor of 6! but not 7!, and so that since the algorithm found the factor p, it must have been found in at most 6 steps; recall that the method took all six! The only other proper factor of n is the prime $q = 83$, $q - 1 = 82 = 2 \cdot 41$. Here, $q-1$ is a divisor of 41! but not of 40!, so the algorithm would take at most 41 steps to find the factor 83, if it does. (Thankfully, we didn't need to do this.)

Why did we actually obtain a proper factor of $n = 15\,023$ in our example? Recall that n has prime factorization $181 \cdot 83$, and since $180 \mid 6!$, $a_6 \equiv a^{6!} \equiv 1 \bmod 181$, so that $181 \mid (a_6 - 1, n)$. Then $(a_6 - 1, n) = 181$ if and only if $83 \nmid (a_6 - 1, n)$, or $a^{6!} \not\equiv 1 \bmod 83$.

Let's consider whether this is the case. If a is not a unit modulo 83 (i.e., a is a multiple of 83), then $a^m \not\equiv 1 \bmod 83$ for every integer m. If a is a unit modulo 83, then the order of a modulo 83 is a divisor of $82 = 2 \cdot 41$, so is either $2, 41$, or 82. Since $2 \mid 6!$ but $41 \nmid 6!$ and $82 \nmid 6!$, we have that $a^{6!} \equiv 1 \bmod 83$ exactly if the order of a modulo 83 equals 2. Hence we obtain the proper factor 181 if the order of a modulo 83 is not 2. We chose $a = 2$, whose order is in fact 83!

8 The $p+1$ factoring algorithm

Now we turn to another factoring algorithm that is a type of analog of Pollard's $p-1$ algorithm, called the $p+1$ *factoring algorithm.* Notice that in Pollard's method, once we check that $(a,n)=1$, then all powers of a are units modulo n, so all our computations are equivalent to working in the group of units $\langle \mathbb{Z}/n/\mathbb{Z}\rangle^\times$.

In the $p+1$ algorithm, we perform a similar process, but work in a new group called a *circle group.* If you're familiar with the complex numbers, the circle subgroup of this group consists of all complex numbers of norm 1, which trace out the unit circle in the complex plane (which is where the nomenclature comes from).

Recall that a **field** is a commutative ring with at least two elements, in which every nonzero element is a unit. We know that if p is prime, $\mathbb{Z}/p\mathbb{Z}$ is a field with p elements. In fact, there are other fields that have finitely many elements. We will work in the group of units of a field with p^2 elements, where p is prime.

The finite field $\mathbb{F}_{p^2}$. Fix a prime p, and an integer d that is not a square modulo p. Then the ring $\mathbb{F}_{p^2}$ can be defined as the set of elements of the form

$$a + b\sqrt{d}$$

where a, b are integers, and $(\sqrt{d})^2 = d$. Two elements $a+b\sqrt{d}, a'+b'\sqrt{d} \in \mathbb{F}_{p^2}$ are equal if and only if $a \equiv a' \bmod p$ and $b \equiv b' \bmod p$. Notice that there is a unique representation of each element $\mathbb{F}_{p^2}$ as $a + b\sqrt{d}$, where $0 \leq a, b < p$, so $\mathbb{F}_{p^2}$ contains exactly p^2 elements.

Addition and multiplication are defined as you might expect:

$$(a + b\sqrt{d}) + (a' + b'\sqrt{d}) = (a + a') + (b + b')\sqrt{d}$$
$$(a + b\sqrt{d}) \cdot (a' + b'\sqrt{d}) = aa' + ab'\sqrt{d} + a'b\sqrt{d} + bb'd = (aa' + bb'd) + (ab' + a'b)\sqrt{d}$$

It is apparent from these formulas that both addition and multiplication are commutative, and it is straightforward to check that that both are associative. This ring has additive identity $0 = 0 + 0\sqrt{d}$ and multiplicative identity $1 = 1 + 0\sqrt{d}$.

Given an element $z = a + b\sqrt{d}$ of $\mathbb{F}_{p^2}$, its **imaginary part** is $\mathrm{Im}(z) = b$, and its **conjugate** is defined as $\overline{z} = a - b\sqrt{d}$, so that $\mathrm{Im}(\overline{z}) = -b$ and $\overline{\overline{z}} = z$. If z is nonzero, its multiplicative inverse is $\frac{\overline{z}}{a^2-b^2d} = \left(\frac{a}{a^2-b^2d}\right) - \left(\frac{b}{a^2-b^2d}\right)\sqrt{d}$, which follows from the fact that

$$z \cdot \overline{z} = (a + b\sqrt{d})(a - b\sqrt{d}) = a^2 - b^2 d.$$

Notice that $z \cdot \overline{z}$ can be thought of as an element of $\mathbb{Z}/p\mathbb{Z}$.

Though it appears that the field $\mathbb{F}_{p^2}$ depends on the choice of d, any field as defined above has the same ring structure after renaming the elements; we say that they are *isomorphic*.

The circle group $T(p)$. The circle group is the subgroup of the group of units $\mathbb{F}_{p^2}^* = \mathbb{F}_{p^2} \smallsetminus \{0\}$ consisting of all elements $z = a+b\sqrt{d}$ of $\mathbb{F}_{p^2}^*$ for which $z \cdot \overline{z} = a^2 - b^2 d \equiv 1 \bmod p$.

In fact, the circle group has $p + 1$ elements, in the $p + 1$ factoring algorithm, this group can be thought of as taking the role of the group of units $(\mathbb{Z}/p\mathbb{Z})^\times$ (which has $p-1$ elements) in Pollard's $p - 1$ factoring algorithm.

Example. Take $p = 5$. Notice that since $0^2 = 0, 1^2 = 1, 2^2 = 4, 3^2 \equiv 4 \bmod 5$, and $4^2 \equiv 1 \bmod 5$, we know that $d = 2$ is not a square modulo 5. Then $\mathbb{F}_{5^2} = \mathbb{F}_{25}$ consists of the 25 elements $a + b\sqrt{2}$, where $0 \leq a, b \leq 4$.

If $z = 3 + 4\sqrt{2}$, then its imaginary part is $\mathrm{Im}(z) = 4$, and its conjugate is $\overline{z} = 3 - 4\sqrt{2} = 3 + 1\sqrt{2} = 3 + \sqrt{2}$. If $w = 2 + 3\sqrt{2}$, then

$$z + w = (3 + 4\sqrt{2}) + (2 + 3\sqrt{2}) = 5 + 7\sqrt{2} = \boxed{5 + 2\sqrt{2}}$$
$$zw = (3 + 4\sqrt{2})(2 + 3\sqrt{2}) = 6 + 9\sqrt{2} + 8\sqrt{2} + 12 \cdot 2 = 30 + 17\sqrt{2} = 0 + 2\sqrt{2} = \boxed{2\sqrt{2}}$$

The circle group $T(5)$ consists of the nonzero elements $z = a + b\sqrt{2}$ of $\mathbb{F}_{25}$ for which $z \cdot \overline{z} = a^2 - 2b^2 \equiv 1 \bmod p$. The multiplicative identity $1 = 1 + 0\sqrt{2}$ is clearly in this group, and so is -1. A more interesting element is, for instance, $2 + 2\sqrt{2} \in T(5)$ since $(2 + 2\sqrt{2})(2 - 2\sqrt{2}) = 4 - 2 \cdot 4 = -4 = 1$ since $-4 \equiv 1 \bmod 5$. The circle group has

$p+1=6$ total elements–try finding the others!

In what follows, we use analogous definitions to those in the definition of $\mathbb{F}_{p^2}$. Given integers a, b, and d, $z = a + b\sqrt{d}$ is a complex (and perhaps real) number, $\mathrm{Im}(z) = b$, and $\overline{z} = a - b\sqrt{d}$. Given an integer $n > 1$, we write $z \ \% \ n$, to mean the number $a' + b'\sqrt{d}$, where $a' = a \ \% \ n$ and $b' = b \ \% \ n$.

The $p+1$ factoring algorithm. Our goal is to find a proper factor of a composite integer n. Fix integers a, b, and d, and let $z = a + b\sqrt{d}$. If $1 < (z \cdot \overline{z}, n) < n$, then this greatest common divisor is a proper factor of n.

Otherwise, successively compute the following:

$$\begin{aligned}
&z_2 \ \% \ n, \text{ where } z_2 = z^2\\
&z_3 \ \% \ n, \text{ where } z_3 = z_2^3 = (z^2)^3 = z^{2\cdot 3}\\
&z_4 \ \% \ n, \text{ where } z_4 = z_3^4 = (z^{2\cdot 3})^4 = z^{2\cdot 3\cdot 4}\\
&\quad\vdots\\
&z_k \ \% \ n, \text{ where } z_k = z_{k-1}^k = (z^{2\cdot 3 \cdots (k-1)})^k = z^{k!}\\
&\quad\vdots
\end{aligned}$$

At each step, compute $(\mathrm{Im}(z_k), n)$, and continue until $(\mathrm{Im}(z_k), n) \neq 1$. If this greatest common divisor is not n, then it is a proper factor of n. If $(\mathrm{Im}(z_k), n) = n$, then the algorithm fails, and one can start the algorithm again with another value of z.

Notice how this is analogous to Pollard's $p-1$ method! In fact, if n has a prime factor p such that the prime factorization of $p+1$ consists of small primes, then the $p+1$ algorithm is fairly efficient. To prove the following proposition that shows this, we need to apply the following variant of **Lagrange's theorem**: Suppose that x is an element of a finite group G under multiplication. If G has m elements, then $x^m = 1$ in G. When $G = (\mathbb{Z}/p\mathbb{Z})^\times$, which has $p-1$ elements, we see that this is precisely Fermat's little thoerem!

Proposition. Suppose that n is a composite integer with prime factor p, and fix an integer B for which $(p+1) \mid B!$. Suppose that in the $p+1$ algorithm, d is not a square modulo p, and a and b are not both zero modulo p. Then the algorithm terminates (either finds a proper factor of n, or fails by finding the factor n) in at most B steps.

Proof. Given integers a, b, and d that is not a square modulo p, notice that as an element of $\mathbb{F}_{p^2}$, $z = a + b\sqrt{d}$ is nonzero. Consider the element $w = z \cdot \overline{z}^{-1}$. As an exercise, check that $\overline{w} = \overline{z} \cdot z^{-1}$. Then

$$w \cdot \overline{w} = (z \cdot \overline{z}^{-1})(\overline{z} \cdot z^{-1}) = 1$$

so w is in the circle group $T(p)$. Since $T(p)$ has $p+1$ elements, the version of Lagrange's theorem stated above implies that $w^{p+1} = 1$ in $T(p)$. Since $(p+1) \mid B!$, we then have that $w^{B!} = 1$ as well; i.e.,

$$(z \cdot (\overline{z})^{-1})^{B!} = z^{B!} \cdot (\overline{z})^{-B!} = 1,$$

forcing $z^{B!} = (\overline{z})^{-B!}$, which you can check equals $\overline{z^{B!}}$. Since an element of $\mathbb{F}_{p^2}$ equals its conjugate exactly if its imaginary part is zero, we conclude that $\operatorname{Im}\left(z^{B!}\right) \equiv 0 \bmod p$.

This means that $p \mid \operatorname{Im}(z_B)$, and since p is also a divisor of n, $p \mid (\operatorname{Im}(z_B), n)$, so $(\operatorname{Im}(z_B), n)$ is a factor of n. □

Notice that since before applying the $p+1$ algorithm, we do not know the prime factors p of n, so we cannot necessarily choose an integer d that is not a square modulo such a factor p. However, recall that in determining the formula for square roots modulo primes p such that $p \equiv 3 \bmod 4$, we proved that -1 is never a square modulo such a prime. Hence, if n has any prime factor congruent to 3 modulo 4, then the algorithm will find a multiple of this factor using $d = -1$. In this case, we often use i to denote the element $\sqrt{-1}$ of $\mathbb{F}_{p^2}$.

In general, for a randomly chosen d, there is a $\frac{1}{2}$ chance that d is not a square modulo any prime factor p of d, so by running the algorithm multiple times, we should find a d for which the algorithm terminates fairly quickly.

Example ($p+1$ factoring algorithm). Consider $n = 851$. Let's choose $d = -1$, and $z = 1 + 2i$, where $i = \sqrt{-1}$. We first find $z \cdot \overline{z} = 1^2 - 2^2 \cdot -1 = 5$, and find that $(5, 851) = 1$. Then we proceed as follows:

$$\begin{aligned}
z_2 &\equiv (1+2i)^2 \equiv -3+4i && \text{mod } 851 \text{ and compute } (4, 851) = 1\\
z_3 &\equiv (-3+4i)^3 \equiv 117+44i && \text{mod } 851 \text{ and compute } (44, 851) = 1\\
z_4 &\equiv (117+44i)^4 \equiv 32\,125\,393 + 242\,017\,776i && \\
&\equiv 143 + 184i && \text{mod } 851 \text{ and compute } (184, 851) = 23
\end{aligned}$$

We have found the factor 23 of n, and dividing out, $n = 23 \cdot 37$!

Note that $p = 23$ was a factor of n in this exercise, and since $p \equiv 3 \bmod 4$ and $p + 1 = 24 = 2^3 \cdot 3$, which divides $4!$, the algorithm was guaranteed to terminate in at most 4 steps.

Virtual Class Notes, Week 8 (March 23 - 27).

9 The Miller-Rabin primality test

Recall that before Spring Break, we stated the Miller-Rabin primality test. This test, when conclusive, allows one to conclude that a given integer n is composite.

Method 9.1 (Miller-Rabin primality test)**.** Take an odd integer $n \geq 3$, so that $n-1$ is even. Factor out the highest power of 2 as possible from $n-1$, writing $n - 1 = 2^k \cdot q$, where q is odd. (Note that k and q are unique!)

Then n is **composite** if for some fixed $a \in \mathbb{Z}$, $1 < a < n$, the following properties hold:

1. $a^q \not\equiv 1 \bmod n$, and
2. $a^{2^i q} \not\equiv -1 \bmod n$ for all $i = 0, 1, \ldots, k-1$.

Proof. Assume that $p = n$ is prime. We need to show that one of the conditions must fail, so either (1) $a^q \equiv 1 \bmod p$, or (2) one of the following must be true:

$$a^q \equiv -1, a^{2q} \equiv -1, a^{2^2 q} \equiv -1, \cdots, \text{ or } a^{2^{k-1}q} \equiv -1 \bmod p. \qquad (\star)$$

Now, $a^{2^{k-1}q}$ is a square root of 1 modulo p, since

$$(a^{2^{k-1}q})^2 \equiv a^{2^k q} \equiv a^{n-1} \equiv a^{p-1} \equiv 1 \bmod p$$

by Fermat's little theorem.

We know (e.g., you proved it on Midterm 1!) that the only square roots modulo a prime are ± 1. If $a^{2^{k-1}q} \equiv -1 \bmod p$, then the last congruence in ($\star$) holds. Otherwise, we know that $a^{2^{k-1}q} \equiv 1 \bmod p$. In this case, $a^{2^{k-2}q}$ is a square root of 1 modulo p, so $a^{2^{k-2}q} \equiv -1 \bmod p$ or $a^{2^{k-2}q} \equiv 1 \bmod p$. In the first case, the second-to-last equation in ($\star$) holds, and in the second case, $a^{2^{k-3}q}$ is a square root of 1 modulo p!

Hence we can continue in this manner (e.g., formally, by induction), to conclude that if none of $a^{2^{k-1}q}, a^{2^{k-2}q}, \cdots, a^{2q}$ are congruent to -1 modulo p, then then they are all congruent to 1, and a^q is a square root of 1 modulo p. In this case, $a^q \equiv -1 \bmod p$ or $a^q \equiv 1 \bmod p$, so that either the first equation in ($\star$) holds, or the original condition (1) holds! □

Example 9.2 (Miller-Rabin primality test)**.** Let's apply the Miller-Rabin test to $n = 713$, which doesn't have any obvious small factors. Since $n - 1 = 712 = 2^3 \cdot 89$ and 89 is odd, we have $k = 3$ and $q = 89$ in the statement of the test. Let's try the smallest permissible a value, $a = 2$. We compute, using fast exponentiation (which we omit):

$$\begin{aligned} a^q &\equiv 2^{89} \equiv 140 \not\equiv \pm 1 \bmod 713 \\ a^{2q} &\equiv 2^{2 \cdot 89} \equiv (2^{89})^2 \equiv 140^2 \equiv 19\,600 \equiv 349 \not\equiv -1 \bmod 713 \\ a^{2^2 q} &\equiv 2^{2^2 \cdot 89} \equiv (2^{2 \cdot 89})^2 \equiv 349^2 \equiv 121\,801 \equiv 591 \not\equiv -1 \bmod 713 \end{aligned}$$

Since $k - 1 = 2$, these are the only values we need to compute to conclude that $n = 713$

is composite! (In fact, $713 = 23 \cdot 31$.)

Recall that $n = 561$ is a Carmichael number; it is composite and $a^{561} \equiv a \bmod 561$ for every integer a. Hence it is not possible to show 561 is composite by exhibiting a case in which the conclusion of Fermat's little theorem fails. However, the Miller-Rabin test does show 561 is composite! Try doing this yourself, and then check your work in Savin text, where this example is worked out (page 149).

In fact, if n is composite, more than 75% of choices for a, $1 < a < n$, in the Miller-Rabin test are witnesses for the compositeness of n! Hence if we start by trying $a = 2, 3, \ldots$ (which are the easiest to apply to the test, since they are small), we are likely to soon come across a value that works. On the other hand, this means that if n is actually prime, the Miller-Rabin test can tell us that it is *likely* that n is prime: For instance, if we pick 5 random values for a and the Miller-Rabin test is inconclusive, then since $(1/4)^{10} = 1/1024 < .001$, there is a 99.9% chance that n is prime. Moreover, if we check more than $\frac{1}{4}$ of values for a in the range $1 < a < n$ and the test is inconclusive, then we **can conclude that n is prime!**

10 The parity of solutions to the discrete logarithm problem

The idea behind the Miller-Rabin primality test–that the only square roots of 1 modulo a prime are ±1–actually helps us determine the parity of solutions (i.e., whether they are even or odd) to the discrete logarithm problem!

Notice that if p is an odd prime, then $p-1$ is even, so $(p-1)/2$ is a positive integer.

Theorem 10.1 (Parity of solution to discrete logarithm problem)**.** *Fix an odd prime p, a primitive root g modulo p, and a unit X modulo p. If an integer x_0 is a solution to the discrete logarithm problem*

$$g^x \equiv X \bmod p,$$

then

- *x_0 is even if and only if $X^{\frac{p-1}{2}} \equiv 1 \bmod p$, and*
- *x_0 is odd if and only if $X^{\frac{p-1}{2}} \equiv -1 \bmod p$.*

Proof. First notice that since g is a primitive root modulo p, $g^{p-1} \equiv 1 \bmod p$, but $g^t \not\equiv 1 \bmod p$ for $0 < t < p-1$. Using this, we calculate:

- If x is even, so $x = 2k$ for some integer k,

$$X^{\frac{p-1}{2}} \equiv (g^x)^{\frac{p-1}{2}} \equiv (g^{2k})^{\frac{p-1}{2}} \equiv g^{k(p-1)} \equiv (g^{p-1})^k \equiv 1^k \equiv 1 \bmod p.$$

- If x is odd, so $x = 2k+1$ for some integer k,

$$X^{\frac{p-1}{2}} \equiv (g^x)^{\frac{p-1}{2}} \equiv (g^{2k+1})^{\frac{p-1}{2}} \equiv g^{k(p-1)+\frac{p-1}{2}} \equiv (g^{p-1})^k \cdot g^{\frac{p-1}{2}} \equiv g^{\frac{p-1}{2}} \not\equiv 1 \bmod p.$$

Finally, notice that since $X^{p-1} \equiv 1 \bmod p$ by Fermat's little theorem, $X^{\frac{p-1}{2}}$ is a square root of 1 modulo p, so it must be 1 or -1 modulo p. □

Example 10.2 (Finding the parity of a discrete logarithm solution)**.** In fact, $g = 75$ is a primitive root modulo the prime $p = 101$. Suppose we are interested in the discrete logarithm problem

$$75^x \equiv 40 \bmod 101.$$

Is a solution x_0 odd or even?

Using the theorem above, taking $X = 40$, since $(p-1)/2 = 50$, we calculate

$$X^{(p-1)/2} \equiv 40^{50} \equiv 40^{32} \cdot 40^{16} \cdot 40^2 \equiv 79 \cdot 68 \cdot 85 \equiv 100 \equiv -1 \bmod 101$$

Hence any solution x_0 is odd! In fact, using the Baby-step, giant-step method for solving discrete logarithms, we can verify that the solution $1 < x_0 < 101$ is $x_0 = 31$.

It makes sense that if a solution x_0 to the discrete logarithm problem $g^x \equiv X \bmod p$ is odd (or even, respectively), then all solutions are odd (even): Since g has order $p-1$, if x_0 is one solution, then all solutions have the form $x_0 + k(p-1)$, where k is any integer. (Check this!)

Thursday, March 5. Today we recalled from last time that composite integers $n > 1$ with the property that $a^n \equiv a \bmod n$ for all $a \in \mathbb{Z}$ are called **Carmichael numbers** (or "Fermat pseudoprime"), and that 561 as $3 \cdot 11 \cdot 17$ is the smallest one. We proved that 561 is a Carmichael number using the facts that (1) factors into a product of distinct primes, and (2) for each prime factor $p = 3, 11, 17$, $p-1$ is a divisor of $560 = 561 - 1$.

We listed a few of the next Carmichael numbers, and noted that there are infinitely many Carmichael numbers with any number of prime factors. In all our examples, all of the prime factors have multiplicity one.

These observations led us to state the following criterion for determining whether a given number is a Carmichael number:

Korselt's criterion for Carmichael numbers: A composite integer $n > 1$ is a Carmichael number if and only if the following hold:

1. n is squarefree ($m^2 \nmid n$ for all $m > 1$, or equivalently, $p^2 \nmid n$ for all primes p).

2. If p is prime and $p \mid n$, then $(p-1) \mid (n-1)$.

We proved that Korselt's criterion is valid; an interesting feature is that chose special integers a, and used the fact that $a^n \equiv a \bmod n$ to deduce (1) and (2): a is chosen to be prime for (1), and a is chosen to be a primitive root modulo a prime factor of n for p.

The upshot of our discussion is that although if we can find $a \in \mathbb{Z}$ for which $a^n \not\equiv a \bmod n$, we can deduce that $n > 2$ is composite, such an a doesn't always exist for composite n.

Fortunately, we can refine this idea, and use our theory of square roots modulo primes, to obtain a much more effective primality test! We noticed that if p is an odd prime, then $a^{(p-1)/2}$ must be ± 1 modulo p since it is a square root of $a^{p-1} \equiv 1 \bmod p$. This idea goes into the following test for determining that a given number is composite:

Miller-Rabin primality test: Take $n > 2$ odd, and factor out as many 2s as possible from the even number $n-1$, writing $n - 1 = 2^k \cdot q$ for $k \geq 1$ and odd q. If for some integer a, $1 < a < n$, the following hold, then n must be composite:

1. $a^q \not\equiv 1 \bmod n$, and
2. $a^{2^i q} \not\equiv -1 \bmod n$ for all $i = 0, 1, \ldots, k-1$.

Finally, we returned Midterm 1.

Tuesday, March 3. Today, we took Midterm 1.

Thursday, February 27. We started class by reviewing the baby-step, giant step method for solving the discrete logarithm problem, and noticed that if there is a solution, then a solution can be found in at most $2m$ steps, where $m = \lceil\sqrt{p-1}\rceil$ steps! We went through an example.

Next, we discussed digital signatures, and described the process for RSA digital signitures. In the same setup as the RSA cryptosystem (same public and private keys), if Bob wants Alice to sign a message $x \in \mathbb{Z}$, then Alice computes $y = x^d \,\%\, n$, her signed document. Bob can then verify that it was indeed Alice who signed (since only Alice knows the decryption exponent d), but finding $y^d \,\%\, n$, which, if Alice actually signed it, will be congruent to $x^{ed} \equiv x \bmod n$, the original message. We investigated why another party cannot "forge" Alice's signature.

Finally, we started approaching the question on how to decide whether a given integer n is prime or composite. Checking whether integers $2, 3, \ldots$ up to $\sqrt{n}$ divide n is very time consuming when n is large. (We can restrict to only checking for prime factors, but this assumes we know which integers are prime to begin with!)

By the (general version) of Fermat's little theorem, if n is prime, then $a^n \equiv a \bmod n$ for all $a \in \mathbb{Z}$. Hence if $a^n \not\equiv a \bmod n$ for some integer a, then n is definitely composite. In this case, we call a a *witness* for the fact that n is composite.

This provides is an efficient way to verify that a given composite integer is actually composite: Check whether $2^n \equiv 2 \bmod n$, $3^n \equiv 3 \bmod n$, etc., until we find a congruence that fails. Unfortunately, though, there exist integers n that are not prime, but that $a^n \equiv a \bmod n$ for all integers a. Therefore, there is no witness in this case, but n is not prime.

Composite integers $n > 1$ with the property that $a^n \equiv a \bmod n$ for all $a \in \mathbb{Z}$ are called **Carmichael numbers**. The smallest Carmichael number is 561. We started justifying that this number is indeed a Carmichael number by factoring 561 as $3 \cdot 11 \cdot 17$. We'll finish this next Thursday, and then prove a criterion for testing whether an integer is a Carmichael number in general.

Tuesday, February 25. We started class today with announcements about Midterm 1, additional office hours, and quiz corrections.

After a short quiz on Euler's theorem, we recalled that the security of the RSA cryptosystem relies on the fact that it is usually very difficult to factor large numbers. We proceeded to describe our second public-key cryptosystem, which relies on the fact that the *discrete logarithm* problem is difficult to solve: After fixing a large prime p and integers g, X (where g can have large order modulo p), find an integer solution x to the equation

$$g^x \equiv X \bmod p.$$

This cryptosystem, the ElGamal cipher, can be thought of as an application of the Diffie-Hellman key exchange. Like RSA, one party (Alice) wants to receive messages from anyone, after publishing public keys. One interesting difference, though, is that each party wanting to send Alice a message must also choose their own secret key.

The ElGamal process is as follows: Alice picks a large prime p, and an integer g, preferably that has large order modulo p. She picks a private key $x \in \mathbb{Z}$, calculates $X = g^x \,\%\, p$, and publishes the public-key triple

$$(p, g, X).$$

Then Bob, or anyone who wants to send Alice a message, chooses their own private key $y \in \mathbb{Z}$. Then he calculates $Y = g^y \,\%\, p$ and $k = X^y \,\%\, p$ (this should look familiar from Diffie-Hellman!). Then given a (chunk of) plaintext $m \in \mathbb{Z}$, he turns it into the ciphertext $km \,\%\, p$. He then sends Alice this ciphertext with header "Y;" that is, he sends

$$Y; km \,\%\, p.$$

Finally, to decrypt the message, notice that as in Diffie-Hellman, Alice can find k since

$$k = X^y \,\%\, p = g^{xy} \,\%\, p = Y^x \,\%\, p$$

and she knows x, her private key, and Y, sent as the header from Bob. Hence she can calculate k^{-1} modulo p, and multiplies this by the ciphertext

$$k^{-1}(km) \equiv m \bmod p,$$

so that its least nonnegative residue modulo p is the original message m.

We went though an example of encryption/decryption using ElGamal in detail, and then noticed that it makes sense for Alice and Bob to choose private keys x and y, respectively, relatively prime to $p-1$ (which can be checked quickly via the Euclidean Algorithm).

We noticed that if an eavesdropper could find the key k, and break the code, if they could solve the discrete logarithm problem: (p, g, X) are public, so if one could find x_0 satisfying $X \equiv g^{x_0} \bmod p$, then since Y is published as Bob's header, one can compute

$$Y^{x_0} \equiv (g^y)^{x_0} \equiv (g^{x_0})^y \equiv X^y \equiv k \bmod p.$$

We discussed one method for solving the discrete logarithm problem

$$g^x \equiv X \bmod p \tag{10.2.1}$$

where p is a prime, g is a unit modulo p, and $X \in \mathbb{Z}$. The algorithm is called the baby-step, giant-step method, and takes less than $2\sqrt{p}$ steps (while computing $g, g^2, g^3, \dots$ could take many more in general). Fix the smallest integer m for which $p-1 < m^2$. If x_0 is a solution to the discrete logarithm problem, we can assume $0 \leq x_0 < p-1$ by Fermat's little theorem. Apply the division algorithm to x_0 and m to obtain

$$x_0 = mq + r$$

where $0 \leq r < m$. Notice that if $q \geq m$, then $x_0 = mq + r \geq m^2 + r \geq m^2 > p-1$, which is not the case, so we can assume that $0 \leq q, r < m$. The process goes as follows:

Baby-step, giant-step method for solving the discrete logarithm problem. To find an integer solution x_0 to (10.2.1), after choosing m as above, proceed as follows:

Baby steps. List the least nonnegative residues modulo p of

$$1, g, g^2, g^2, \dots, g^m.$$

Giant steps. List the least nonnegative residues modulo p of

$$X(g^{-m})^i \quad \text{for} \quad i = 1, 2, \dots$$

until one matches with a least nonnegative residue on the baby step list.

Then if g^j from the first list is congruent to $X(g^{-m})^i$ on the second, we have that $X \cdot g^{-mi} \equiv g^j \bmod p$, so $X \equiv g^j \cdot g^{mi} \equiv g^{mi+j} \bmod p$, and $x_0 = mi + j$ is a solution to the discrete logarithm problem.

Notice that in our setup above, since $X \equiv g^{x_0} \equiv (g^m)^q g^r \bmod p$ for some $0 \le q, r < m$, $g^r \equiv X(g^{-m})^q \bmod p$. In other words, for some j on the first list, and $0 \le i < m$ on the second list,

$$g^j \equiv X(g^{-m})^i \bmod p$$

so we know we will hit a match in fewer than m giant steps.

Thursday, February 20. Given relatively prime integers a and $m > 1$, we started class by defining the **order** of a modulo m as the smallest positive integer d for which $a^d \equiv 1 \bmod m$. Analogously, the **order** of $[a]_m \in (\mathbb{Z}/m\mathbb{Z})^\times$ is the smallest integer d for which $[a]^d = [1]$.

The existence of such an integer d comes directly from **Euler's theorem**, which says that if $(a, m) = 1$, then

$$a^{\varphi(m)} \equiv 1 \bmod m.$$

In other words, if $[a]_m \in (\mathbb{Z}/m\mathbb{Z})^\times$, then $[a]^{\varphi(m)} = [1]$.

We went through several examples of finding orders modulo different m, and applying Euler's theorem. In fact, if $(a, m) > 1$, then a cannot have an order modulo m as defined above, by a homework problem! We proved a **lemma** saying that if a and $e \ge 1$ are integers and $a^e \equiv 1 \bmod m$, then e is a multiple of the order of a modulo m. We saw that this, in combination with Euler's theorem, can help determine orders of integers!

On a related note, we turned to methods of **efficient exponentiation modulo an integer**. First, we noticed that using Fermat's little theorem of Euler's theorem, we can reduce to computing an exponent smaller than $\varphi(m)$ (which equals $p-1$ if $p = m$ is prime). E.g., since $51^{100} \equiv 1 \bmod 101$ by FLT,

$$51^{303} \equiv 51^{3\cdot 100+3} \equiv (51^{100})^3 \cdot 51^3 \equiv 1^3 \cdot 51^3 \equiv 51^3 \equiv 38 \bmod 101.$$

On the other hand, we turned to the question of finding the least nonnegative residue of 54321^{12345} modulo the prime 29989. Since $54321 \equiv 24332 \equiv -5657 \bmod 29989$, we could attack the "simpler" problem of finding the least nonnegative residue of 24332^{12345} or $(-5657)^{12345} = -5657^{12345}$ modulo 29989. Let's find the latter by first finding 5657^{2345} % 29989.

Our procedure for **fast(er) exponentiation** first requires writing the exponent in base 2. In our case,

$$12345 = 2^{13} + 2^{12} + 2^5 + 2^4 + 2^3 + 2^0.$$

Then

$$5657^{12345} = 5657^{2^{13}} \cdot 5657^{2^{12}} \cdot 5657^{2^5} \cdot 5657^{2^4} \cdot 5657^{2^3} \cdot 5657.$$

Next, we can iteratively find each term in the product modulo 29989:

$$\begin{aligned}
5657^{2} &\equiv 32001649 \equiv 3386 && \text{mod } 29989\\
5657^{2^2} &\equiv (3386)^2 \equiv 11464996 \equiv 9198 && \text{mod } 29989\\
5657^{2^3} &\equiv (9198)^2 \equiv 4235 && \text{mod } 29989\\
5657^{2^4} &\equiv 1803 && \text{mod } 29989\\
5657^{2^5} &\equiv 11997 && \text{mod } 29989\\
&\vdots\\
5657^{2^{12}} &\equiv 15464 && \text{mod } 29989\\
5657^{2^{13}} &\equiv 3010 && \text{mod } 29989
\end{aligned}$$

Note that in our procedure, we don't "skip" powers, so that the integer in each steps has a relatively small number of digits. Finally, we find that

$$5657^{12345} \equiv 3010 \cdot 15464 \cdot 11997 \cdot 1803 \cdot 4235 \cdot 5657 \quad \text{mod } 29989.$$

Again, iteratively multiplying and reducing, we find the least nonnegative residue to be 118, so that $54321^{12345} \equiv -118 \equiv 29871 \text{ mod } 29989$

Note that in class, we started calculating the residue of 4941^{12345} *modulo* 29989*. If you want to check your final answer, it is* 12047*.*

Next, we fully described the RSA Cryptosystem. One party, which we will call Alice, chooses two large distinct primes p and q, and makes $m = pq$ public (i.e., n is a *public key*), though p and q are kept secret. She then calculates $\varphi(m) = (p-1)(q-1)$, and picks an integer e relatively prime to $\varphi(m)$; this is called the **encryption exponent**. (She can easily verify that her exponent is valid by performing the Euclidean algorithm). Notice that the **public keys for RSA** are the modulus m, and encryption exponent e.

To encrypt a message represented as an integer x, anyone who wants to send Alice a message calculates

$$x^e \ \% \ m.$$

Anyone can do this, since m and e are public! However, only Alice can decrypt the message. The decryption procedure is as follows: If $y \in \mathbb{Z}$ is the encrypted message, then she takes

$$y^d \ \% \ m,$$

where d is the **decryption exponent**, which is the inverse of e modulo $\varphi(m)$. She can easily find this via back-substitution in the Euclidean algorithm.

We checked that this procedure works as planned First, we noticed that since $ed \equiv 1 \text{ mod } \varphi(m)$, $ed - 1 = \varphi(m)k$ for some $k \in \mathbb{Z}$. Hence if we encrypt x as $x^e \ \% \ m$, and then decrypt, Alice obtains an integer congruent to

$$(x^e)^d \equiv x^{ed} \equiv x^{ed-1} \cdot x \equiv x^{\varphi(m)k} \cdot x \equiv (x^{\varphi(m)})^k \cdot x \equiv 1^k \cdot x \equiv x \text{ mod } m$$

where x is the orginal message! Notice that we applied Euler's theorem to make this conclusion.

We went through an example using the primes $p = 41$ and $q = 43$.

Tuesday, February 18. We started class by defining the **least nonnegative residue** of an integer a modulo $m > 1$, often denoted $a\%m$, as the smallest nonnegative integer congruent to a modulo m.

Next, we reviewed the process of "flipping coins over the telephone," and decided that unless Alice has a process for finding square roots modulo other primes, she should choose primes congruent to 3 modulo 4, where we have a formula. We also discussed the fairness of this process for Alice and for Bob.

For p a prime, we defined a **primitive root modulo** p to be a unit a modulo p for which no pair of integers among

$$a, a^2, \ldots a^{p-1}$$

are congruent modulo p. We found all primitive roots modulo 5, and then stated a **theorem** saying the primitive roots exists modulo every prime p! In fact, there are $\varphi(p-1)$ primitive roots modulo p, where φ denotes the Euler phi function.

After this, we presented the **Diffie-Hellman Key Exchange**. The goal here is to create a secret key that only two parties know. Two public keys are published, a large prime p, and a primitive root g modulo p.

The first party, Alice, picks a secret integer x and computes $X = g^x\%p$, passing X over the public channel to the second party, Bob. Similarly, Bob picks a secret y and passes $Y = g^y\%p$ to Alice. Notice that anyone (e.g., the eavesdropper "Eve") has access to X and Y, but finding x and y from these is difficult; solving an equation of the form $X \equiv g^x \bmod p$ for x is called a **discrete logarithm problem**.

Finally, Alice computes the secret key as $k = Y^x \ \% \ p$ (she chose the secret integer x, obtained Y from Bob, and p is public) and Bob computes it as $k = X^y \ \% \ p$ similarly. Moreover, we confirm that

$$Y^x \equiv (g^y)^x \equiv (g^x)^y \equiv X^y \bmod p$$

so that Alice and Bob indeed have the same key k! We went through an entire example illustrating the key exchange, using a relatively small prime.

Next, we introduced the RSA Cryptosystem, our first public key crypotosystem. We described all public and private keys, and next time we will describe the procedures of entcryption and decryption.

Finally, we had a quiz on square roots modulo integers.

Thursday, February 13. We used the conclusions that we made last time (via the Chinese Remainder Theorem) to find square roots mod $p \cdot q$, where p and q are *distinct* primes. In particular, an integer a has a square root modulo pq if and only if it has a square root modulo p and modulo q. Using this, we found that 53 has no square root modulo 55 since it has no square root mod 5. On the other hand, 34 has square roots $12, 32, 28$, and 43 modulo 55 (i.e., $\pm 12, \pm 28$; these come from square roots of $34 \equiv 34 \bmod 5$, 2 and 3, and $34 \equiv 1 \bmod 11$, 1 and 10, via the CRT.

We concluded that there can be either zero or four square roots modulo pq if p and q are distinct primes, unless one of them is 2 or $a \equiv 0 \bmod pq$.

Next, we discussed flipping coins over the telephone. We describe the process: First, Alice chooses *large* primes p and q, and sends their product, n to Bob. Bob can't factor n; he picks a random large integer a and calculates $c \equiv a^2 \bmod n$, sending c to Alice. Alice then finds the four square roots of c modulo n using the the factorization $n\ pq$ and the Chinese Remainder Theorem. These square roots are $\pm a$ and $\pm b$ for some integer b (but she does not know which is a!) She chooses one of these, and sends it to Bob (her "guess"); say it is a If $x = \pm a \bmod n$, Alice "wins" the coin flip, and Bob wins otherwise.

We did an example of how this process works using two primes. Then we discussed how Alice can ensure that Bob doesn't cheat, and vice versa. For next time, make sure to verify that Alice can find all square roots of c modulo pq! Also, try #1 from Homework 2 to make sure that you can follow the process of finding square roots modulo pq.

After this, we described the Caesar/shift cipher, and translated it into mathematics using modular arithmetic.

Finally, we started defining the notion of a primitive root modulo a prime, but needed to fix it! We'll start here next time.

Tuesday, February 11. Today we had a reminder about office hours, and the extended deadline on the programming portion of our first homework. We also quickly discussed the solutions to the quiz from last time. From here, we recalled that last time, we showed that if p is a prime congruent to 3 modulo 4, then -1 *has no square root* modulo p.

We also showed that if an integer a *has* a square root b modulo any prime p, then its square root are precisely $\pm b$ (so there are exactly two unless $a = 0$ or $p = 2$).

Next, we stated the following **proposition**: Fix a prime p such that $p \equiv 3 \bmod 4$, and an integer a. If $p \mid a$, then a has one square root modulo p, namely, 0. Otherwise, exactly one of a or $-a$ has a square root modulo p, and this square root is $\pm a^{p+1/4}$. We proved this proposition.

Then we used the proposition to find the square roots of 5 modulo 11, and to *fail* to find the square roots of 2 modulo 11–i.e., we proved that 2 is *not* a square modulo 11,.

Finally, we posed the question of when an integer a has a square root modulo pq, if p and q are *distinct* primes. We used the CRT to show that a square root exists modulo pq if and only if a square root exists modulo p, and a square root exists modulo q! We will apply this next time.

Thursday, February 6. We started class today with a short quiz on the definition of congruences modulo an integer, the units of $\mathbb{Z}/m\mathbb{Z}$, and the statement of the Euclidean Algorithm.

We then continued class by noticing that the CRT can be applied iteratively to solve systems of congruences modulo integers that are *pairwise relatively prime.*

Next, if R and S are rings, we defined their **product** $R \times S$, another ring. Then we investigated the map

$$\begin{aligned} \psi : \mathbb{Z}/6\mathbb{Z} &\to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \\ [a]_6 &\mapsto ([a]_2, [a]_3) \end{aligned}$$

and determined that it is well-defined (actually a function between the two sets!), and a bijection. Then we unraveled this fact to show that this is equivalent to the CRT with $m = 2$ and $n = 3$!

In general, the CRT is equivalent to the bijectivity of the analogous map

$$\begin{aligned} \psi : \mathbb{Z}/mn\mathbb{Z} &\to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}\ n\mathbb{Z} \\ [a]_{mn} &\mapsto ([a]_m, [a]_n) \end{aligned}$$

under the assumption that m and n are relatively prime.

We stated **Fermat's little theorem**: Suppose that p is a prime and a is an integer not divisible by p, then

$$x^{p-1} \equiv 1 \bmod p.$$

We saw that this is equivalent to the statement that

$$x^p \equiv x \bmod p$$

for *every* integer x.

We also noticed why the following **property of primes** holds: If a and b are integers, and p is prime, then

$$p \mid ab \implies p \mid a \text{ or } p \mid b.$$

However, this does not always hold if p is composite!

Next, given integers a and $m > 1$, we defined a **square root of a modulo m** as an integer solution to the equation

$$x^2 \equiv a \bmod m.$$

We found that 4 has square roots 2 and 3 modulo 5. These can also be described as ± 2 and ± 3 modulo 5, noticing immediately that in general, if b is a square root of a modulo m, then so is $-a$. We found, however, that 3 has no square roots mod 5; i.e., 3 is *not a square* modulo m.

We found that ± 1 are the only square roots of 1 modulo 17, or even modulo any prime p. Also, modulo any integer, we at least have two square roots of 1 (± 1) and one square root of 0 (0).

In fact, 71 has a square root modulo 77. Howe to we find it?

We considered all primes congruent to each integer modulo $2, 3$, and 4. We also roughly described Dirichlet's theorem on the distribution of primes modulo different integers. Then we saw that if $p \equiv 3 \bmod 4$, then -1 cannot be a square modulo p!

Tuesday, February 4. We started class by reviewing the topics from last week, while going through a series of examples to see how everything connects with one another.

We set up the following problem: If we have an army of an unknown number of soldiers, but we know that the remainder when divided by 15 is 14, and by 17 is 1, than can we decide how many soldiers are in the army (assuming it appears that there are fewer than, say, 250)? This problem translates to finding a solution to the system of congruence equations $x \equiv 14 \bmod 15$ and $x \equiv 1 \bmod 17$.

This lead to the statement of the **Chinese Remainder Theorem** (CRT) in terms of congruences: If m and n are relatively prime integers, the system of equations $x \equiv a \bmod m$ and $x \equiv b \bmod n$ has an integer solution regardless of the integers a and b. Moreover, the solution is "unique modulo mn," meaning the following: (1) if x_0 is a solution, then if $x_0 \equiv y_0 \bmod m$, then y_0 must be a solution, and (2) if z_0 is a solution, then $z_0 \equiv x_0 \bmod m$.

We saw that *not* every system of congruences has a solution in the case that the moduli are relatively prime. We also calculated that our army has 239 soldiers.

After this, we proved the existence of a solution in the CRT, and proved part (1) in uniqueness. Part (2) is part of your first homework!

Tuesday, January 28 and Thursday, January 30. This week, Professor Marge Bayer was a guest lecturer. On Thursday, we had a short quiz on groups.

In class, we defined the **greatest common divisor** of two integers m, n, denoted $\gcd(m, n)$, or just (m, n), as the smallest positive *common divisor* (i.e., an integer d for which $d \mid m$ and $d \mid n$). We proved that if m and n are integers, and $m = nq + r$ for some integers q, r, then $(m, n) = (n, r)$.

After this, we stated the **Euclidean algorithm**, and explained why the previous result shows that it is a valid algorithm for computing a greatest common divisor. We also did an example of carrying out the Euclidean algorithm. **Bézout's theorem** says that if m, n are integers and $d = (m, n)$, then there exist integers a, b for which

$$am + bn = 1.$$

Often we call this equation **Bézout's identity**. We used "back substitution" in the Euclidean algorithm to find integers a and b in our example.

From here, we stated the **existence and uniqueness of prime factorization**, and proved it by induction.

We defined a **ring**, and gave examples; for instance, the integers, $\mathbb{Z}/m\mathbb{Z}$, rings of polynomials, and the collection of square matrices. A **unit** of a ring as an element that has a multiplicative inverse. We wrote out the multiplication table for $\mathbb{Z}/6\mathbb{Z}$, and found the units. We also noted the units of $\mathbb{Z}/4\mathbb{Z}$ and $\mathbb{Z}/9\mathbb{Z}$, and conjectured that the units of $\mathbb{Z}/m\mathbb{Z}$ are $[a]$, where $(a, m) = 1$ We proved this conjecture.

Finally, for a positive integer m, we defined the **Euler phi function** $\varphi(m)$ as the number of integers $1, 2, \ldots, m$ relatively prime to m. We found formulas for $\varphi(p)$ and $\varphi(p^k)$, if p is prime and k is a positive integer. Then we stated the fact that $\varphi(mn) = \varphi(m)\varphi(n)$, and started investigating why this might hold.

Thursday, January 23. We started class today by defining an **equivalence relation** on a set. After giving several examples (and non-examples!), we defined the **equivalence class** $[a]$ of an element a of the set S.

After this, we defined what it means for an integer a to **divide** anther integer b (often written $a \mid b$): $ak = b$ for some integer k. Then we defined what it means for two integers a, b to be **congruent modulo** another integer $m > 1$ (written $a \equiv b \bmod m$):

$$m \mid (b - a).$$

When $m = 2$, we figured out that two integers are congruent exactly if they are both even or both odd. In general, two integers are congruent modulo m if and only if they have the same remainder after dividing by m!

We proved that congruence modulo m is an equivalence relation. We call the equivalence class of an integer a its **congruence class**, and often denote it $[a]_m$, or just $[a]$ if the modulus m is understood. We described the congruence classes modulo $m = 2$ (the set of all even integers, and the set of all odd integers) and $m = 3$. We noticed that there are exactly m, and they can be written as $[0] = [m], [1], [2], \ldots, [m - 1]$. We defined $\mathbb{Z}/m\mathbb{Z}$ as the set of equivalence classes of the integers under congruence modulo m.

Finally, we defined operations of addition and multiplication on $\mathbb{Z}/m\mathbb{Z}$:

$$[a]_m + [b]_m = [a + b]_m$$
$$[a]_m \cdot [b]_m = [ab]_m$$

However, through examples, we noticed that it is not clear that this operation is not obviously *well-defined*, meaning that if $[a] = [a']$ and $[b] = [b']$, we must have that $[a + b] = [a' + b']$ and $[ab] = [a'b']$. We checked the first by hand, and left the second as homework.

Finally, we checked that $\mathbb{Z}/m\mathbb{Z}$ is group under addition, but is *not* a group under multiplication!

Tuesday, January 21. Today, we started class by going over the syllabus, and the material on the course website. We went into detail about the course expectations.

Next, we introduced the notion of a **group**, and while studying these objects, introduced some mathematical notation. Please interrupt me in lecture if you cannot remember what certain notation means! We gave several examples of groups, including the sets of integers $\mathbb{Z}$, rational numbers $\mathbb{Q}$, and real numbers $\mathbb{R}$ under addition. We noticed, however, that the inverse property does not hold if we instead consider these sets under multiplication. To rectify this, if S is a set with binary operation multiplication, we use the notation $S^\times$ to denote the subset of S of elements that have (multiplicative) inverses. Then as long as S satisfies the associative properties and has a (multiplicative) identity, $S^\times$ is a group under the operation of multiplication. For instance $\mathbb{R}^\times = \mathbb{R}\backslash\{0\}$, $\mathbb{Q}^\times = \mathbb{Q}\backslash\{0\}$, and $\mathbb{Z}^\times = \{1, -1\}$.

We also found a group in which the operation is not always *commutative*: The set of $n \times n$ matrices with real entries and nonzero determinant, under matrix multiplication. This is an example as above: the set of all $n \times n$ matrices that have multiplicative inverses. In fact, the subset of these with determinant 1 forms a **subgroup** of this group, meaning a subset that is itself a group under the same operation and identity.

After this, we proved that in a group, the identity is unique (there is only one), and each element of the group has a unique inverse. If the operation is denoted $\cdot$ or $*$, then we often denote the unique inverse of an element a as a^{-1}. After this, we showed that **cancellation** holds in a group: If a, b, and c are elements of a group $(G, *)$, then if $a * b = a * c$, then $b = c$. On the other hand, we found an example of matrices (that do not have nonzero determinant!) for which cancellation holds.